

RRQ OUTLINE

Title

On (date), Congress reauthorized Individuals with Disabilities Education Act (?). The revised law is very different from the previous version in at least one important respect. Whereas practitioners were previously encouraged to use IQ-achievement discrepancy to identify children with learning disabilities (LD), they now may use something called “responsiveness-to-intervention,” or RTI. In other words, RTI is a new, alternative method of LD identification. It is also a means of providing early intervention to all children at risk for school failure. The reauthorized IDEA now permits districts to use as much as 15% of its special education monies to fund early intervention activities. All this would seem to have implications for the number and type of children identified; the kinds of educational services provided; and who delivers them. RTI may be especially important to *Reading Research Quarterly* readers because roughly 80% of those with the LD label have been described as reading disabled (Lyon, 1995). With RTI, there may be a larger role for reading specialists, which, in turn, might affect pre-service and in-service professional development activities conducted by universities and school districts. Yet, with all this said, much needs to be understood to ensure that RTI provides effective early intervention and represents a valid means of LD identification. Below, we explain important features of RTI; why it has been promoted as a substitute for IQ-achievement discrepancy; and what remains to be understood about it before it may be seen as a valid means of LD identification.

What Is RTI?

Explaining the “R” in RTI

Selecting at-risk students. Practitioners use different assessments at different points in the RTI process. Before evaluating students’ responsiveness to interventions, a group of “at-risk” students is typically identified from which non-responders are likely to emerge. Identification of this group usually occurs in the first month of the school year. Practitioners may choose among several strategies to identify at-risk children. They can look at all students’ performance on the previous year’s high-stakes test in reading or math and choose a criterion such as “all below the 25th percentile are at risk.” Or, in the current school year, they may administer a standardized, norm-referenced measure to all students in a given grade, and those scoring below, say, the 25th percentile would be designated at risk. Perhaps the best strategy from a measurement perspective is to assess every student in the grade on a brief screening tool that demonstrates diagnostic utility for predicting end-of-year performance on reading and math assessments (in the elementary grades) or on local graduation requirements (at the secondary level).

Monitoring at-risk students. Once at-risk students are selected, their responsiveness to general education instruction is monitored. Again, there is more than one way of doing this. At the end of a relatively brief period (e.g., 8 weeks) of classroom instruction, the at-risk students may be administered a screening tool or a brief

standardized achievement test in the area of risk. Responsiveness may be defined as “a score above the 16th percentile.” An arguably better criterion would require practitioners to compare the at-risk students’ performance to (a) local or national normative estimates for weekly improvement or (b) criterion-referenced figures for weekly improvement. If “a” and “b” are unavailable, then “responsiveness” may be operationalized as some improvement (i.e., a slope greater than the standard error of estimate of the slope). At-risk children judged unresponsive to classroom instruction are given more intensive instruction at a second tier, or level, either in the classroom or outside the classroom (see below). And their performance during this more intensive, second-tier instruction may be assessed in a manner similar to how their performance was assessed during first-tier instruction.

Much of RTI assessment, then, is progress monitoring, which assists practitioners’ efforts both to provide early intervention and to identify special-needs children. Regarding early intervention, progress monitoring should be understood in part as formative evaluation--formative in the sense that teachers use the data to determine whether they need to change their curricula, materials or instructional procedures. Progress monitoring also generates diagnostic information that helps practitioners make classification and program placement decisions (e.g., moving a student from tier 1 to tier 2). Finally, and most importantly, progress monitoring is *dynamic* because the associated metric is change in student learning—change in level or change in rate. It contrasts with traditional assessment’s focus on more *static* learner characteristics such as IQ that are presumed unalterable.

Explaining the “I” in RTI

Focus on reading instruction. Most look to RTI as a means of delivering early intervention to address academic problems, not school behavior problems. Specifically, the interventions—at least those currently described in writing by practitioners and researchers--typically target reading problem and, more specifically, early reading problems (e.g., Al Otaiba & Fuchs, in press; McMaster, Fuchs, Fuchs, & Compton, in press; O’Connor, 2002; Vaughn, Linan-Thompson, & Hickman, 2002; Vellutino et al., 1996). This is not accidental. Many of the same policymakers behind RTI have also been responsible for Reading First. Reading First, a major component of No Child Left Behind (P.L. 107-110, 2001), reflects the influence of two well-known research syntheses on early reading (Snow, Burns, & griffin, 1998; NRP, 2000). It requires schools to use scientific knowledge to guide their selection of core reading curricula, and to use valid screening measures and progress monitoring to identify students in need of more intensive reading instruction.

Multi-tiered instruction. As indicated, RTI is also multi-tiered. Different versions have anywhere from 2 to 4 levels. The nature of the academic interventions change across the tiers, usually becoming more intensive as one goes from tier 1 to tier 2 and so forth. Increasing the intensity of instruction can be defined as (a) using more systematic and explicit instruction; (b) conducting it more frequently; (c) adding to its duration; (d) creating smaller and more homogenous student groupings; or (e) deploying instructors

with greater expertise. Some (e.g., Grimes, 2002) regard these tiers as substituting for the comprehensive evaluation now afforded all children suspected of having a learning disability. Others (e.g., Division of Learning Disabilities of the Council for Exceptional Children, date; Telzrow, McNamara, & Hollinger, 2000) see the tiers as supplementary, or as a component of a more comprehensive and traditional evaluation. The first group views RTI mostly in terms of providing prevention, and they advocate more tiers. The second group views RTI more as an identification and classification procedure; they argue for fewer tiers.

Problem-solving. To date, practitioners conducting RTI use a “problem solving” approach to intervention. Researchers are partial to the use of standard treatment protocols. To explain the problem-solving approach, we turn to the work of practitioners in the Heartland (IA) Educational Agency. Heartland is the largest of the state’s 15 educational agencies (Ikeda, Tilly, Stumme, Volmer, & Allison, 1996), with responsibility for providing training and technical assistance to 56 districts and 350 schools (Ikeda & Gustafson, 2002a). Statewide school reform began in 1985 and has continued in earnest since 1992 (Ikeda & Gustafson, 2002a). As part of reform making, Heartland staff developed a four-level problem-solving model partly to “provide educational assistance in a timely manner” (Grimes, 2002, p. 8). According to Ikeda and Gustafson, at Level 1, a teacher confers with a student’s parent(s) to try to resolve academic or behavior problems. At Level 2, the teacher and his or her school’s Building Assistance Team (BAT) meet to identify and analyze problems and to help the teacher select, implement, and monitor the effectiveness of an intervention. An absence of success at this level triggers the involvement of Heartland staff, which defines Level 3. Heartland staff are mostly doctoral-level or masters-level school psychologists and special educators who use behavioral problem solving to refine or redesign the intervention and coordinate its implementation from that point on. Finally, at Level 4, special education assistance and due process protections are considered.

At each level, the problem-solving process is meant to be the same: practitioners determine the magnitude of the problem; analyze its causes; design a goal directed intervention; conduct it as planned; monitor student progress; modify the intervention based on student responsiveness; and evaluate its effectiveness and plot future actions (cf. Grimes, 2002). Throughout this problem-solving process, and across the four levels, “data about a student’s responsiveness to intervention becomes the driving force” (Grimes, 2002, p. 4). Teachers and Heartland staff are directed to compare the student’s performance level and learning rate with what is expected of students in the same classroom. It is the student’s relative classroom performance, rather than test performance, that determines responsiveness/ unresponsiveness and, eventually, special education eligibility. In other words, Heartland’s multilevel approach is meant to provide teachers with repeated opportunity and increasing levels of support to help students become responsive to instruction. Students who remain unresponsive are eligible for special education.

According to Ikeda and Gustafson (2002b), Heartland’s recursive, multilevel problem-solving model has several advantages over the traditional psychometric

approach to eligibility decision making. First, the model directs practitioners to collect more thorough and instructionally relevant assessment data. Because cut scores (e.g., a standard score of 80 on an achievement test) have been deliberately excluded from consideration, the problem-solving model “force[s] decisions to be made using multiple sources that show a convergence of evidence” (Ikeda & Gustafson, p. 6, emphasis in original; also see Ikeda et al., 1996). Second, say Ikeda and Gustafson, use of the model leads to a match between the presenting problem, assessment methods, and interventions. Third, the model is non-categorical. Children found eligible for special education services are not labeled LD, mentally retarded, or behavior disordered. This facilitates the provision of services to at-risk students who might otherwise not receive them.

The problem-solving approach to intervention has been adopted by a large handful of school districts (Canter, 2003), including the Minneapolis Public Schools (see Marston, Muyskens, Lau, & Canter, 2003). Its popularity among practitioners is no doubt partly due to its idiopathic nature: For each child, an effort is made to personalize assessment and intervention. But this individualized approach is a potential weakness as well as a potential strength. The problem-solving approach presupposes considerable expertise among practitioners in assessment and intervention. They must be skillful in numerous types of assessment and intervention; they must have the clinical judgment and experience to know which assessments and interventions to apply; and they must have the knowledge, discipline, and opportunity to carefully and accurately measure the effectiveness of selected interventions, which are sometimes a unique hybrid of two or more evidence-based practices that in combination have not been researched. Which leads us to the standard treatment protocol approach.

Standard treatment protocol. A standard treatment protocol is an alternative to the problem-solving approach to intervention. Whereas the problem-solving approach differs from child to child, a standard treatment protocol is just that...standard. Implementation usually involves an intensive fixed duration trial (e.g., 10-15 weeks) of an unchanging treatment delivered in small groups or one-on-one (e.g., Al Otaiba & Fuchs, in press; McMaster et al., in press; Vaughn, Linan-Thompson, & Hickman-Davis, 2002; Vellutino et al., 1996). If the student responds to a treatment trial, she is seen as remediated and disability-free, and she is returned to the general-education classroom for instruction. If she is unresponsive, a disability is suspected and further evaluation is warranted. This approach is illustrated by the work of Vellutino and colleagues (e.g., Vellutino et al., 1996).

Vellutino et al. (1996) asked first grade teachers to nominate their poorest readers at the beginning of the school year. At the start of the second semester of first grade, Vellutino et al. assigned the children to tutoring and contrast groups. The tutored children received a 30-minute, one-to-one intervention for five days each week for most of the semester. This amounted to between 70-80 tutoring sessions. Instructors were certified and experienced teachers who had received 30 hours of training in Vellutino’s tutoring program. Tutoring focused on phonemic awareness, decoding, sight word practice, comprehension strategies, and reading connected text. In fall of second grade, tutored

students who were below the 40th percentile on the Basic Skills Cluster participated in an additional 8-10 weeks of tutoring.

Between winter of first grade and spring of second grade, students were repeatedly administered the Word Attack and Word Identification subtests of the WRMT-R. Responsiveness to intervention was based on slopes derived from linear regression analysis of the WRMT-R data. Vellutino et al. (1996) described four levels of responsiveness: “very limited growth,” “limited growth,” “good growth,” and “very good growth.” Two-thirds of the tutored students demonstrated “good growth” or “very good growth” after one semester of tutoring in first grade. Indeed, they had basically caught up to their normal-reading classmates. Vellutino et al. suggested that these students had not really been reading disabled but “instructionally” disabled. By contrast, the remaining one-third of the tutored readers remained in the lowest 30th percentile on the two subtests of the WRMT-R despite receiving tutoring in both first and second grade. The researchers described these children as “difficult-to-remediate.”

Intervention-as-test. As with assessment, intervention—be it problem solving or standard protocol--serves RTI’s two purposes: to provide struggling students with early and presumably effective instruction, and to provide a valid means of assessing learner needs. The “I” in RTI has, in a sense, become the test stimulus. Children’s level or rate of growth—their degree of responsiveness—is the test performance. Although many RTI proponents are critical of the traditional psychometric approach, they, like the traditionalists, must still prove the validity of their methods; in their case, “intervention-as-test.” A principal means of demonstrating this is by using evidence-based interventions and by ensuring that, in each and every instance, the intervention is implemented with fidelity. In this regard, the standard treatment protocol may have an advantage over problem-solving: Everyone knows what to implement since there is but one protocol; and it is easier to “scale-up” its use among practitioners and to assess its fidelity.

In some places, school psychologists are responsible for intervention implementation; in others, reading specialists; in others, classroom aides.

Case Study of the RTI Process (see Figure 1).

Why RTI?

Special Education Costs

For decades, people have been frustrated by the LD construct generally, and by IQ-achievement discrepancy particularly. One prominent reason: economics. In a sense, LD became too successful for its own good—if success may be defined by the number of children given the label. As shown in Figure 2, shortly after LD was legitimized as a special-education category in Public Law 94-142 (a.k.a. “Education of All Handicapped Children Act of 1975”), the proportion of children with LD in the general population skyrocketed from less than 2% in 1976-1977 to more than 6% in 1999-2000. This has

proved very expensive to school districts because, on average, it costs 2-3 times more to teach children with disabilities. Not long ago, New York City was spending \$1.67 billion, or 22 cents of every school dollar, on special education, partly because of a 1980 consent decree issued by a federal judge in a class action lawsuit filed on behalf of Jose P., a Puerto Rican student with disabilities. In a relatively recent year, the New York City's special education system employed one-quarter of all school employees and provided services to 130,037 students or 13% of the city's one million school children (National Association of State Boards of Education, 1991). The board of education's budget director, Leonard Hellenbrand, was quoted as saying that non-disabled students suffer as a result of special education spending. "Kids that don't have court orders in their hands are dead meat," he said (Dillon, 1994).

IQ-Achievement Discrepancy

IQ-achievement discrepancy—the most widely used method of LD identification—has often been viewed as the “culprit,” which brings us to a second reason for dissatisfaction with the LD construct. The discrepancy approach has been frequently criticized as a-theoretical (Lyon, date; Wilson, date). Moreover, according to some, this absence of theory has encouraged or permitted states and school districts within states to specify discrepancy differently. Today, discrepancy varies in terms of how it is computed (e.g., standard IQ score minus standard achievement score vs. the regression of IQ on achievement), its size (e.g., 1.0 SD vs. 2.0 SDs), and which specific IQ and achievement tests are used. Not surprisingly, these varying definitional features and criteria have led to large inconsistencies in LD prevalence between states (e.g., Scruggs & Mastropieri, 2002).

Such inconsistency in the definition of IQ-achievement discrepancy and varying prevalence rates—as well as the outright noncompliance by some school-level personnel with state and district guidelines (cf. Gottlieb, Alter, Gottlieb, & Wishner, 1994; MacMillan, Gresham, & Bocian, 1998; Shepard & Smith, 1983)—have contributed to a widespread view that the LD designation is whatever teachers and parents want it to be (e.g., Coles, 1987; Reynolds, Wang, & Walberg, 1987; Skrtic, 1991; Sleeter, 1998; Ysseldyke, Algozzine, & Epps, 1983). A more damaging assertion, perhaps, is that the IQ-achievement discrepancy approach fails to distinguish a qualitatively different and more deserving subgroup of students from a much larger group of low achievers. Recent studies suggest that young, poor readers with and without an IQ-achievement discrepancy perform similarly on many reading-related cognitive tasks (e.g., Fletcher et al., 1994; Foorman, Francis, & Fletcher, 1995; Stanovich & Siegel, 1994), as well as demonstrate phonological-processing deficits, which are correctable with appropriate instruction (e.g., Fletcher, 1995; Morris et al., 1998; Siegel, 1989, 1999; Stanovich, 1999; Torgesen, Morgan, & Davis, 1992; Vellutino et al., 1996). Thus, say the critics, thanks to the IQ-achievement discrepancy approach, the LD label is not just arbitrarily assigned, it is unfairly withheld from children who are as needy and deserving as those to whom the label is given.

Concerns that discrepancy is a-theoretical and arbitrary, and that some of its basic assumptions have not been unsupported by research, have crystallized for many into two major criticisms of it. First, it represents a “wait-to-fail” model antithetical to early intervention; that is, children must fall dramatically behind their peers in academic achievement to qualify as “LD.” This usually requires repeated failure over a long time. Indeed, the modal age at which children receive the label is 12 years—or fifth or sixth grade (see Figure 3). Hence, the derisive “wait-to-fail” characterization.

Second, critics of discrepancy say that the low achievement of so-called children with LD is presumed to reflect disability when, more times than not, it reflects poor teaching. The discrepancy approach, in short, provides no fail-safe to guard against this wrongful inference. RTI, advocates promise, will promote appropriate use of evidence-based instruction in tiers 1 and 2 and, by so doing, will decrease the numbers of children incorrectly identified as disabled and in need of special education.

Unanswered Questions and Issues

There is considerable hype surrounding RTI. One such example was a recent posting on a listserv of the American Psychological Association’s Division 16 (School Psychology): “There is strong empirical support that the RTI approach is sufficient for a comprehensive evaluation of LD, if the appropriate assessment and intervention procedures are put in place.” Such exaggeration notwithstanding, much still needs to be understood about RTI.

Problem Solving and Standard Protocol: Implications for “Responsiveness” to Instruction

Problem solving and standard protocol approaches differ conceptually and with respect to technical issues. Conceptually, each involves a different implicit meaning of “responsiveness/non-responsiveness.” Use of a standard treatment protocol provides a rigorous test for non-responders and the presence of disability. Students like those in the Vellutino et al (1996) study, or in the Vaughn et al. (2002) investigation, participate in a research-backed, intensive, and iterative instructional process. In such circumstances, it makes little sense to point to poor or inadequate instruction as a cause of non-responsiveness. It makes more sense to consider disability as a cause. At the same time, use of a standard treatment protocol raises the question, Is it possible that some children who are responsive to instruction in a second or third tier of a multi-tier approach still have disabilities and, once returned to their classroom instruction without the intensity and systematicity of the standard treatment protocol, again demonstrate the same learning problems that first marked them as candidates for participation in the standard treatment protocol? In short, whereas the standard treatment protocol approach is likely to identify “true” non-responders, is it also likely to identify “false” negatives? Evidence for this may be found in the Vaughn et al. study. A subset of children who met criteria for dismissal from intensive tutoring subsequently failed to perform adequately back in their classrooms and eventually required additional tutoring.

By contrast, an at-risk student's responsiveness to classroom instruction with individualized adaptations (i.e., with teachers adapting instruction to accommodate the unique learning needs of individual students) suggests that adequate learning will continue without further intervention. Students participating in generally effective classroom instruction with adaptations, whose learning is much worse than that of their peers, are likely to require the intensity of instruction special education is meant to provide. Moreover, defining "intervention" and "responsiveness/non-responsiveness" in the mainstream classroom presumes that disability should be assessed as it occurs under "normal" conditions: in the normal or mainstream setting. This parallels contexts in which other psychological conditions are diagnosed. Ruling out disability only after intensive effort improves a condition seems akin to concluding a patient never had cancer because surgery restored her to health.

Regarding technical issues, problem solving and standard treatment protocol approaches pose different challenges. As described, problem solving occurs in the classroom; standard treatment protocols are typically implemented outside the classroom in a small-group context. Relying on general education to assess responsiveness to instruction has the advantage of a normative framework referenced to the typical population. That is, responsiveness to generally effective instruction can be estimated for all students so that a normative profile can be generated to describe the full range of response. With general education instruction as the intervention, traditional cut points (e.g., 1.5 standard deviations below the mean) may be used to define disability. Such an approach requires measurement of all students. By contrast, it seems unlikely that a normative framework may be applied to the standard treatment protocol approach. Thus, logistics and logic seem to argue against exposing the full range of students to an intensive tutoring regimen for the purpose of producing a normative profile. In all likelihood, practitioners would need to rely on a normative framework restricted to very poor readers, a proposition requiring empirical validation.

In comparison to the standard treatment protocol approach, problem solving is usually associated with a lower bar to determine non-responsiveness and easier access to special education. Assuming that special education is effective, this helps ensure that all children with special needs receive appropriate services. Yet, relatively easy access to special education can, in some cases, reflect a "rush to judgment" and identification of "false positives," or children who are incorrectly identified and labeled. The standard treatment protocol approach, by contrast, tends to provide more intensive instruction, to which many children respond positively. However, it is more likely to produce "false negatives," or students with disabilities who improve during intensive tutoring only to be returned to general education where they fail once again. In selecting between these two approaches, it may be necessary to determine whether one's primary intent is identification or prevention.

Measuring and Defining Non-Responsiveness

Regardless of which RTI approach is adopted, two components of the assessment process must be specified. First methods must be determined for measuring students'

responsiveness to instruction. Second, once student responsiveness has been quantified, a criterion must be applied for defining non-responsiveness. Below such a criterion, students are identified as having reading disabilities.

Various methods are available for specifying these two assessment components. As described, Vellutino et al. (1996) tested students on subtests of the Woodcock Reading Mastery Tests several times over the course of a multi-year study. To establish a cut-point for responsiveness, they rank-ordered slopes representing children's growth in responsiveness to tutoring, performed a "median split" on the slopes, and designated the bottom half as non-responsive. Similarly, Torgesen and colleagues (2001) evaluated student performance at the end of treatment on the subtests of the Woodcock Reading Mastery Tests, designating non-responsiveness as failing to achieve "normalized" status; that is, a word-reading standard score of 90 or better. Good, Simmons, and Kame'enui (2001), like Torgesen et al., also specified non-responsiveness in terms of posttreatment status. However, their approach involves a criterion-referenced "benchmark" associated with future reading success.

Finally, Speece and Case (2001) took yet a different tack. They adopted frequent measurement using curriculum-based measurement so that non-responsiveness could be identified earlier in the school year than was possible with the Vellutino et al., Torgesen et al., or Good and associates methods. Speece and Case applied a "dual discrepancy" criterion (also see Fuchs & Fuchs, 1998; Fuchs, Fuchs, & Speece, 2002). Non-responders were students whose slope and level of performance fell at least 1 standard deviation below their class mean. This dual-discrepancy approach could also be determined with respect to school, district, or national norms or using benchmark cut-points associated with future reading performance.

These alternative methods produce different prevalence rates of reading disability and different subsets of non-responsive children. This is very important because a major criticism of IQ-achievement discrepancy as a method of LD identification has been the unreliability of the diagnosis. Practitioners relying on an assortment of assessment procedures in an RTI framework may produce similarly unreliable diagnoses. To develop more consistent identification procedures, researchers must explore the soundness of various methods. At the same time, however, different assessment methods demonstrate differential utility in distinguishing responsive and non-responsive groups on different components of beginning reading. For this reason, consistency in identifying non-responders across the various components of beginning reading skill is an important criterion for selecting a valid assessment approach. (In our own research—Fuchs, Fuchs, and Compton...)

Research should address these questions:

1. How should "non-responsiveness" be operationalized? What method of non-responsiveness should be used? Median split on performance level or on growth? Torgesen's "below 90th percentile? Good's benchmark for words correct per minute? A dual-discrepancy that accounts for performance level and growth? And

what measures should be used to define reading and math at different grade levels?

2. Implicit in RTI is that the intervention is a “test.” As with traditional psychometric models, RTI practitioners must be concerned about the validity of their interventions. Without it, RTI as a means of LD identification will not succeed. How to ensure fidelity and effectiveness of interventions? Standard protocols have validity data; not so problem solving. Yet, standard protocol is limited because it is standardized. The weakness and strength, then, of the standard protocol approach is standardization. (Tension or trade-off between standardization and individualization.)
3. What is the appropriate role of special education in a district that uses a multi-tier RTI model?