The California School Psychologist

2005, Volume 10

CONTENTS

The California School Psychologist, Vol. 10, pp. 21-32, 2005

Using Response-to-Intervention to Enhance Outcomes for Children

Amanda M. VanDerHeyden and Shane R. Jimerson
University of California, Santa Barbara

Response-to-Intervention (RTI) models have substantial promise for screening, intervention service delivery, and to serve as catalysts for system change to enhance the educational outcomes of children. RTI represents a more flexible service delivery model; however, it is essential to articulate how RTI can be effectively implemented and demonstrated to be functionally meaningful. This article provides a brief review of probable advantages and possible challenges of implementing RTI and also provides an overview of data emerging from the Screening to Enhance Equitable Placement (STEEP) problem-solving model. Further research is necessary to clearly delineate the purpose(s), operationalize procedures and judgments, and evaluate the decision-making utility of the RTI models in practice.

Key Words: RTI, STEEP, Implementation, Models

Along with the 2002 Reauthorization of the Elementary and Secondary Education Act — No Child Left Behind (NCLB) — policy-makers, parents and educational professionals are facing the most significant changes in federal educational policy in perhaps 30 years (CASP Board of Directors, 2003; p. 1).

Response to Intervention (RTI) has emerged as a topic of great interest to researchers and practitioners working in schools. Recently RTI has garnered both attention and momentum of policy-makers in response to growing research and practical implementation (Fletcher, & Reschly, 2005; Gresham et al., 2005; Kavale, Kaufman, Naglieri, & Hale, 2005; Shrank et al., 2005). Extensive field trials implemented in Iowa and Minnesota have demonstrated the viability of RTI within the special education screening and eligibility determination process (Ikeda, Tilly, Stumme, Volmer, & Allison, 1996; Marston, Muyskens, Lau, & Canter, 2003; Reschly & Grimes, 1991). Consensus between special and general education reform agendas has propelled RTI into the forefront as a viable alternative to traditional diagnosis of high-incidence disabilities, particularly learning disability (LD). The criteria in the reauthorized IDEA guidelines for identifying LD state that: (a) a severe discrepancy between achievement and intellectual ability shall not be required; and (b) a response to intervention (RTI) may be considered.

Several lines of research have contributed to the development and popularity of RTI as an alternative assessment approach. The reliability, validity, and utility of the discrepancy-based formula for identifying LD has been questioned conceptually, empirically, and pragmatically (Fuchs, Fuchs, Mathes, Lipsey, & Roberts, 2001; Gresham & Witt, 1997). The problem of disproportionate identification of children with LD by race and gender has continued to be an exigent issue (Donovan & Cross, 2002; Holtzman & Messick, 1982). Simultaneously, the utility of curriculum-based measures (CBM) for (a) identifying children not likely to benefit from the general education curriculum without assistance, (b) predicting important long-term outcomes, and (c) tracking individual student growth and informing instructional programming has been established. Intervention research, most particularly in the area of early reading, has demonstrated the power of early intervention to remediate current and prevent future deficits, while indicating a lack of differentiated instruction and effectiveness for children placed

Address correspondence to Amanda VanDerHeyden; University of California; Gevirtz Graduate School of Education; Santa Barbara, CA 93106-9490. E-mail: amanda@education.ucsb.edu.

into special education under the high-incidence categories (Kavale & Forness, 1999).

These data and others have led to consensus among scientists and policymakers concerning the need for change. Whereas there has been some consensus concerning the need for change, there has not been consensus on how this change can best be achieved. Changes in the identification process will likely fall somewhere along the continuum of weak or ineffective pre-referral interventions implemented with low integrity to demonstrating special benefits for those ultimately identified and placed into special education classrooms (Fuchs & Fuchs, 1998). Whereas RTI has considerable promise as a tool within special and general education, it is a vulnerable construct if misapplied. The following explores probable advantages of RTI models, the challenges that have yet to be resolved with the wide-scale use of and decision-making under RTI, and describes one RTI model (i.e., Screening to Enhance Equitable Placement) that has been used successfully in several schools.

Probable Advantages of RTI

Properly conceived and implemented RTI offers several advantages over other methods of identification. In this section, we present probable advantages of RTI models. Associated challenges are presented in the next section. It is important to address the challenges because there is the ever-pressing concern that the advantages are possible but not absolute without effective implementation.

Improved treatment validity. One of the primary advantages involves its direct link to treatment or consequential validity. Messick (1995) and others (e.g., Hayes, Nelson, & Jarrett, 1987) have argued that the value of an assessment is tied to the consequences of its use with children. Because RTI provides assistance to children to determine whether more assistance should be provided, adults are focused in their efforts to properly articulate a concern, develop targeted intervention to resolve the concern, and collect information to determine whether the concern has been adequately addressed or whether different solution efforts need to be implemented. This approach changes the goal of assessment from what some have described as "admiration" of the problem (Gresham, VanDerHeyden, & Witt, in submission) to problem-solving. RTI provides an opportunity for psychoeducational assessment and assessors to "aim at the target" (Macmann & Barnett, 1999) of producing real improvement for children in the classrooms in which they must function and learn.

Contextualized decision-making. RTI emphasizes the pre-referral conditions (child and environment) and this context becomes part of the decision-making equation. RTI models require that child assessment become context specific and represent some estimate of the degree to which the child is likely to attain important educational outcomes in the general education setting in the absence of special services (this is ultimately the judgment teams are being asked to make). This approach of assessing child performance within context allows practitioners to quantify (a) the state of instructional affairs in the child's regular education environment and (b) potential learning given optimal instructional conditions. This information allows practitioners to improve general education programming and also determine whether children are likely to continue to struggle without special help or whether instruction has been inadequate. Hence, children are likely to receive help in the general classroom environment much more quickly under RTI models. If the intervention is effective, future deficits may be prevented at the individual child level (Lennon & Slesinski, 1999) and teachers may use the same strategies to assist other children in the same classroom.

Improved identification accuracy for LD. Because intervention becomes a specified, operationalized variable under RTI models, false positive identification errors should be reduced dramatically. For this advantage to be realized, implementation must occur correctly and decision-making procedures, shown to be reliable and meaningful, must be adequately specified. As data-gathering proceeds, it will be important to identify acceptable standards against which to judge the value of an RTI model. The

present discrepancy-based classification of LD is not an acceptable standard for comparison as it has been found to have classification agreement rates barely above chance levels (Macmann et al., 1989). The discrepancy model has been criticized particularly for failing to distinguish between low achievers and children with LD (Gresham & Witt, 1997). When implementing RTI, because instruction is manipulated to judge its effect on learning directly, low-achieving children who are low-achieving because they have not had adequate instruction (previously false positive identification errors) are less likely to be identified as having LD. Further, ensuring adequate instruction as a prerequisite to individual child assessment will positively affect disproportionality (Marston, et al., 2003; VanDerHeyden & Witt, in press). The use of universal screening, with adequately specified criteria for selecting at-risk groups of children, delivering intervention strategies of sufficient power, and judging the effects of remedial efforts can protect against over- and under-identification errors. Removing the current reliance on teacher identification and requiring direct measures of child performance in context will enhance identification accuracy (Macmillan, 1998; Macmillan et al., 1998; Macmillan & Speece, 1999; Marston, Mirkin, & Deno, 1984; Shinn, Tindal, & Spira, 1987; VanDerHeyden & Witt, 2005).

More effective intervention. RTI is likely to facilitate less restrictive interventions and placements for children. Requiring universal screening and structured intervention services in general education classrooms conducted with sufficient control to permit reliable and meaningful decision-making will result in a broader range of supports within general education. The school psychologist's role is likely to shift to a greater emphasis on instructional consulting and contextual decision-making. This shift is likely to create a need for graduate programs to enhance the skill sets of practitioners to provide such services in schools. This shift also will have tremendous benefits for systems and children. If effectively implemented, RTI allows school psychologists to bring their expertise to bear on assessment strategies at the classroom level and assist teachers to use data formatively to enhance their instructional programming. Because interventions are delivered in a structured fashion, they are more likely to be effective. Structured interventions are those that (a) allow for some estimation of the likelihood of the intervention being effective under specified stimulus conditions, (b) allow for some estimation of the degree of match between the stimulus conditions for an individual child and the intervention selected, (c) allow for estimation of the integrity with which the intervention is delivered, and (d) allow for sensitive ongoing progress monitoring data.

Possible Challenges of RTI Models

For RTI to provide maximum benefits several challenges need to be resolved. Empirical studies in the field are needed to determine which components are needed for maximum effectiveness.

Decision-making criteria under RTI models must be operationalized and validated through research. Others have noted that determining the purpose of RTI will be critical to determining how implementation should proceed (Fuchs, 2003). Selection of the reference group necessarily reflects the identified purpose of RTI. For example, if the focus is on prevention, then less stringent initial cutpoints for identification will allow practitioners to identify more children for supplemental services. More stringent initial cutpoints reduce the costs of intervention services potentially, but also reduce the degree to which general prevention effects can be obtained. Also, use of at-risk comparison samples will inflate the incidence estimates under RTI models (Vaughn, Linan-Thompson, & Hickman, 2003). Several approaches to identification through RTI have been examined, including, monitoring response to general education services (Case, Speece, & Molloy, 2003; Speece, Case, & Molloy, 2003), monitoring response to intervention provided to a small subset of the population following identification through a problem-solving model of assessment (VanDerHeyden et al., 2003), and monitoring response to supplemental instruction (Vaughn et al., 2003).

Effective intervention delivery poses new challenges for teams. Adequate control of relevant intervention variables will be a key challenge under RTI models. The selected intervention must be one that, if properly implemented, is likely to effectively remediate the problem to which it is applied. The intervention, once properly specified, must be implemented with integrity and with sufficient frequency, intensity, and duration to permit effectiveness. These variables are necessary to eliminate confounds when a lack of response is observed. To conclude that the child has demonstrated an unsuccessful RTI and thus, should be considered for special education services, the intervention must have been properly identified and sufficiently implemented to rule out alternative explanations for the outcome (e.g., poor integrity, insufficiently powered intervention). Additional variables needing further research include which child behaviors are most important to measure (short term growth on instructional level task versus child response under conditions that differ from intervention conditions and more closely reflect conditions in the regular classroom), how child change or lack of change in RTI is best reflected (level of performance, trend of performance, combined level and trend of performance). Identifying which behaviors to measure and under what conditions to measure them (Daly et al., 2004; Fuchs, 2003; Good, Simmons, & Kame'enui, 2001) will be central to predicting generalized outcomes and attaining treatment validity under RTI. There is some precedent for using level and trend of performance during controlled intervention (Fuchs & Fuchs, 1998; VanDerHeyden et al., 2003) or general education instruction (Case et al., 2003; Speece et al., 2003) to judge RTI. The degree to which the age of the child, prior instructional experience, and other variables (i.e., intervention power) influence slope will be important areas for future investigation. Because there is no incontrovertible index of LD, research findings may reflect the diagnostic realities that exist. Hence, the "real" effect of RTI may be to prevent the many false positive errors in identification that previously occurred as a result of de-contextualized assessment and inadequate pre-referral protections.

IMPLICATIONS OF IMPLEMENTING RTI MODELS

What value does the concept of RTI bring to the equation of prevention efforts in early academic skill areas? The concept of RTI offers the promise of a controlled demonstration of potential growth rates. Knowing potential growth rates allows practitioners, program administrators, and parents to understand child performance within the learning context. RTI allows those who work with children to examine current learning conditions and how those differ from optimal learning conditions implemented during an RTI trial, child growth given current learning conditions relative to child growth given optimal learning conditions during an RTI trial. Child-environment fit can be evaluated to enhance child performance in the regular learning environment or a decision can be made to provide supplemental or specialized services to augment the regular learning environment. System change can occur when RTI indicates that the current learning conditions differ from conditions arranged during an RTI trial and that the majority of children exposed to RTI show optimal growth.

Fuchs, Mock, Morgan, and Young (2003) categorized existing published RTI efforts into two approaches. One approach was described as the standard protocol-based intervention approach illustrated by the work of Torgeson and colleagues (Torgeson et al., 2001) and Vellutino and colleagues (Vellutino et al., 1996, 1998). The second approach was described as problem-solving models of assessment, illustrated by the large scale efforts in Minnesota (Marston et al., 2003) and Iowa (Ikeda, Tilly, Stumme, Volmer, & Allison, 1996). D. Fuchs et al. (2003) noted the promise of standard approaches to intervention and characterized problem-solving models as more susceptible to individual differences in implementation. In their analysis, D. Fuchs and colleagues predicted that the standard approach would be more costly and time-consuming but would be likely to maintain consistently a higher level of quality and therefore, be more effective.

It is possible that a standard approach could be used within a problem-solving model of assessment (VanDerHeyden et al., 2003; VanDerHeyden & Witt, 2005), allowing for consistency in quality of intervention but doing so in a way that maintains efficiency by providing the standard intervention to a small subset of children deemed to be most at-risk. VanDerHeyden et al. (2003) found that 11% of children screened were identified to participate in individual intervention. Using a standard protocol to deliver an intervention that was designed to control for multiple causes of poor performance and based on principles of effective instruction, a high rate of intervention success was obtained within 5-9 sessions. Specifically, only 5% of children screened failed to respond successfully to intervention. Further, the children who did not respond successfully were proportionately similar to population base rates by race and gender. The intervention used for this assessment required 10 minutes per day to implement and included modeling, guided practice with immediate corrective feedback and prompting, timed independent practice for a score with incentives for improved performance, and delayed error correction. Successful interventions were those that produced a rate of growth comparable to the average growth of children not at risk in the school and level of performance that reached the instructional criterion imposed in the study (Deno & Mirkin, 1977) on a criterion-level, novel probe. Even with this stringent criterion for success, 95% of children screened ultimately responded successfully to this short-term intervention indicating that their learning difficulties could likely be addressed in general education (VanDerHeyden et al., 2003). This type of hybrid approach, although very simple, may allow for more accurate and more efficient identification of children who require more individualized problem-solving efforts. This approach merits further investigation. In the section that follows this screening model is described in greater detail.

SCREENING TO ENHANCE EQUITABLE PLACEMENT PROBLEM-SOLVING MODEL

Witt, Daly, and Noell (2000) described a problem-solving model of assessment called Screening to Enhance Equitable Placement (STEEP). The STEEP model has similarities to other problem-solving models that have been described in the literature (Fuchs & Fuchs, 1998; Good & Kaminski, 1996; Shinn, Collins, & Gallagher, 1998). Most problem-solving models use a series of scripted procedures to "filter" children at a school and identify those who might need special assistance to meet important academic objectives. Most models rely on universal screening using sensitive measures (often curriculum-based measurement), application of decision rules to identify at-risk children, supplemental instruction provided in general education (e.g., whole class, small group, and extra tutoring), followed by individualized intervention with frequent assessment to monitor progress and evaluate child performance relative to environmental requirements to determine that (a) the problem has been successfully resolved or (b) the problem merits further intervention. The components of the STEEP model are described below (and summarized in Table 1). Where the STEEP model is used, consideration for special education services requires that a child pass through three gates: poor performance relative to same-class peers and a functional instructional criterion (Deno & Mirkin, 1977); poor performance given powerful incentives; and poor performance given individual, protocol-based intervention performed with integrity for 10 consecutive days. Each gate is described in the following section.

Schoolwide universal screening. Curriculum-based assessment and measurement (CBM) probes are administered classwide in reading, math, and writing. CBM is a process of using a brief "test" derived from the curriculum to assess children. In math, for example, the child would typically complete a page of math problems representing one or more skills that the teacher has recently taught. Teachers can be trained to reliably administer CBM probes. In one hour, it can be determined how all children in a class are performing on basic math, reading, and writing skills. These estimates are

Table 1.
Summary of the Components of Screening to Enhance Equitable Placement (STEEP)

Stage I. Schoolwide Universal Screening
Curriculum-based assessment and measurement (CBM) probes are administered classwide (i.e., reading, math, and writing), performance of individual students is examined.
About 15% of children screened proceed to Stage II, the performance/skill deficit assessment.

Stage II. Performance/Skill Deficit Assessment
Students are offered a reward for exceeding their previous performance and are then re-examined using the classwide academic assessment probe that had been previously administered. Progress is monitored to determine whether other intervention strategies or assessments are warranted.
About 11% of children proceed to Stage III, individual intervention.

Stage III. Individual Intervention
Students exhibiting skill deficits (in classes where the majority of the class is performing at or above the instructional range) participate in daily individual intervention performed by the classroom teacher in the regular classroom setting during the regular school day. The student's instructional level is determined by sampling until the student scores in the instructional range. Progress is monitored to determine whether other intervention strategies or assessments are warranted.
About 3-5% of children do not respond successfully to short-term, protocol-based intervention delivered with integrity in the regular classroom setting.

relatively accurate and have been found to correlate well with performance on more comprehensive tests such as the Woodcock Johnson Psychoeducational Battery, Revised (WJ-R; Woodcock & Johnson, 1989) or Iowa Test of Basic Skills (ITBS; Hoover, Hieronymus, Frisbie, & Dunbar, 1993). Following schoolwide screening, problems are categorized as classwide (class median score falls below the instructional range described by Deno & Mirkin, 1977) or individual child problem (classwide median falls at or above the instructional range and individual child scores below the 16th percentile for his or her class and in the frustrational range described by Deno & Mirkin, 1977). Thus, two anchors are applied initially to define the problem (local anchor is classwide performance, broader anchor is instructional level performance that has been linked to functional competence). The instructional standards used for reading are 40-60 words read correctly per minute for children in first and second grades and 70-100 words read correctly per minute for children in third, fourth, and fifth grades. For math, the instructional standards are 20-40 digits correct per two minutes for first, second, and third grade students and 40-80 digits correct per two minutes for fourth and fifth grade students. When a classwide problem is identified, a classwide intervention is prescribed. The first step in performing classwide intervention involves finding the instructional level of the *class* by using CBM. Classwide intervention can take many forms but the STEEP model has used the following protocol most frequently: modeling the target skill, guided practice with frequent opportunities to respond and immediate feedback, and timed independent practice to yield a score for progress monitoring, followed by delayed error correction with a verbal rehearsal strategy. Classwide intervention is delivered at a difficulty level that matches the instructional level of the majority of students in the class using paired peer practice (e.g., classwide peer tutoring). The intervention requires 10 minutes daily. The classwide intervention is performed for 10 consecutive school days. Using a response-to-instruction model, the

children who continue to perform below the instructional standard when the class median has reached mastery *and* who are unlikely (based on their current growth rate) to reach the instructional range within the next week (i.e., another week's growth of that student's average growth during intervention would not increase the student's performance to the instructional range within one week) are identified and referred for the next phase, the performance/skill deficit assessment.

The above process illustrates what occurs if a classwide problem is identified. Using classwide intervention where classwide problems exist introduces variation into the performance of the children to permit more accurate identification of which individual children may be at risk. More accurate identification is facilitated because trend and level data can be examined to identify at-risk individuals (Fuchs & Fuchs, 1998; VanDerHeyden & Witt, 2005). Philosophically, where a classwide problem is observed, classwide intervention is appropriate, communicating that many children need help. Pragmatically, classwide intervention is more efficient than working individually with many low-performing children. If a classwide problem is ruled out following the classwide assessment, then children who perform below the 16th percentile for their classes and fall below the instructional range move to the next level of assessment, a brief assessment of the effect of incentives upon performance (i.e., performance/skill deficit assessment). When STEEP is used, approximately 15% of children are identified through the schoolwide screening to participate in further assessment (VanDerHeyden et al., 2003). The school-based team conducts the performance/skill deficit assessment.

Performance/skill deficit assessment. During the performance/skill deficit assessment, the consultant provides the student with a copy of the classwide academic assessment probe that had been previously administered. Students are told that they can earn a reward of their choice from the treasure chest by "beating their last score." This score is written in the top left-hand corner of the student's paper. Students are allowed to sample briefly the items in the treasure chest. The treasure chest is a small transparent box containing several small tangible items (e.g., pencils, balls, stickers, bracelets, and coupons for free time). The probe is then re-administered. The performance/skill deficit assessment for math and writing can be administered to groups of three to five students simultaneously, whereas the performance/skill deficit assessment of reading is administered individually in a quiet space on the school campus. This component requires no more than five minutes per assessment. Children whose performance improves to the instructional range to earn an incentive do not participate in further assessment. Children whose performance does not improve to the instructional range will participate in an individual intervention in their classroom. Approximately 11% of the total cases screened are found to exhibit a skill deficit that merits individual intervention or the third gate (VanDerHeyden et al., 2003).

Individual intervention. At this point (as soon as one week after the schoolwide screening), those children exhibiting skill deficits, in classes where the majority of the class is performing at or above the instructional range, participate in daily individual intervention performed by the classroom teacher (or teacher designee) in the regular classroom setting during the regular school day. In this stage, a standard protocol-based intervention that requires approximately 10 minutes is applied to estimate the child's potential for learning given relatively simple intervention procedures. The student's instructional level is determined by sampling backward through successively lower level materials until the student scores in the instructional range. Protocol-based interventions consist of four basic steps: modeling, guided practice with immediate error correction, independent timed practice with slightly delayed error correction, and the opportunity to earn a reward for "beating the last highest score." The interventions are protocol-based, designed to address multiple possible causes of poor performance, and designed to produce evidence (i.e., permanent products) that they occurred to allow for estimation of treatment integrity. Use of a standard intervention for screening has many benefits over the more

precise, but more time-consuming approach of individualized assessment and intervention that would occur with most problem-solving models. Because many children will respond to a standard intervention, the need to draft an individual intervention for every child is reduced to the smaller subset of children who do not respond to the standard protocol. Improved efficiency in the process may enhance the integrity with which the procedures occur. The school-based consultant (usually the school psychologist) collects the data weekly, quantifying two critical variables: the degree to which the intervention occurred correctly and the child's performance on a novel, instructional-level probe of the target skill and a novel, criterion-level probe of the target skill. The school-based consultant enters the data into a database and graphing tools automatically generate graphs for the teacher, principal, and consultant. If problems occurred in implementing the intervention, then the consultant re-trains the teacher. The purpose of this brief intervention is to measure the child's RTI. To measure RTI, a minimum of 10-15 consecutive intervention sessions, conducted with integrity, are required. Intervention progress for a student (i.e., slope derived from ordinary least squares regression) is evaluated relative to data pertaining to "normal" progress as well as data pertaining to the progress of children receiving specialized intervention. Structured decision rules are applied to determine if the child is a "responder" or "nonresponder" to the intervention or whether additional data are needed. Data on nonresponders are made available to the school-based team to assist in determining whether a child should receive an eligibility evaluation. Some estimates indicate that about 3-5% of children screened fail to respond sufficiently to brief intervention performed with integrity for 5-9 days (VanDerHeyden et al., 2003). Hence, if the decision rules of STEEP are applied in a school, the school-based team can expect approximately 5% of children to require extensive individualized intervention and potentially a full eligibility evaluation.

Data benefit both special and regular education. As a side-effect of completing each step, data are generated that are useful for both regular and special education decision-making and programming. For example, routine schoolwide screening permits identification of classwide, gradewide, and schoolwide problems that can be addressed through curriculum revision and supplemental instruction, and allows for monitoring to ensure growth for all students in the regular education curriculum. Once STEEP has been completed a list of children who may require special education services has been generated that is more accurate than teacher referral (VanDerHeyden et al., 2003), baseline performance prior to receipt of special services has been specified, and strategies have likely been identified that effectively produce growth for the child. For those who do not respond successfully to structured intervention, additional assessment can be performed to determine individual strengths and needs and inform educational programming and placement decisions.

Implementation data. VanDerHeyden, Witt, and Naquin (2003) reported the technical properties of decisions made based on STEEP relative to other methods of identification, especially teacher referral. All first and second grade children (N = 182 children; 364 screened cases in reading and math) were screened using several screening tools (schoolwide CBM, teacher identification, Comprehensive Inventory of Basic Skills, Revised (Brigance, 1999) subtests in reading and math, and state reading test). Any child who was identified through any of the screenings (n = 101 reading and math cases) participated in a more comprehensive assessment that included individual curriculum-based assessment with 5-9 sessions of individual intervention (RTI criterion). Additionally, all children in second grade were administered the ITBS and a randomly selected subset of children who failed at least one of the screenings were administered the WJ-R. Any child who failed the CBM screening participated in the remaining steps of STEEP for reading or math with one change. Only a single instructional session was included as part of STEEP in the study. Each case was coded as STEEP-positive (if the child

performed in the bottom 16% of the class and in the frustrational range, was not improved to the instructional range to earn an incentive, and was not improved to the instructional range given a single instructional session). All remaining cases were coded as STEEP-negative. To permit a series of accuracy analyses, the RTI criterion, performance on the ITBS, and performance on the WJ-R were used as standards for comparison. Because the results were similar across the outcome measures, only the analyses using the RTI criterion were reported. VanDerHeyden et al. found that STEEP produced many fewer errors in identification relative to the other screening devices. Use of STEEP produced improvements in both under- and over-identification errors. Further, accuracy of identification with STEEP was stable across environments, whereas teacher referral accuracy varied according to the general achievement level of the class (VanDerHeyden & Witt, 2005). STEEP was more accurate than the other screening devices, and was also more similarly accurate across race and gender. Alarmingly, prior to intervention 56% of children falling below the 16th percentiles in their classes were of minority ethnicity (initial cutpoint for determining risk), whereas only 15% of children attending the school were of minority ethnicity. Following 5-9 intervention sessions, 29% of those who did not respond to intervention were of minority ethnicity. Before intervention, 57% of the minority students in the school appeared in the high-risk category, whereas following intervention, only 7% of the minority students at the school fell into the high-risk category (nonresponders to intervention). One would expect that about 5% of children, irrespective of race might fail to respond to intervention. Hence, the use of RTI models has potential to reduce disproportionate identification (VanDerHeyden & Witt, in press).

STEEP has been operating in the Vail School District in Arizona for two years. During that time, the percentage of children in the district identified as LD (previously on an upward trend), has been reduced from 6% of children in 2001-2002 to 3.5% of children in 2003-2004. Additionally, gains on the high-stakes tests have been observed (VanDerHeyden & Burns, in press). Additionally, the number of children being evaluated has been reduced by greater than 50% and the percentage of those who are evaluated and qualify has been increased from a stable baseline of 50% to greater than 90% in 2003-2004.

PROBLEM SOLVING AND CALIFORNIA

In its position paper addressing the reauthorization of IDEA the California Association of School Psychologists (CASP) discuss critical constructs and principles (CASP Board of Directors, 2003) that will be important to consider in implementing response to intervention models. These include appropriate: (a) funding for educational services; (b) support for all children; (c) scientific framework for assessment and intervention; (d) clarification of learning disabilities; (e) empirically validated practices must be employed for both assessment and intervention; (f) eligibility for services must be established using objective, empirically valid definitions and assessment methodologies for all handicapping conditions; and (g) appropriately trained and credentialed personnel are essential at all stages of assessment and intervention. The CASP position paper highlights the anticipated implications for school psychologists in each of the above areas and emphasizes that, “Highly trained and skilled school psychologists are essential to the success of proposed educational reform of both general and special education” (CASP Board of Directors, 2003; p. 4). The STEEP problem-solving model described above is congruent with the three-tiered process for screening, early identification, and eligibility model delineated by the CASP Board of Directors (2003; see Table 2). The STEEP problem-solving process would occur during the tier-one phase of the CASP three-tiered process. Those students who did not respond to the STEEP strategies, would then proceed to the tier-two phase of the proposed CASP three-tiered process.

Table 2.
Detailed Description of a Three-Tiered Process for Screening, Early Identification, and Eligibility for Special Education Services in the Category of Learning Disabilities

• *Tier-one begins when pupils fail to make expected progress within general education.*

- Tier 1 interventions are preventative in that the process is invoked as soon as a child's acquisition of academic skills (particularly reading) is behind that of his/her peers.
- Students identified as functioning below the instructional range must receive timely, persistent, research-based interventions addressing the identified deficiency.
- The effect of curricular and instructional suggestions made to the classroom teacher in Tier 1 must be measured through relevant and frequent data collection.
- For children who fail to make adequate progress, a systematic and multidisciplinary problem-solving process is invoked in Tier 2.

• *Tier-two involves a multidisciplinary, problem-solving process for pupils who fail to make adequate progress following Tier 1 interventions.*

- The problem-solving process in Tier 2 must involve both general and special education teachers and support personnel, but still falls within the administrative and fiscal responsibility of general education.
- The Student Study Team (or other mechanism employed in Tier 2) has the following essential features: (a) a commitment to collaborative problem-solving regarding a student's difficulties within the general education classroom; (b) problem definition/functional analysis of the problem; (c) generation of empirically supported intervention strategies employing the resources available in the school, home and community to implement and sustain these interventions; and (d) systematic evaluation of intervention effects.
- When a student's rate of progress continues to be problematic, a referral for a comprehensive psycho-educational evaluation is appropriate.

• *Tier-three is used to define the special education referral process for pupils who continue to fail to make adequate progress.*

- The evaluation is for determining eligibility under IDEA as learning disabled, and utilizes all previously collected data concerning rate of progress, results of direct observation, and additional individual assessment.
- The Tier 3 assessment would be a comprehensive multidisciplinary assessment of the child's educational needs.
- The psycho-educational evaluation shall include an individual assessment based on current theory and research regarding learning disabilities.
- Cognitive assessment relative to the area of academic need is essential in the identification of learning disabilities, as well as other developmental or "high incidence" disabilities (e.g., ADHD).
- If the results of this comprehensive evaluation indicate that a student's instructional needs cannot be met solely in the general education program, an IEP Team meeting may be convened to determine appropriate supports and services in special education.

Note. The above table is from the CASP Board of Directors (2003) and is reprinted with permission.

CONCLUSION

RTI models have considerable promise for screening, intervention service delivery, and catalysts for system change. Research is needed to articulate purposes, operationalize procedures and judgments, and evaluate the decision-making utility of the models in practice. RTI represents a more flex-

ible service delivery model, but in order to ensure that children are reliably and effectively provided with assistance and achieve the objective of providing the "right assistance at the right time" (Pasternak, 2002), then it will be critical to articulate how RTI can be judged (which behaviors to measure, how frequently, for how long, under what stimulus conditions, and compared to what reference group using what units of measurement) and demonstrate that this judgment is functionally meaningful.

REFERENCES

Baker, S. K., Collins, V. L., & Gallagher, S. (1998). Curriculum-based measurement and its use in a problem-solving model with students from minority backgrounds. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement. The Guilford school practitioner series* (pp. 143-174). New York: Guilford.

Brigance, A. (1999). *Comprehensive Inventory of Basic Skills* (Rev. ed). North Billerica, MA: Curriculum Associates.

Case, L. P., Speece, D. L., & Molloy, D. E. (2003). The validity of a response-to-instruction paradigm to identify reading disabilities: A longitudinal analysis of individual differences and contextual factors. *School Psychology Review, 32*, 557-582.

CASP Board of Directors. (2003). *Critical constructs and principals regarding the reauthorization of IDEA: Position paper of the California Association of School Psychologists*. Available on-line at http://www.casponline.org/

Daly, E. J., Chafouleas, S. M., Persampieri, M., Bonfiglio, C. M., & Lafleur, K. (2004). Teaching phoneme segmenting and blending as critical early literacy skills: An experimental analysis of minimal textual repertoires. *Journal of Behavioral Education, 13*, 165-178.

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

Donovan, M. S., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, D.C.: National Academy Press.

Fletcher, J. M., & Reschly, D. J. (2005). Changing procedures for identifying children with learning disabilities: The danger of perpetuating old ideas. *The School Psychologist, 59*(1), 10-15.

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

Fuchs, L. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities Research and Practice, 18*, 172-186.

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

Fuchs, D., Fuchs, L. S., Mathes, P. G., Lipsey, M. W., & Roberts, P. H. (2001). Is "learning disabilities" just a fancy term for low achievement? A meta-analysis of reading difficulties between low achievers with and without the label. Paper written for the Office of Special Education Programs, U. S. Department of Education, and presented at the OSEP's LD Summit conference. Washington, D.C. Retrieved December 9, 2004, from http://www.air-dc.org/ldsummit/download/Fuchs%20Final%2008-10-01.pdf

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11*(4), 326-336.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

Gresham, F. M., Reschly, D. J., Tilly, W. D., Fletcher, J., Burns, M., Christ, T., Prasse, D., Vanderwood, M., & Shinn, M. (2005). Comprehensive evaluation of learning disabilities: A response to intervention perspective. *The School Psychologist, 59*(1), 26-29.

Gresham, F., VanDerHeyden, A. M., & Witt, J. C. (2005). IQ-achievement discrepancy in the identification of reading disabilities: Conceptual, measurement, and policy issues. Manuscript submitted for publication.

Gresham, F. M., & Witt, J. C. (1997). Utility of intelligence tests for treatment planning, classification, and placement decisions: Recent empirical findings and future directions. *School Psychology Quarterly, 12*, 249-267.

Hayes, S. C. Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42*, 963-974.

Holtzman, W. H., & Messick, S. (1982). *Placing children in special education: A strategy for equity.* Washington, D.C.: National Academy Press.

Hoover, H. D., Hieronymus, A. N., Frisbie, D. A., & Dunbar, S. B. (1993). *Iowa test of basic skills*: *Form M*. Riverside Publishing.

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

Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984-1024). New York: Wiley.

Kavale, K. A., Kaufman, A. S., Naglieri, J. A., & Hale, J. B. (2005). Changing procedures for identifying children with learning disabilities: The danger of poorly supported ideas. *The School Psychologist, 59*(1), 16-25.

Lennon, J. E., & Slesinski, C. (1999). Early intervention in reading: Results of a screening and intervention program for kindergarten students. *School Psychology Review, 28*, 353-364.

Macmann, G. M., & Barnett, D. W. (1999). Diagnostic decision making in school psychology: Understanding and coping with uncertainty. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 519-548). New York: Wiley.

Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of classification agreement. *The Journal of Special Education, 23*, 127-149.

MacMillan, D. L. (1998). Unpackaging special education categorical variables in the study and teaching of children with conduct problems. *Education and Treatment of Children, 21*(3), 234-245.

MacMillan, D. L., Gresham, F. M., & Bocian, K. M. (1998). Discrepancy between definitions of learning disabilities and school practices: An empirical investigation. *Journal of Learning Disabilities, 31*(4), 314-326.

MacMillan, D. L., & Speece, D. L. (1999). Utility of current diagnostic categories for research and practice. In R. Gallimore & L. P. Bernheimer (Eds.), *Developmental perspectives on children with high-incidence disabilities: The LEA series on special education and disability* (pp. 111-133). Mahwah, NJ: Lawrence Erlbaum.

Marston, D., Mirkin, P., & Deno, S. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education, 18*(2), 109-117.

Marston, D., Muyskens, P., Lau, M., & Canter, A. (2003). Problem-solving model for decision-making with high-incidence disabilities: The Minneapolis experience. *Learning Disabilities Research and Practice, 18*, 187-200.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Reschly, D. J., & Grimes, J. P. (1991). State department and university cooperation: Evaluation of continuing education in consultation and curriculum-based measurement. *School Psychology Quarterly, 20*, 519-526.

Shinn, M. R., Tindal, G. A., & Spira, D. A. (1987). Special education referrals as an index of teacher tolerance: Are teachers imperfect tests? *Exceptional Children, 54*, 32-40.

Shrank, F. A., Teglasi, H., Wolf, I. L., Miller, J. A., Caterino, L. C., & Reynolds, C. R. (2005). The American Academy of School Psychology reply to the response-to-intervention perspective. *The School Psychologist, 59*(1), 30-33.

Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research and Practice, 18*, 147-156.

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34*, 33-58.

VanDerHeyden, A. M., & Burns, M. K. (in press). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention*.

VanDerHeyden, A. M., & Witt, J. C. (2005). Quantifying the context of assessment: Capturing the effect of base rates on teacher referral and a problem-solving model of identification. *School Psychology Review, 34*, 161-183.

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32*, 204-227.

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

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

Vellutino, F. R., Scanlon, D. M., & Tanzman, M. S. (1998). The case for early intervention in diagnosing specific reading disability. *Journal of School Psychology, 36*, 367-397.

Witt, J. C., Daly, E., & Noell, G. (2000). *Functional assessments*. Sopris West: Longmont, CO.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson psychoeducational battery-revised*. Allen, TX: DLM Resources.