Running head: RESPONSE TO INTERVENTION AND LEARNING DISABILITIES

Response to Intervention in the Identification of Learning Disabilities: Empirical Support and Future Challenges

Frank M. Gresham

University of California-Riverside

Amanda VanDerHeyden

University of California-Santa Barbara

Joseph C. Witt

Louisiana State University

All correspondence should be directed to Frank M. Gresham, Graduate School of Education, University of California-Riverside Riverside, CA 92521 frank.gresham@ucr.edu

REVISED: January, 2005

Abstract

The recent reauthorization of the Individuals With Disabilities Education Act now allows for the use of response-to-intervention (RTI) as an alternative to discrepancy-based models for identification of learning disabilities (LD).This article presents an alternative approach for the identification of children with LD based on a RTI model. RTI can be delivered within a problem solving or a standard protocol approach using data-based decision making. In a RTI approach, children may be deemed eligible for special education if the amount of service exceeds to scope and resources available in general education. Four advantages of a RTI model are described: (a) early identification, (b) use of a risk model rather than a deficit model, (c) reduction of identification biases, and (d) strong focus on student outcomes. Methodological, technical, and logistical issues in RTI are discussed along with future directions for research and practice.

Response to Intervention in the Identification of Learning Disabilities: Empirical Support and Future Challenges

Historically, the field of learning disabilities could not be characterized as one that has been built on the strongest base of empirical support or unanimity of professional agreement regarding causes, identification practices, and remedial strategies. A central and unresolved issue facing the field pertains to policies and procedures that assist schools in determining which children need special education and related services, what types of services are needed, and with what intensity or strength should such services be delivered to effectively remediate academic difficulties. At the heart of these issues lies a fundamental but critical question: Which children have a learning disability and therefore are legally entitled to special education and related services and which children do not? That is, to what extent (if any) are low achieving children who are given the label of "learning disabled" different from low achieving children who are not given that label? Unfortunately, the process by which public schools qualify students as having a learning disability is often perceived as being unfair, logically inconsistent, and confusing (Bocian, Beebe-Frankenberger, MacMillan, & Gresham, 1999).

The purpose of this paper is to present an alternative approach in determining which children are and are not eligible for special education under the label of specific learning disabilities. This alternative approach is based on a response-to-intervention (RTI) model, which can be conducted within the context of either a problem solving, or standard protocol model using data based decision-making. In a RTI model, children are exposed to multi-tiered interventions in general education settings to determine *which* children need *what* services, delivered with *how much* intensity. If the amount of service exceeds the scope and resources of general education, then the child may be deemed eligible to receive additional resources through

special education. Methodological, technical, and logistical issues involved with implementing a RTI approach is presented along with future directions for research and practice.

IQ-Achievement Discrepancy: Origins and Validity Problems

The Isle of Wight studies can perhaps be viewed as the watershed event in canonizing the notion of IQ-achievement discrepancy in defining LD and solidifying its subsequent use in determining LD eligibility (Rutter, 1989; Rutter & Yule, 1975). These studies defined two types of underachievement in reading: general reading backwardness (GRB) and specific reading retardation (SRR). GRB was characterized as children having reading achievement scores that were consistent with their IQ (less than two standard errors of estimate predicted from their IQ score). SRR was characterized as deficient reading achievement that was equal or greater than two standard errors of estimate predicted from their IQ score. Thus, the GRB group was similar to what has been called "garden variety" low achievers (nondiscrepant low achievers) and the SRR group was similar to what we now call LD (discrepant low achievers).

Rutter and Yule (1975) examined the distribution of predicted achievement scores and found an overrepresentation of children with GRB in the lower tail of the distribution of predicted reading scores thereby representing a "hump." The Isle of Wight studies appeared to support the IQ-achievement discrepancy approach in defining reading disabilities and differentiating the SRR (discrepant low achievers) from GRB (nondiscrepant low achievers) group. More recent critical analyses have called into question the validity of Rutter and Yule's conclusions (see Fletcher, Francis, Shaywitz, Lyon, Foorman, Stuebing, et al., 1998).

The origins of IQ-achievement discrepancy in federal law can be traced back to 1976 when Congress ordered the Bureau of Education for the Handicapped (the forerunner of the Office of Special Education programs) to establish classification criteria for specific learning

disabilities (SLD) that would provide guidance to states and limit prevalence of the disability (1975 EHA statute cited in the Federal Register, November 29, 1976). What followed was an intense debate about how to best diagnose SLD. Cognitive and perceptual processing were rejected at that time because ample evidence at the time indicated that these measures were psychometrically weak, did not produce more accurate identification of SLD, did not provide valid and effective implications for instruction, and did not improve accuracy of predictions about outcomes (Torgesen, 2002). Thirty years after this debate, the field has still not seen the development of a substantial or convincing body of evidence showing that the use of cognitive and perceptual processing measures: (a) improves the accuracy of LD identification, (b) produces reasonable control over prevalence, (c) contributes to the design and implementation effective interventions, or (d) enhances predictions of important outcomes (Gresham 2002; Reschly & Ysseldyke, 2002; Silver, 2001;Torgesen, 2002; Vaughn & Fuchs, 2003).

SLD classification criteria had to be developed and published in the Federal Register by December 31, 1977 or a de facto prevalence cap of 2% automatically would have gone into effect. A controversial solution in the form requiring "…a severe discrepancy between achievement and intellectual ability" with the areas of achievement listed (e.g., reading recognition, reading comprehension, mathematical calculation) was published just before the deadline (Federal Register, December 29, 1977). This action produced, in effect, a contradiction between the statutory definition of LD (the law) that emphasized psychological processing and classification criteria (the regulations) that ignored psychological processing and emphasized general intellectual functioning and achievement.

The most fundamental assumption in the identification of LD is that IQ-achievement discrepancy is a valid marker for the presence of a specific learning disability. That is, students

displaying unexpected underachievement (discrepant underachievers) relative to their IQ are *different* on a variety of factors as compared to students not showing such discrepancies (nondiscrepant underachievers).

Several lines of converging research evidence strongly suggest that IQ-achievement discrepancy is not a valid marker for the presence of LD (Fletcher, Shaywitz, Shankweiler, Katz, Liberman, Steubing et al., 1994; Rogers, 1983; Share, McGee, McKenzie, Williams, & Silva, 1987; Stanovich & Siegel, 1994; Vellutino, Scanlon, & Lyon, 2000; Vellutino, Scanlon, Sipay, Small, Chen, & Denkla, 1996). These lines of research test directly the hypothesis that discrepant low achievers (LD) are different from nondiscrepant low achievers ("garden variety" low achievers). Overall, these studies showed that discrepant (LD) and nondiscrepant (LA) children do not differ on measures of reading achievement, cognitive abilities, phoneme awareness, short-term memory, visual analysis, and word retrieval.

Several meta-analyses have contrasted low achievers (LA) and LD groups on a variety of measures (Fuchs, Fuchs, Mathes, & Lipsey, 2002; Hoskyn & Swanson, 2000; Steubing et al., 2002). Hoskyn and Swanson (2000) reviewed 19 studies that contrasted LA (nondiscrepant) from LD (discrepant) groups. Based on an analysis of 274 weighted effect sizes, these authors found small differences between LA and LD groups on measures of cognitive skills related to reading such as pseudoword reading (g=.29), real-word phonetic analysis (g=-.02), speech-related phonological processing (g=.27), automaticity (g=.05), and spelling (g=.19). These authors concluded from their regression model analysis that both LA and LD groups share a general phonological processing deficit, a finding consistent with other findings in reading disability research (Perfetti, Beck, Bell, & Hughes, 1987; Stanovich & Siegel, 1994; Torgesen, 1999; Torgesen & Burgess, 1998; Torgesen, Burgess, & Rashotte, 1996).

More recently, Steubing et al. (2002) conducted a meta-analysis of 46 studies that investigated the validity of the IQ-achievement discrepancy for children with LD in the domains of achievement, behavior, and cognitive skills. These studies contrasted LD and LA children and included most of the 19 studies reviewed by Hoskyn and Swanson (2000) meta-analysis. Steubing et al. reported statistically insignificant effect sizes between LD and LA groups in the domains of achievement (*d*=-.12) and behavior (*d*=-.05). These authors reported a small difference (*d*=.30) in the cognitive domain, however, those measures most closely related to reading showed statistically insignificant differences between LD and LA groups (phonological awareness=-.13; memory=.10; rapid naming=-.12; vocabulary=.10). These authors concluded that defining LD on the basis of IQ-achievement discrepancy has little evidence for validity.

Summary. The IQ-achievement discrepancy approach to determining which students have and do not have LD has numerous conceptual and measurement problems that seriously call into question its continued use in making eligibility determinations. At the heart of the criticism of discrepancy is the fact that discrepant and nondiscrepant low achievers are not meaningfully different in domains of achievement, behavior, and processes related to reading. Moreover, discrepant and nondiscrepant low achievers do not differ in their response to instruction nor does the discrepancy approach inform instructional decisions (Fuchs & Fuchs, 1998; Gresham & Witt, 1997; Gresham, 2002; Vaughn & Fuchs, 2004; Vellutino et al., 2000). An alternative to the discrepancy approach is known as the *response to intervention* approach that is part of the reauthorization of the Individual with Disabilities Education Act (IDEA) recently passed by Congress and signed into law by the President. The advantages and potential of this approach to defining LD are discussed in the remainder of this article along with measurement issues, implementation challenges, and decision rules that must be resolved in the future.

Response to Intervention: An Alternative to IQ-Achievement Discrepancy

History of Response to Intervention

The basis of the response to intervention (RTI) approach can be traced back to the National Research Council (NRC) investigation (see Heller, Holtzman, & Messick, 1982) in which the validity of the special education classification system was evaluated on the basis of three criteria: (a) the quality of the general education program, (b) the value of the special education program in producing important outcomes for students, and (c) the accuracy and meaningfulness of the assessment process in the identification of a disability. Vaughn and Fuchs (2003) indicated that the first two criteria emphasize the quality of instruction whereas the third criterion involves judgments of the quality of instructional environments and the student's response to instruction delivered in those environments. The third criterion described in the NRC study is consistent with Messick's (1995) evidential and consequential bases for test use and interpretation. Speece (2002) described problems with IQ-achievement discrepancy in terms of unintended social consequences such as difficulty of young children to qualify under this criterion as well as overrepresentation of males and minority children using this approach. In addition to these difficulties, the discrepancy approach does not inform instructional decisions that might be made to improve student outcomes (Gresham, 2002). Heller et al. (1982) argued that only when all three criteria are met can a special education classification be considered *valid*.

Fuchs and Fuchs (1997, 1998) operationalized the NRC criteria by using a curriculum based measurement (CBM) approach that measures a student's responsiveness or unresponsiveness to intervention delivered in the general education classroom. In earlier work, Fuchs (1995) compared the RTI approach to the practice used in medicine whereby a child's

growth over time is compared to that of a same-age group. A child showing a large discrepancy between his or her height and that of a normative sample may be considered a candidate for certain types of medical intervention (e.g., growth hormonal therapy). In education, a child showing a discrepancy between the current level of academic performance and that of same-age peers in the same classroom might be considered a candidate for special education. It is important to note that a low-performing child who shows growth rates similar to that of peers in a low-performing classroom would not be considered a candidate for special education because the child is deriving similar educational benefits (although low) from that classroom (Fuchs, 1995).

Fuchs and Fuchs (1998) argued for a reconceptualization of the LD identification process based on a treatment validity notion. This approach does not classify students as LD unless and until it has been demonstrated empirically that they are not benefiting from the general education curriculum. Unlike traditional LD assessment that assesses students at one point in time using ability, achievement, and processing measures, the treatment validity approach repeatedly assesses the student's progress in the general education curriculum using CBM. Special education is considered only if a child's performance shows a *dual discrepancy* (DD) in which performance is below the level of classroom peers and the student's learning rate (growth) is substantially below that of classroom peers.

The CBM-DD model for determining LD eligibility determination consists of three phases. Phase I involves the documentation of adequate classroom instruction and dual discrepancies. It involves the collection of periodic CBM assessments for all students in a general education classroom. During this phase, overall classroom performance is compared to the performance relative to other classrooms or district norms. If classroom performance is

adequate, then individual student data are evaluated to determine the presence of a dual discrepancy based on: (a) a difference of 1 standard deviation between the student's CBM median score and that of classmates and (b) a difference of 1 standard deviation between a student's CBM slope of improvement (growth) and that of classmates. Students meeting these criteria and who do not have accompanying exclusionary conditions (e.g., mental retardation, sensory disabilities, autism), move on to Phase II of the process.

Phase II of this process involves implementation of a prereferral intervention focusing on remediating the student's dual discrepancy. CBM data are collected to judge the effectiveness of the intervention with the provision that the teacher implements a minimum of two interventions over a 6-week period. If a student does not show adequate progress (in terms of level and slope), the student enters Phase III of the process.

Phase III involves the design and implementation of an extended intervention plan. This phase represents a special-education diagnostic trial period in which the student's responsiveness to a more intense intervention is measured. This phase lasts about 8 weeks, after which a team reconvenes and makes decisions about the student's placement. The team could decide that the intervention was successful and an individual education plan (IEP) would be developed and the plan continued. Alternatively, the team could decide that the intervention was unsuccessful in eliminating the dual discrepancy and consider alternative decisions such as changing the nature and intensity of the intervention, collecting additional assessment information, considering a more restrictive placement, or changing to a school having additional resources that better address the student's needs.

In the above CBM-DD model, in order to qualify a student under LD and to provide special education services, a three-pronged test must be passed: (a) a dual discrepancy between

the student's performance level and growth (1 standard deviation for each) and of peers must be documented, (b) the student's rate of learning with adaptations made in the general education classroom is inadequate, and (c) the provision of special education must result in improved growth.

Speece and Case (2001) provided further validity evidence for the CBM-DD model in identifying students as LD. Children were identified as being at risk for reading failure if their mean performance on CBM reading probes placed them in the lowest quartile of their classes. A contrast group was identified that included five students from each classroom based on scores at the median (2 students) and the 30^{th}, 75^{th}, and 90^{th} percentiles (1 student at each level). At risk students were placed into one of three groups: CBM dual discrepancy (CBM-DD), regression-based IQ-reading achievement (IQ-DS), and low achievement (LA). Students in the CBM-DD group were given 10 CBM oral reading probes administered across the school year. Slopes (based on ordinary least squares regression for each child and classroom) were calculated, and each student's performance level was based on the mean of the last two data points. Children were placed in the CBM-DD group (n=47) if their slope across the year and level of performance at the end of the year was greater than 1 standard deviation below that of classmates. Students were placed in the IQ-DS group (n=17) if their IQ-reading achievement discrepancy was 1.5 or more standard errors of prediction (approximately a 20-point discrepancy). Children were placed in the LA group (n=28) if their total reading score was less than a standard score of 90.

Speece and Case showed that students in the CBM-DD group were more deficient on measures of phonological processing and were rated by teachers as having lower academic competence and social skills and more problem behaviors than students in the IQ-DS and LA groups. The CBM-DD and IQ-DS groups were not different on a standardized measure of

reading achievement demonstrating the specificity of the CBM-DD model (is this right, Frank?). These data offer further support for the CBM-DD model to identify students as LD, specifically those with phonological deficits. In later commenting on this study, Speece, Case, and Molloy (2003) stated:

> . . . by focusing on both level and growth in reading achievement as indexed by CBM, a valid group of children who experience reading problems was identified. Although much simpler identification methods would be preferred, other analyses indicated that single indicators of reading difficulty (LSF, ORF, phonological awareness) were not sensitive indicators of either DD or status as problem readers . . . The dual-discrepancy method would require major challenges in the way children are identified; however, our initial evidence suggests that benefits may outweigh the costs of change. (p. 150)

More recently, the RTI approach received further attention as a viable alternative to IQ-achievement discrepancy from the LD Initiative that was sponsored by the Office of Special Education Programs (U.S. Department of Education) that culminated in a national conference held in Washington, DC in August 2001 entitled the *LD Summit*. Nine white papers were written and presented over a two-day period. One paper (Gresham, 2002) specifically addressed the literature on *responsiveness to intervention* which was responded to by four professionals within the field of LD (Fuchs, 2002; Grimes, 2002; Vaughn, 2002; Vellutino, 2002). Subsequent to the LD Summit, the President's Commission on Excellence in Special Education (2002) emphasized RTI as a viable alternative to IQ-achievement discrepancy in the identification of LD.

President Bush signed into law the reauthorization of IDEA in November 2004. The law now reads with respect to specific learning disabilities:

> Specific learning disabilities: (A) General: Notwithstanding section 607 of this Act, or any other provision of law, when determining whether a child has a specific learning disability as defined under this Act, the LEA shall *not be required* to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability in oral expression, listening comprehension, reading recognition … (B) Additional Authority: In determining whether a child has a specific learning disability, a LEA *may use* a process which determines if a child *responds to a scientific, research based intervention.* (Emphases added)

Clearly, the reauthorized version of IDEA does not require nor does it eliminate IQ-achievement discrepancy as a basis for identifying children with LD. Moreover, it allows, but does not require, school districts to use a response-to-intervention approach to identify LD.

Advantages of RTI

There are at least four advantages of adopting a RTI approach in the identification of LD: Early identification of learning problems, (b) conceptualizing learning problem in terms of a risk model rather than a deficit model, (c) reduction of identification biases, and (d) focuses on student outcomes (Fletcher et al., 2002; Gresham, 2002; Vaughn & Fuchs, 2003). Each of these will be discussed briefly in the following sections.

Early identification of learning problems. Perhaps the most compelling reason for adopting a RTI approach is that it offers the opportunity of providing help to struggling children immediately. The current use of IQ-achievement discrepancy to identify LD has been termed a "wait-to-fail" approach because it requires that a student fail severely enough and long enough for the teacher to made a decision to refer. The developmental odds of being classified as LD increase dramatically from 1st to 4th grades. For example, between 1st and 2nd grades, the LD rate

doubles, from 2nd to 3rd grades, the rate doubles again, and between 3rd and 4th grades, the odds go up by a factor of 1.5 (U.S. Department of Education, 2002). Thus, between 1st and 4th grades, the odds of being labeled LD increase linearly by 450%. Discrepancy approaches penalize younger children because they are much less likely to demonstrate a discrepancy than older children (Fletcher, Francis, Shaywitz, Lyon, Foorman, Steubing et al., 1998).

The discrepancy approach delays the delivery of potentially effective, *early* interventions that could remediate learning difficulties. Teacher referral often results in unacceptability high rates of false positives (inaccurate identification of students as LD) and false negatives (failure to identify students who are LD) (MacMillan, Gresham, & Bocian, 1998; MacMillan & Siperstein, 2002; Vaughn & Fuchs, 2003). A RTI approach to the identification of learning difficulties can assist in closing the gap between identification and intervention (Vaughn & Fuchs, 2003). Students can be screened easily for early literacy problems and can be provided with effective early interventions to remediate academic difficulties (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Jenkins & O'Connor, 2002; O'Connor, 2000).

Risk versus deficit models. Historically, the field of LD has operated under a deficit model in which underlying cognitive and processing deficits are identified and specifically designed instructional strategies are recommended to remediate those deficits (Mann, 1979; Ysseldyke, 2001). Vaughn and Fuchs (2003) noted that the field of LD has simply not been successful at reliably identifying underlying processing deficits and linking that assessment information to effective instructional strategies. Current approaches to LD assessment rely heavily on aptitude by treatment interaction (ATI) logic in which instructional treatments are matched to aptitude strengths presumably to produce better outcomes. After 20 years of disappointing research, Cronbach (1975) abandoned ATI as a basis for applied psychology and

recommended a process akin to what is now called problem solving and short-run empiricism (see Reschly & Ysseldyke, 2002; Tilly, Reschly, & Grimes, 1999).

RTI operates under a risk model in which early identification of learning difficulties is emphasized. Under this model, *all* students are screened for potential learning and behavioral difficulties early in their school careers (e.g., kindergarten to 1st grade). Those students identified as being at risk are given supplemental instruction or behavioral support that has been shown to be an effective practice through research (i.e., evidence-based practice) to remediate these learning and behavioral difficulties.

The most important concept in any RTI model is the idea of matching the intensity of the intervention to the severity and resistance of the problem. This approach characterizes interventions that differ in terms of their nature, comprehensiveness, and intensity as well as the degree of unresponsiveness of behavior to those interventions (Gresham, 2004). The RTI approach offers better opportunities to integrate services between general and special education (Vaughn & Fuchs, 2003).

Reduction of identification biases. Special education eligibility in the public schools almost always begins with a general education teacher's decision to refer a child for special education consideration. The decision to refer a child for LD assessment is almost always based on academic difficulties, particularly in reading (Bocian, Beebe, MacMillan, & Gresham, 1999). The principle guiding teacher referral is one of *relativity*—that is, what is the child's academic performance relative to the modal performance of the class or the gap between the referred child's reading level and that of members of the lowest reading group.

A teacher's decision to refer is influenced not only by academic deficiencies, but other factors such as gender, socioeconomic status, or minority group membership (MacMillan &

Siperstein, 2002; Reschly, 2002; VanDerHeyden & Witt, in press). Donovan and Cross (2002) argued that a RTI approach to the referral process has the potential of reducing and perhaps eliminating the disproportionate overrepresentation of certain minority groups in special education that result from biases in the teacher referral process. Also, it is well established that there is a bias in overidentifying boys and underidentifying girls as LD by the current teacher referral process (Donovan & Cross, 2002; Shaywitz, Shaywitz, Fletcher, & Escobar, 1990). The power of iterative problem-solving efforts implemented within an RTI model of identification to reduce disproportionate identification by race and sex, and its superiority to other methods of identification such as teacher referral, has been empirically demonstrated (VanDerHeyden & Witt, in press).

Focus on student outcomes. RTI is based on the premise that measures and domains assessed should be determined by their relationships to child outcomes. Useful and appropriate measures and domains have a documented relationship to *positive* child outcomes, not just predictions of failure. Measures without such relationships do little for children and may cause harm because they deflect attention from measures and domains that can be used to produce positive outcomes (Reschly & Tilly, 1999). RTI emphasizes *direct measurement* of achievement, behavior, and the instructional environment as the core foci of a comprehensive evaluation of LD. RTI is concerned primarily with the assessment of measurable and changeable aspects of the instructional environment that are related to child outcomes. Assessment in a RTI model concentrates on factors that are related to achievement and interventions to improve rate and level of skill development.

The RTI approach is based on the assumption that a significant proportion of children who are or might be identified as LD may be more accurately characterized as “instructional

casualties" (Vaughn, Linan-Thompson, & Hickman, 2003). Clay (1987) argued that many children "learn to be learning disabled" because they are not exposed to early fundamental literacy skills in kindergarten and 1st grade (e.g., phoneme awareness, print concepts, letter-sound correspondence). In addition, many of these children are exposed to marginally effective general education reading curricula and instruction that have either not been scientifically validated or that have been implemented with poor integrity (National Reading Panel, 2000).

RTI involves analyses of prior and current instructional opportunities and the application of evidence-based instructional strategies related to positive child outcomes. Instructional variables assessed include alterable factors such as: time allocated for instruction, academic learning time, pacing of instruction, number of opportunities to respond, sequencing of examples and nonexamples of skills and so forth (Denton, Vaughn, & Fletcher, 2003; Carnine, 1997; National Reading Panel, 2000; Witt, VanDerHeyden, & Gilbertson, 2004). An essential component of a RTI model involves the direct measurement of treatment integrity of instructional interventions delivered in a general education classroom.

RTI Models

There are two basic approaches to delivering intervention services in a RTI model: (a) problem solving approaches and (b) standard protocol approaches (Fuchs, Mock, Morgan, & Young, 2003). These two approaches are described below. Emerging models that combine the two approaches described below may be particularly promising for use in school settings. (Barnett, Daly, Jones, & Lentz, 2004; Duhon, Noell, Witt, Freeland, Dufrene, & Gilbertson, 2004; Noell, Gansle, Witt, Whitmarsh, Freeland, LaFleur, Gilbertson, & Northrup, 1998; VanDerHeyden, Witt, & Naquin, 2003).

Problem Solving

Problem solving emanates from the behavioral consultation model first described by Bergan (1977) and later revised by Bergan and Kratochwill (1990). Behavioral consultation takes place in a sequence of four phases: (a) problem identification, (b) problem analysis, (c) plan implementation, and (d) plan evaluation. The goal in behavioral consultation is to define the referral problem in specific, operational terms, to identify the environmental conditions related to the referral problem, to design and implement an intervention plan with integrity, and to evaluate the effectiveness of the intervention (Bergan & Kratochwill, 1990).

Despite the popularity of problem solving approaches, Fuchs et al. (2003) suggested that practitioners and researchers have infrequently evaluated the outcomes of this approach to intervention and, more importantly: "The few that have done so generally failed to produce persuasive evidence that classroom-based interventions (1) are implemented with fidelity and (2) strengthen students' academic achievement or improve classroom behavior." (p. 163) Whereas it is true that there will need to be increased precision in the implementation of intervention procedures if they are to be used to make decisions about students, there is a considerable research base showing that interventions developed within a problem-solving model can be implemented with integrity and doing so improves students' academic performance.

Noell, Witt and their colleagues, borrowing from the performance management literature in business and industry (Balcazar, Hopkins, & Suarez, 1985), have demonstrated that teachers can reliably implement a variety of academic and behavioral interventions. Specifically, these researchers have shown that follow-up contact by a consultant that includes review of student progress and treatment implementation can lead to near perfect intervention implementation (see Mortenson & Witt, 1998; Noell, Duhon, et al., 2002; Noell, Witt, LaFleur, Mortenson, Ranier, &

LeVelle, 2000; Noell, Witt, Slider, Connell, Gatti, Williams, Koenig, & Resetar, in press; Witt, Noell, LaFleur, & Mortenson, 1997). The type of follow-up and support procedures in this research range from traditional performance management procedures such as daily performance feedback (Witt et al., 1997) or weekly performance feedback (Mortesen & Witt, 1998) to the use of procedures derived from the school-based consultation literature such as behavioral consultation or social influence (Noell, Witt, Slider, Connell, Gatti, Williams, Keonig, Resetar, & Duhon, in press).

The school-based performance management literature also serves to partially validate the claims of those such as Fuchs et al. (2003) with the consistent finding that a teacher who is asked to conduct an intervention without follow-up support typically results in poor intervention implementation. RTI is not a circumscribed event such as administering a test; rather, it is a multi-component process that, at the most basic level, involves someone implementing an intervention and then interpreting the results.

This above series of studies along with a huge literature in applied behavior analysis strongly suggests that behavioral interventions developed in a behavioral consultation model can (a) be implemented with integrity, but only with specific performance feedback from consultants and (b) if implemented correctly, produce substantial changes in student behavior (see Alberto & Troutman, 2003; Elliott, Witt, Kratochwill, & Stoiber, 2002; Kazdin, 2001).

A fundamental and unresolved issue is the extent to which typical practicing school psychologists can or will achieve similar levels of integrity with implementation and student outcomes using behavioral consultation methods. It is clear that RTI delivered within a problem-solving model will bring with it more responsibility on the part of professionals such as school psychologists to increase the precision and intensity of intervention efforts. There will also be a

need to move from the clear process of interpreting static test results to the ongoing process of interpreting time series intervention data (Barnett, Daly, Jones, & Lentz, 2004).

Standard Protocol Approaches

Another approach in a RTI model is the use of validated treatment protocols that can be implemented with students having academic difficulties in a given domain (e.g., reading). Many students classified as LD may fail to acquire basic academic skills not because of some underlying processing disorder, but because they have not been given adequate opportunities to learn (Gresham, 2002). The current use of IQ-achievement discrepancy and processing assessment does not screen out those children whose reading difficulties might be due to either inadequate schooling or limited exposure to effective reading instruction (Clay, 1987; Foorman et al., 1998; Vellutino, Scanlon, Sipay, Small, Pratt, Chen, & Denkla, 1996). Vellutino et al. (1996) suggested that exposure to validated reading instruction for a period of time should be used as a "first cut diagnostic aid" in distinguishing between reading problems caused by cognitive deficits versus those caused by experiential deficits (e.g., poor reading instruction).

Vellutino et al.'s (1996) research is an exemplary illustration of the standard protocol approach in distinguishing between cognitive and experiential deficits for children with reading difficulties. These authors conducted a longitudinal study with 183 kindergarten children composed of poor readers (n=118) and normal reader controls (n=65). Poor readers were selected on the basis of scoring below the 15th percentile on measures of word identification or letter-sound correspondence using nonsense words. Selected children in the poor reader group (n=74) were given daily one-to-one tutoring (30 minutes per day) for a total of 15 weeks over 70-80 sessions (35-40 hours of tutoring). Vellutino et al. calculated growth curves for each child from kindergarten to second grade. Slopes from these analyses were rank-ordered and used to

place children into 1 of 4 groups: Very Limited Growth, Limited Growth, Good Growth, and Very Good Growth. Approximately two-thirds of the sample showed either Very Good Growth or Good Growth indicating that they had caught up to their peers who did not have reading difficulties suggesting that these children were instructional casualties because they had not received sufficient reading instruction and/or other necessary preliteracy experiences.

Torgesen and colleagues described another standard protocol approach in which two carefully designed instructional approaches were compared for a sample of 8 to 10 year old children identified by their schools as LD (Torgesen, Alexander, Wagner, Rashotte, Voeller, & Conway, 2001). One intervention was the Auditory Discrimination in Depth (ADD) program that emphasized discriminations among phonemes, monitoring/representation of sound sequences in spoken syllables, and self-monitoring skills (Lindamood & Lindamood, 1998). The other intervention was the Embedded Phonics (EP) program that provided direct, explicit instruction in word-level reading skills as well as extensive opportunities to read and write meaningful text. Both intervention programs were provided to students on a 1:1 basis, in two 50-minute sessions per day, 5 days per week for 8 to 9 weeks. Students were assessed 1 to 2 years following completion of the intervention. Total time of intensive reading instruction for both groups totaled 67.5 hours. Following training, all students received 8 weeks of generalization training consisting of a single 50-minute session each week.

Both the ADD and EP interventions were equally effective in remediating students' reading difficulties based on the Woodcock-Johnson Broad Reading Cluster score. These interventions *normalized* the reading skills of approximately one-half to two-thirds of the students depending on the outcome measure used. About 40% of the LD students were returned to full-time general education and were no longer considered in need of special education.

Approximately 25% of students in this investigation were nonresponders to the intensive reading interventions with a mean standard score of 70 (2nd percentile) on Word Attack, Word Identification, and Passage Comprehension. These "treatment resisters", like the 25-30% of nonresponders in the Vellutino et al. (1996) study, may be the "real" LD students in need of special education and related services support that is not feasible within the context of a general education classroom.

Standard protocol approaches such as the ones described above have convincing empirical evidence that they can be used to effectively remediate reading difficulties in most (but not all) young poor readers. The primary advantage of the standard protocol approach compared to the problem-solving approach is that it may afford better quality control of instruction. Given that these protocols are scripted, these protocols can be used to ensure the integrity of instruction. Fuchs et al. (2003) aptly point out, however, that the standard protocol approach has been used almost exclusively by researchers and not by school practitioners. This research-to-practice gap represents a challenge in the area of RTI.

Technical Issues in Measuring RTI

The RTI approach to defining LD creates a number of technical measurement issues that differ substantially from the measurement issues involved in IQ-achievement discrepancy. The most fundamental issue in a RTI approach deals with the question of "adequate responsiveness." That is: How does one define an adequate response to intervention and how does one measure it? Two basic approaches have been proposed for indexing response to intervention: (a) final status and (b) growth models (see Fuchs, 2003). A variation of a growth model known as dual discrepancy (Fuchs & Fuchs, 1997, 1998; Speece & Case, 2001) was described earlier in this

paper. These approaches present their own measurement and technical problems that must be resolved by future research. These approaches are described in the following sections.

inal Status

One of the most straightforward ways of determining adequate response to intervention is to evaluate where the student is at the end of intervention. Students showing adequate functioning at posttest might be considered "treatment responders" and therefore not in need of special education services. What constitutes adequate functioning? There is no right or wrong answer, however, several guidelines might be suggested. For example, one might consider a student to have adequately responded to intervention if he or she is now functioning in the normal range on a norm-referenced measure of achievement (e.g., 25th percentile or higher). Another approach based on CBM might be whether or not the student meets an established benchmark criterion (i.e., level of performance that predicts functional competence in the skill area) for a particular skill at a given grade level (e.g., reading 40 words correctly per minute in 1st grade). There are well-established benchmark criteria for oral reading fluency using CBM that can be used in decision-making in a RTI model (see Fuchs, 2000; Shinn, 2002).

Torgesen et al. (2001) used final status in a sample of school-identified LD students to determine if students responded adequately to intensive one-to-one reading instruction (67.5 hours). These researchers showed that between one-half to two-thirds of students receiving the intensive reading intervention "normalized" their skills depending on the measure used. For example, students achieved scores in the normal range on the Word Attack (*M*=93.4) and Passage Comprehension (*M*=92.4) subtests of the Woodcock Reading Mastery Test-Revised (Woodcock, 1987). Additionally, students attained scores in the normal range on 5 of 6 phonological measures including: Phoneme Elision (*M*=99.5), Digit Memory (*M*=90), Nonword

Repetition (*M*=102), Rapid Automatized Naming Digits (*M*=90), and Rapid Automatized Letters (*M*=94.5). Interestingly, the reading interventions were most effective on measures of reading comprehension with 80 to 85% of students performing in the average range at the end of intervention.

The Torgesen et al. investigation provides useful information regarding how one might define *inadequate responders* based on the RTI concept. About 25% of the school-identified LD students (a sample likely to have a greater number of nonresponders relative to a general population) in this study were inadequate responders to the intensive intervention with mean standard scores of about 70 on Word Attack, Word Identification, and Passage Comprehension. Approximately 40% of these students who the schools previously identified as LD were returned to general education and deemed no longer in need of special education and related services.

The problem in using final status as the criterion in a RTI model is that it ignores the concept of growth that is a fundamental aspect of academic learning (Fuchs, 2003). For instance, a student can make very good growth as measured by slope estimates, but may not meet normative or even benchmark criteria as indexed by level estimates. Similarly, students can make relatively poor growth but may have started the intervention relatively close to the criterion standard. The concept of *growth* is an essential aspect of RTI and is considered in the following section.

Growth Models in RTI

All interventions seek to produce a discrepancy between baseline and post-intervention levels of performance and this notion forms the basis of a RTI approach (Gresham, 2002). The final status approach described above essentially uses this logic in which one compares pretest and posttest levels of performance. The effects of intervention, at least in group design studies,

are determined using some form of repeated measures analysis to compute simple mean differences on dependent variables for groups. Although these types of analyses can tell us whether or not an intervention produced mean differences for *groups*, these are insufficient data to model *individual change* adequately. What is needed in a RTI model is a way to model growth trajectories of students while they are being exposed to intervention.

Vellutino et al. (1996) used a growth analysis in a longitudinal study of 183 kindergarten children composed of poor readers (*n*=118) and normal reader controls (*n*=65). Poor readers were selected on the basis of scoring below the 15th percentile on measures of word identification or letter-sound correspondence using nonsense words. Children in the poor reader group were given 15 weeks of daily one-to-one tutoring (30 minutes per day) over 70 to 80 sessions. Using hierarchical linear modeling analyses, growth rates were calculated for each child from kindergarten to 2nd grade. Slopes from these analyses were rank-ordered and used to place children into 1 of 4 groups: Very Limited Growth, Limited Growth, Good Growth, and Very Good Growth. Approximately half the sample showed Very Limited Growth or Limited Growth (inadequate responders) and the other half of the sample showed either Good Growth or Very Good Growth (treatment responders).

Vellutino et al. (1996) also showed that IQ-achievement discrepancy scores did not reliably distinguish between difficult-to-remediate (Very Limited Growth or Limited Growth) and readily remediated (Very Good Growth or Good Growth) children and did not predict response to intervention. Vellutino et al. argued that this type of approach be used as a "first cut" diagnostic aid to determine which children's learning problems result from experiential deficits such as inadequate preliteracy experiences or inadequate instruction. It could be argued that half of the poor readers that showed Very Good Growth or Good Growth in this study might have

been inaccurately classified as LD (false positive errors) using IQ-achievement discrepancy because they would not have been exposed to effective, evidence-based reading instruction.

Figure 1 depicts three hypothetical growth curve models that might be expected in a RTI approach. The solid line represents the average growth over time one might expect from the classmates who exhibit no reading difficulty. For those beginning the year exhibiting a reading disability, one might hypothesize that some number, when provided with an evidence-based reading intervention implemented with integrity, will accelerate their progress and actually "catch up" (Hypothetical A) with those students showing no reading disability. These cases represent probable cases where instruction may have been inadequate and when taught well, they "catch up." Such cases might be viewed as *instructional casualties*, not "true" disabilities (Vaughn et al., 2003).

Another subgroup of children (Hypothetical B) begin the year well behind the nondisabled readers, but when provided with evidence-based reading instruction, they progress at the same rate as nondisabled readers, but fail to close the gap. That is, they show similar parallel slopes as the nondisabled readers (reflecting reading growth), but they never reach the desired grade-level reading performance.

The third group (Hypothetical C) resembles the "nonresponders" or "treatment resisters" in a number of reading research studies (e.g., Torgesen et al., 2001; Vellutino et al., 1996). They begin the year well behind the nondisabled readers, but despite exposure to evidence-based reading instruction delivered with integrity, they continue to fall further and further behind their nondisabled peers (i.e., they show both flat slopes and lower levels of reading performance). This research suggests that some 4 to 6% of students in a poor reading population are expected to

exhibit this pattern of inadequate response to instruction (Foorman, Francis, Fletcher, Schatschneider & Mehta, 1998; Torgesen et al., 2001; Vellutino et al., 1996, 2000).

Unresolved Issues in RTI

In terms of reading, we have a convincing body of evidence to suggest that many children with reading difficulties can be effectively remediated by intense exposure to evidence-based reading instruction. This evidence is based on research sponsored by the National Institute of Child Health and Human Development (NICHD) over the past 20 years that indicates reading difficulties are caused by weaknesses in the ability to process phonological aspects of language (Liberman, Shankweiler, & Liberman, 1989; Stanovich & Siegel, 1994; Torgesen, 1996; Vellutino, 1987; Vellutino & Scanlon, 2002). What the field does *not* have at this time is the availability of validated treatment protocols for other academic achievement areas such as mathematical calculation, mathematical reasoning, written expression, and so forth. This requires further research of the same quality as the aforementioned NICHD reading research. Other unresolved questions for RTI are as follows: (a) Is RTI a legitimate basis for determining LD? (b) What is a comprehensive assessment in a RTI model? and (c) How do we handle due process concerns in RTI? These are addressed briefly below.

RTI and Disability

Apart from the need for controlled outcome research in other academic areas, there is still the question of whether RTI is a legitimate basis for ruling in or ruling out the presence of a learning *disability*. Does the fact that a child responded adequately to an intervention rule out that he or she did not have a disability? Does this mean that the child's learning difficulties were caused exclusively by poor instruction? Does the fact that the child did not respond adequately to an intervention mean that he or she has a "true" learning disability?

These questions are perhaps of more interest to researchers than to treatment consumers such as parents or teachers. One might argue that the accuracy of a cancer diagnosis is not confirmed by how a patient responds to radiation and chemotherapy treatment regimens. That is, the diagnosis of cancer is made independently of treatment. While this is true, we do not enjoy the same luxury in defining a learning disability. Our field has always faced daunting conceptual and measurement difficulties in assessing processing or other cognitive abilities and relating these meaningfully to intervention (Bradley, Danielson, & Hallahan, 2002; Reschly & Ysseldyke, 2002; Torgesen, 2002). A legitimate argument from a RTI perspective is that if a child's learning difficulties have been remediated ("normalized"), then the issue of whether or not that child had a learning disability in the first place is rendered moot.

Comprehensive Assessment in RTI

What does a comprehensive assessment look like in a RTI approach to LD? An in-depth presentation of this topic would constitute an entire paper and therefore will not be comprehensively presented here. Briefly, RTI advocates argue that a comprehensive assessment is related to child outcomes and it gathers relevant functional information (see Gresham et al., 2004; Witt et al., 2004). Useful and appropriate measures and domains must have a documented relationship to *positive* child outcomes and not just predictions of failure. RTI uses *direct measurement* of achievement, behavior, and the instructional environment as the core foci of a comprehensive LD evaluation. The emphasis of assessment in a RTI approach is on assessment of measurable and changeable aspects of the instructional environment that are related to child outcomes in domains of achievement and behavior.

Comprehensive evaluation in a RTI approach necessarily focuses on teachable skills related to the curriculum that informs decision makers about what to teach and how to teach it.

Evaluation in a RTI model collects representative, direct, and low inference measures that concentrates on referral concerns and answers the assessment questions. Comprehensive assessment in a RTI approach also involves the direct measurement of the treatment integrity of instructional interventions delivered in either the general education classroom or in small group or individual instruction. A learning disability is conceptualized as: (a) low level of performance in a relevant domain in relation to peers, (b) slow growth rates compared to peers despite high quality, evidence-based interventions delivered with integrity, (c) documented adverse impact on educational performance, (d) documented need for special education, and (e) exit criteria defining goals for the special education program (Gresham et al., 2004). Child achievement and behavior outcomes in natural settings drive decisions at every step in the RTI comprehensive evaluation of LD.

Due Process

A final issue that is presently unclear is how due process would operate within a RTI approach. Would due process begin with problem-solving based interventions in general education? Should problem solving be ineffective, would due process concerns be addressed beginning with intensive academic interventions? Would due process commence only when it is clear that special education eligibility and placement is forthcoming? In some ways an RTI approach may introduce subjectivity because "professional" or "clinical" judgment may be needed more frequently with an RTI based system than under a traditional system. RTI calls for making a series of decisions pertaining to important issues such as whether an intervention is evidenced based, whether it is being implemented with integrity, whether it has continued long enough and ultimately whether it was effective. Reasonable people can disagree about any of these types of decisions. Further schools have never been overly conscientious about keeping

data on intervention integrity or even effectiveness. Within due process or civil proceedings the requirement to intervene in a "defensible" manner is an area of law that has yet to develop.

Vaughn and Fuchs (2003) point out that if due process is initiated later in the RTI process, students could be caught in a cycle where they are languishing in general education and are not receiving needed special education services. Alternatively, if due process is initiated too early, it could be expensive in terms of time and resources. At this point, it is unclear how due process should operate in a RTI model.

Conclusion

A RTI approach to determining eligibility for LD is based on a student's inadequate responsiveness to an evidence-based intervention rather than IQ-achievement discrepancy. The RTI approach appears to be a viable alternative to defining LD in light of the myriad of difficulties associated with discrepancy-based approaches for eligibility determination. Part of the appeal of the RTI approach as a decision-making tool is that it allows one to rule out inadequate instruction as a cause of insufficient academic achievement and allows for decision-making within an ecologically valid setting (e.g., the general education classroom).

A RTI approach may protect against faulty decision-making (Macmann & Barnett, 1999; MacMillan et al., 1998), unsubstantiated causal inferences (Neisworth & Bagnato, 1992), and use of assessment tools that do not inform instruction (Gresham & Witt, 1997; Macmann, Barnett, Lombard, Belton-Kocher, & Sharp, 1989). The RTI approach will require the direct measurement of behaviors necessary for successful performance using low inference assessment tools. Outcomes of interventions will be judged based on whether or not they produce acceptable levels of performance.

There is much more research needed to further operationalize the RTI approach to eligibility determination. This is particularly true for using RTI in other academic areas such as r mathematics and written expression. Some argue that RTI should not be embraced at this time because it has an insufficient research base to justify its use (Fuchs et al., 2003). Certainly those who work with and care about children seek to provide services and supports in the most effective manner possible. Efficient service delivery is also important so that resources can be allocated where they are most needed. Many of those who support the potential of RTI for eligibility determination are the very people who have criticized the use of the empirically unsupported practice of IQ-achievement discrepancy and processing deficits in LD classification for many years (Gresham et al., 2004; Reschly & Ysseldyke, 2002). RTI, although presently imperfect, offers the unique advantage of doing something about children's learning difficulties rather than simply "admiring" their problems based on highly inferential processing assessment procedures that do not inform instruction.

> The discussion of RTI is often contentious because it raises questions about very basic ideas in psychoeducational practice that the field has not resolved. Eligibility for specialized services lies at the vortex of many issues central to the field about how learning occurs and what limits there are to human potential for learning. School psychology originated because of a need to sort and serve students of a wider range of experience and ability due to federal mandate (first through compulsory schooling and second through IDEA). One could argue that the field has developed under contingencies arranged primarily through litigation and promoted by advocacy-based arguments rather than evidence-based arguments.

Public policy and law often precede the capabilities and motivation of schools to implement best practices. Functional behavioral assessment (FBA) provides a good example. The 1997 mandate under IDEA for FBA was a timely and appropriate response to converging evidence supporting the power of this technology to produce greater freedom and habilitative experience among individuals displaying problem behaviors that competed with more appropriate, adaptive behaviors in school settings. Yet, the application of FBA has been influenced greatly by such "local" factors as resource availability, theoretical orientation, and practitioner skills such that the *actual* application has born little resemblance to the *prescribed* application (see Gresham, Watson, & Skinner, 2001). Probably because of this weak translation, FBA has not produced the effects of which it is capable in many if not most educational settings and, like many previous mandated policies, it has become less relevant to practice and policy.

Current evidence suggests that RTI can be implemented responsibly while the evidence base accumulates and that iterations and modifications are not only inevitable, but also desirable. The benefits of RTI far outweigh potential costs for children and will only facilitate refinements toward a model supported by converging sources of evidence. The changes and challenges presented by a RTI model necessarily will force school psychologists to "think outside" of their test kits and into an intervention-based practice of school psychology.

References

Alberto, P.A., & Troutman, A.C. (2003). *Applied behavior analysis for teachers* (6th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Balcazar, F., Hopkins, B.L., & Suarez, Y. (1986). A critical, objective review of performance feedback. *Journal of Organizational Behavior Management, 7*, 65-89.

Barnett, D.W., Daly, E.J., Jones, K.M., & Lentz, F.E. (2004). Response to intervention: Empirically-based special service decisions for increasing and decreasing intensity using single case designs. *The Journal of Special Education, 38,* 66-79.

Bergan, J., & Kratochwill, T.R. (1990). *Behavioral consultation and therapy.* New York: Plenum Press.

Bocian, K., Beebe, M., MacMillan, D., & Gresham, F.M. (1999). Competing paradigms in learning disabilities classification by schools and variations in the meaning of discrepant achievement. *Learning Disabilities Research & Practice, 14,* 1-14.

Brooks, P., & Baumeister, A. (1977). A plea for consideration of ecological validity in the experimental analysis of psychology of mental retardation: A guest editorial. *American Journal of Mental Deficiency, 81*, 407-416.

Carnine, D., Silbert, J., & Kame'enui, E. (1997). *Direct instruction in reading* (3rd ed.). New York: Merrill.

Clay, M. (1985). *The early detection of reading difficulties* (3rd ed.). Auckland, Nezealand: Heinemann.

Clay, M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies, 22*, 155-173.

Cronbach, L. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist, 30,* 116-127.

Denton, C.A., & Vaughn, S. (2003). Bringing research-based practice in reading intervention to scale. *Learning Disabilities Research & Practice, 18,* 201-211.

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

Duhon, G.J., Noell, G.H., Witt, J.C., Freeland, J.T., Dufrene, B.A., & Gilbertson, D.N. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33,* 429-443.

Elliott, S.N., Witt, J.C., Kratochwill, T.R., & Stoiber, K.C. (2002). Selecting and evaluating classroom interventions. In M. Shinn, H. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 243-294). Bethesda, MD: National Association of School Psychologists.

Fletcher, J., Francis, D., Shaywitz, S., Lyon, G.R., Foorman, B., Stuebing, K., et al., (1998). Intelligent testing and the discrepancy model for children with learning disabilities. *Learning Disabilities Research & Practice, 13,* 186-203.

Fletcher, J., Shaywitz, S., Shankweiler, D., Katz, L., Liberman, I., Steubing, K., Francis, D., Fowler, A., & Shaywitz, B. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86,* 6-23.

Foorman, B., Francis, D., Fletcher, J., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90,* 37-55.

Francis, D., Shaywitz, S., Steubing, K., & Shaywitz, B. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology, 88,* 3-17.

Fuchs, L.S. (1995). Incorporating curriculum-based measurement into the eligibility decision-making process: A focus on treatment validity and student growth. Paper presented for the National Academy of Sciences Workshop on Alternatives to IQ Testing. Washington, DC.

Fuchs, L.S. (2002). Three conceptualizations of "treatment" in a response-to-treatment framework for LD identification. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 521-529). Mahwah, NJ: Lawrence Erlbaum.

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 & Practice, 18,* 157-171.

Fuchs, L., & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children, 30,* 1-16.

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

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

Gresham, F.M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 467-519). Mahwah, NJ: Lawrence Erlbaum.

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.

Grimes, J. (2002). Responsiveness to interventions: The next step in special education identification, service, and exiting decision making. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 531-547). Mahwah, NJ: Lawrence Erlbaum.

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

Hoskyn, M., & Swanson, H.L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *School Psychology Review, 29,* 102-119.

Individuals with Disabilities Education Act Amendments (IDEA) *of 1997*. (1997). Public Law 105-17.

intervention. *School Psychology Review, 18,* 37-50.

Jenkins, J.R., & O'Connor, R.E. (2002). Early identification and intervention for young children with reading/learning disabilities. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 99-149). Mahwah, NJ: Lawrence Erlbaum.

Liberman, I., Shankweiler, D., & Liberman, A. (1989). The alphabetic principle and learning to read. In D. Shankweiler & I. Liberman (Eds.), *Phonology and reading disability: Solving the reading puzzle* (pp. 1-33). Ann Arbor, MI: University of Michigan Press.

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

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

MacMillan, D., & Siperstein, G. (2002). Learning disabilities as operationally defined by schools. In R. Bradley, L. Danielson, & D. Hallahan (Eds.). *Identification of learning disabilities: Research to practice* (pp. 287-333. Mahwah, NJ: Lawrence Erlbaum.

MacMillan, D.L., & Siperstein, G.N. (2002). Learning disabilities as operationally defined by schools. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 287-333). Mahwah, NJ: Lawrence Erlbaum.

Mann, L. (1979). *On the trail of process*. New York: Grune & Stratton.

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

Mortenson, B.P., & Witt, J.C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology Review, 27,* 613-627.

National Reading Panel (2000). *Report of the national reading panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Neisworth, J., & Bagnato, S. (1992). The case against intelligence testing in early intervention. *Topics in Early Childhood Special Education, 12,* 1-20.

Noell, G.H., Duhon, G.J., Gatti, S.L., & Connell, J.E. (2002). Consultation, follow-up, and behavior management intervention implementation in general education. *School Psychology Review, 31,*217-234.

Noell, G.H.,Gansle, K.A., Witt, J.C., Wthimarsh, E.L., Freeland, J.T., LaFoeur, L.H., Gilbertson, D.N., & Northrup, J. (1998). Effects of contingent reward an instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31,* 659-663.

Noell, G.H., Witt, J.C., Gilbertson, D.N., Ranier, D.D., & Freeland, J.T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, *12,* 77-88.

Noell, G.H., Witt, J.C., LaFleur, L.H., Mortenson, B.P., Ranier, D.D., & LeVelle, J. (2000). A comparison of two follow-up strategies to increase teacher intervention implementation in general education following consultation. *Journal of Applied Behavior Analysis, 33,* 271-284.

Noell,G.H., Witt, J.C., Slider, N.J., Connell, F.E., Gatti, S.L., Williams, K.L., Koenig, J.L., & Resetar, J.L. (in press). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review.*

O'Connor, R.E. (2000). Increasing the intensity of intervention in kindergarten and first grade. *Learning Disabilities Research & Practice, 15,* 43-54.

Perfetti, C., Beck, L., Bell, L., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly, 33,* 283-319.

President's Commission on Excellence in Special Education (2002). *A new era: Revitalizing special education for children and their families.* Washington, DC: United States Department of Education, Author.

Reschly, D.J., & Tilly, W.D. (1999). Reform trends and system design alternatives. In D. Reschly, W.D. Tilly, & J. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 19-48). Longmont, CO: Sopris West.

Reschly, D.J., & Tilly, W.D. (1999). Reform trends and system design alternatives. In D. Reschly, W.D. Tilly, & J. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 19-48). Longmont, CO: Sopris West.

Reschly, D.J., & Ysseldyke, J. (2002). Paradigm shift: The past is not the future. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV.* Bethesda, MD: National Association of School Psychologists.

Reschly, D.J., Tilly, W.D., & Grimes, J. (Eds.)(1999). *Special education in transition: Functional assessment and noncategorical programming.* Longmont, CO: Sopris West.

Reynolds, C. (1984-85). Critical measurement issues in learning disabilities. *Journal of Special Education, 18,* 451-476.

Rogers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology, 53,* 369-373.

Rutter, M. (1989). Isle of Wight revisited: Twenty-five years of child psychiatric epidemiology. *Journal of the American Academy of Child and Adolescent Psychiatry, 29,* 633-653.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16,* 181-197.

Share, D., McGee, R., McKenzie, D., Williams, S., & Silva, P. (1987). Further evidence relating to the distinction between specific reading retardation and general reading backwardness. *British Journal of Developmental Psychology, 5,* 35-44.

Shaywitz, B., Fletcher, J., Holahan, J., Shaywitz, S. (1992). Discrepancy compared to low achievement definitions of reading disability: Results from the Connecticut Longitudinal Study. *Journal of the American Medical Association, 264,* 998-1002.

Siegel, L. (1988). Evidence that IQ scores are irrelevant to the definition and analysis of reading disability. *Canadian Journal of Psychology, 42,* 201-215.

Siegel, L. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22,* 469-479.Silver, L.B. (2001). Controversial therapies. *Perspectives, 27,* 5-8.

Speece, D.L., & Case, L.P. (2001). Classification in context: An alternative approach to identifying early reading disability. *Journal of Educational Psychology, 93,* 735-749.

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

Stanovich, K., & Siegel, L. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core-variable-difference model. *Journal of Educational Psychology, 86,* 24-53.

Stuebing, K., Fletcher, J., LeDoux, J., Lyon, G.R., Shaywitz, S., & Shaywitz, B. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39,* 469-518.

Swanson, H.L., & Hoskyn, M. (1999). Definition X treatment interaction for students with leaning disabilities. *School Psychology Review, 28,* 644-658.

Swanson, H.L., & Saches-Lee, C. (2000). A meta-analysis of single-subject intervention research for students with LD. *Journal of Learning Disabilities, 33,* 114-136.

Swanson, H.L., Carlson, C., & Saches-Lee, C. (1996). A selective synthesis of intervention research for students with learning disabilities. *School Psychology Review, 25,* 370-391.

Tilly, W.D., Reschly, D.J., & Grimes, J. (1999). Disability determination in problem solving systems: Conceptual foundations and critical components. In D. Reschly, W.D. Tilly, & Grimes (Eds.), *Special education in transition: Functional and noncategorical programming* (pp. 285-301). Longmont, CO: Sopris West.

Torgesen, J. (1999). Phonologically-based reading disabilities: Toward a coherent theory of one kind of learning disability. In R.J. Sternberg & L. Spear-Swerling (Eds.), *Perspectives on learning disabilities* (pp. 231-262). New Haven, CT: Westview Press.

Torgesen, J. (2002). Empirical and theoretical support for direct diagnosis of learning disabilities by assessment of intrinsic processing weaknesses. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 565-613). Mahwah, NJ: Lawrence Erlbaum.

Torgesen, J., & Burgess, S. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional

studies. In J. Metsala & L. Ehri (Eds.), *Word recognition in beginning reading* (pp. 161-188). Hillsdale, NJ: Lawrence Erlbaum.

Torgesen, J., Burgess, S., & Rashotte, C. (1996, April). *Predicting phonologically-based reading disabilities: What is gained by waiting a year?* Paper presented at the Scientific Study of Reading, New York.

Torgesen, J., Rose, E., Lindamood, P., Conway, T., & Garvan, C. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579-594.

U.S. Department of Education (1977). Definition and criteria for defining students as learning disabled. *Federal Register, 42,* 250.

United States Department of Education (2002). *Twenty-fourth annual report to Congress on implementation of the Individuals with Disabilities Education Act.* Washington, DC: Author.

Upah, K., & Tilly, W.D. (2002). Best practices in designing, implementing, and evaluating quality interventions. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-IV* (pp. 483-516). Bethesda, MD: National Association of School Psychologists.

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

Vaughn, S. (2002). Using response to treatment for identifying students with learning disabilities. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 549-554). Mahwah, NJ: Lawrence Erlbaum.

Vaughn, S., & Fuchs, L.S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research & Practice, 18,* 137-146.

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. (1979). *Dyslexia: Theory and research.* Cambridge: MIT Press.

Vellutino, F.R. (1987). Dyslexia. *Scientific American, 256,* 34-41.

Vellutino, F.R. (2002). On the role of intervention in identifying learning disabilities. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 555-564). Mahwah, NJ: Lawrence Erlbaum.

Vellutino, F.R., & Scanlon, D. (2002). The Interactive Strategies approach to reading intervention. *Contemporary Educational Psychology, 27,* 573-635.

Vellutino, F.R., Scanlon, D., & Jaccard, J. (2003). Toward distinguishing between cognitive and experiential deficits as primary sources of difficulty in learning to read: A two-year follow-up of difficult to remediate and readily remediated poor readers. In B. R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 73-120). Baltimore: York Press.

Vellutino, F.R., Scanlon, D., & Lyon, G.R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against IQ-achievement discrepancy of reading disability. *Journal of Learning Disabilities, 33,* 223-238.

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

Vellutino, F.R., Scanlon, D., Sipay, E., Small, S., Pratt, A., Chen, R., & Denkla, M. (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.

Witt, J.C., VanDerHeyden, A.M., & Gilbertson, D.M. (2004). Troubleshooting behavioral interventions: A systematic process for finding and eliminating problems. *School Psychology Review, 33,* 363-383.