Running head: DISCREPANCY AND LEARNING DISABILITIES

IQ-Achievement Discrepancy in the Identification of Reading Disabilities: Conceptual, Measurement, and Policy Issues

Frank M. Gresham

University of California-Riverside

Amanda M. VanDerHeyden

Vail School District

Joseph C. Witt

Louisiana State University

Please direct correspondence to: Amanda VanDerHeyden, Vail School District, 13801 E. Benson Highway, Suite B Vail, AZ 85641

Abstract

IQ-Achievement Discrepancy in the Identification of Reading Disabilities: Conceptual, Measurement, and Policy Issues

Historically, the field of learning disabilities could not be characterized as one that has been built on the strongest base of empirical support and unanimity of professional agreement. 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 such services should be delivered to effectively remediate educational difficulties. At the heart of these issues lies a fundamental, but critical question: Which children are learning disabled and therefore are entitled to special education and related services and which children are not? Unfortunately, the process by which public schools qualify students as learning disabled is often unfair, logically inconsistent, and confusing (Bocian, Beebe, MacMillan, & Gresham, 1999). This state of affairs lead G. Reid Lyon of the National Institute of Child Health and Human Development (NICHD) to characterize the field of learning disabilities as a sociological sponge which functions to mop up the spills of general education (Lyon, 1999).

The purpose of the present paper is to analyze the current use of IQ-achievement discrepancy as a marker for the presence of reading disabilities. We focus on reading disabilities because it is by far the most common reason children qualify for the label of specific learning disabilities (SLD) in the schools (Lyon, 1985; Siegel, 1989; Vellutino, Scanlon, & Lyon, 2000). We provide a brief historical background of the LD field and review traditional school practices in LD identification. Also, we review and critique current assessment practices in identifying children with reading disabled (RD) based on IQ-achievement discrepancy as well as the largely

mythical aptitude X treatment interactions that emanate from cognitive ability testing for children with RD.

We conclude by arguing for a radically different 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 in which 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 needed exceeds the scope and resources of general education, then the child is made eligible to receive additional resources through special education.

Historical Background of LD

The term learning disabilities was first introduced by Kirk (1962) to characterize a group of children who have delayed development, retardation, or disorder in the areas of speech, language, reading, writing, arithmetic, or other school subjects. In this definition, the notion of psychological processing disorders was introduced and how these processing disorders were thought to lead to poor academic achievement (Kavale & Forness, 2000). Kirk's (1962) definition also referred to what are now know as the "exclusionary criteria" in which achievement problems are not due to mental retardation, sensory deprivation, or cultural or instructional factors.

Kavale (2002) suggested that the notion of discrepancy in identifying learning difficulties was introduced by Franzen (1920) with his use of the Accomplishment Quotient or AQ which was computed by computing the ratio of educational quotient (grade level performance) to intelligence quotient. An AQ of less than one indicates that the child is demonstrating unexpected underachievement not unlike the IQ-achievement discrepancy used today. Some 12

years later, Marion Monroe (1932) described another discrepancy model in her book entitled *Children Who Cannot Read* (see Hallahan & Mercer, 2002). To calculate a discrepancy, Monroe (1932) computed a reading index based on an average of four tests measuring reading comprehension, word recognition, word analysis, and word discrimination relative to the average scores based on chronological age, mental age, and arithmetic grade equivalent. Hallahan and Mercer (2002) provided the following example of a boy who is chronologically at grade 3.6 with a mental age of 4.0 and an arithmetic grade equivalent of 3.5. This averages to 3.7. If the child's four reading scores average 2.0 then his reading achievement would be 2.0/3.7 or 54% of expected achievement. Monroe argued that a criterion of 80% or less of expected achievement be used to define reading disabilities and suggested that about 12% prevalence rate of reading disabilities (Hallahan & Mercer, 2002).

Over 30 years later, Barbara Bateman (1965) reintroduced the idea of discrepancy between expected and actual academic performance to define LD. In Bateman's terminology, learning disorders are exhibited by children who show an "educationally significant discrepancy" between estimated potential and actual level of performance. Unlike Monroe (1932), Bateman's (1965) definition provided no guidelines or suggestions for what constituted an "educationally significant discrepancy" and did not suggest ways in which intellectual potential and actual level of performance should be operationalized (Kavale & Forness, 2000). Today, over 30 years after the Bateman (1965) definition, the field of LD has not arrived at a consensus regarding the definitional and measurement issues involved in operationalizing an "educationally significant discrepancy."

The Isle of Wight studies are viewed by many as a watershed event in canonizing the notion of IQ-achievement discrepancy in defining RD (Rutter, 1989; Rutter & Yule, 1975).

These studies defined two types of reading underachievement: general reading backwardness (GRB) and specific reading retardation (SRR). Using Performance IQ of the Wechsler Intelligence Scale for Children and a variety of reading measures, Rutter and Yule (1975) defined the GRB and SRR groups using regression-based approaches not unlike current standard score regression approaches used to define LD. The GRB group was defined as having deficient reading scores that was less than two standard errors of estimate predicted from their Performance IQ. The SRR group also was defined as having deficient reading achievement, which was equal to or greater than two standard errors of estimate predicted from their Performance IQ. 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 SLD (discrepant low achievers).

The Isle of Wight studies appeared to support the IQ-achievement discrepancy approach in defining RD and differentiating them from low achieving groups. However, more recent critical analyses have called into question the validity of this conclusion. Fletcher et al. (1998) argued that over one-third of the GRR group had known or suspected evidence of neurological impairment and had IQ scores in the range of mental retardation. Thus, the separation of the GRB and SRR groups and the larger percentage of lower achievement scores in the GRB group in large measure could be attributed to the neurological involvement and lower IQ scores of the GRB group.

School Identification of LD

Of the 13 categories specified in the Individual With Disabilities Education Act (IDEA, 1997), the LD category accounts for over half (52%) of all children served in special education under this act. The number of students served as LD increased almost 300% from 1976-77 to

2001-2002. During this same time frame, the number of students served under the category of mental retardation decreased by about 60% (U. S. Department of Education, 2002). MacMillan and colleagues in commenting on these statistics indicated: "Were these epidemic-like figures interpreted by the Center (sic) for Disease control, one might reasonably expect to find a quarantine imposed on the public schools of America (MacMillan, Gresham, Siperstein, & Bocian, 1996, p. 169).

What explains this huge increase in LD and concomitant decrease in mental retardation over the past 25 years? Although there is no well-established or universally accepted explanation for these data, the increase in LD can be partly explained by school practices of classifying LD on the basis of absolute low achievement that ignores IQ level or IQ-achievement discrepancy (Gresham, 2002; MacMillan, Gresham, & Bocian, 1998). Analyses of current classification practices suggest that: (a) a relatively small number of children are classified as mildly mentally retarded, (b) a large number of children are served (erroneously) in special education as LD, and (c) some unknown number of children avoid detection, and are not referred by teachers despite their concerns about the child's academic performance (Gresham, 2002; MacMillan, Gresham, Bocian, & Siperstein, 1997; MacMillan, Siperstein, & Gresham, 1996).

Competing Paradigms In LD Identification and Placement

One way of understanding the decision rules that schools used to classify children as LD is through a competing paradigms model consisting of three steps or gates: (a) referral by a general education teacher, (b) psychoeducational assessment by a school psychologist, and (c) a team placement decision based on all available evidence regarding the child (Bocian et al., 1999). It is important to note that sources of information and judgments are differentially weighted at each of the three steps of referral, evaluation, and placement.

At the referral step, general education teachers largely use local or classroom norms to judge whether or not a child deviates significantly from his or her peers in academic achievement. The fundamental principle at this step is one of *relativity* in which the teacher's decision to refer is based on a comparison of the child's academic performance relative to the modal academic performance of the class. This is termed in the literature "teachers as imperfect tests" (Gerber & Semmel, 1984; Gresham, MacMillan, & Bocian, 1997; Gresham, Reschly, & Carey, 1987). If, in the teacher's judgment, this gap between the child's and the class's academic performance cannot be closed using the extant resources in the classroom, the referral decision is highly likely (Bocian et al., 1999).

Children referred by teachers a likely to be subjected to a psychoeducational evaluation if prereferral intervention efforts have not been successful in closing the gap between the child's achievement relative to that of the class and/or the teacher's expectation. This second gate is based primarily on national norms and compares the child's performance in the areas of intelligence, achievement, and processing to that of nationally representative standardization samples. This step is driven by the notion of *acceptability* in that the decision to pass the student through this gate is based on how acceptable or unacceptable the child's level of academic performance is relative to his or her ability (Bocian et al., 1999). A child showing a significant discrepancy between ability and achievement (unexpected underachievement) is likely to be passed onto the third gate of placement.

The final step in this process is that of placement based on a decision of a multidisciplinary team (MDT). Unlike referral, which is based on the individual judgment of the referring teacher, the decision of the MDT is a team decision or judgment. The key concept guiding the MDT decision to place is that of *profitability* that consists of the team's collective

judgment that the child will profit or otherwise benefit for special education services offered at that school site (Bocian et al., 1998).

MacMillan et al. (1998) indicated that the different MDTs differentially weight local norms, national norms, and sociocultural/contextual factors in making a placement decision. As such, MDT decisions will vary in the face of identical information due to the forcefulness or convincingness of different MDT players. This differential weighting of information from different sources at least in part explains the huge variability in LD placement rates across states, districts, and even schools within districts.

Empirical Data on School Identification of LD

There are empirical data that partially explain the large variability in school practices in LD identification. MacMillan and colleagues showed that approximately half of students that general education teachers referred had IQ scores between 71 and 85 with another 16% scoring below an IQ of 70 (MacMillan, Gresham, Bocian, & Lambros, 1998). MacMillan et al. used the American Association on Mental Retardation (AAMR, 1992) criterion of IQ<75 and found that 30% of the referred sample scored below that level.

Into what special education categories are the above children placed? Whereas there are a relatively large number of children that might be psychometrically eligible for placement as mildly mentally retarded (MMR), only 14% of 43 children with IQ<75 were classified under this label by schools (MacMillan, Gresham, Siperstein, & Bocian, 1996). Interestingly, a little less than half (44%) of these MMR cases were labeled as LD by schools despite showing IQ in the ranges of mental retardation (an exclusionary criterion) and not demonstrating a discrepancy between ability and achievement (a regulatory requirement). One might consider schools'

mislabeling of MMR cases as LD to be false positive LD identifications that is inconsistent with state and federal regulations of special education identification.

In commenting on the above situation, MacMillan et al. (1998) suggested: " . . . public school practices for diagnosing children with LD bear little resemblance to what is prescribed in federal and state regulations . . . defining LD."(p. 323). MacMillan and Siperstein (2002) argued that identification models extant in state and federal regulations are "measurement bound" in that they specify cutoff scores, required discrepancies, and other psychometric profiles that should be applied objectively in determining eligibility. This appearance of "objectivity" in the assessment of LD, however, contains a huge amount of subjectivity and judgment that further obscures what we know and do not know about LD (MacMillan & Siperstein, 2002).

LD as a Relative Category

Why has the category of LD been so problematic in terms of who is and is not eligible for services under this label? As we explained above, part of the reason is rooted in the competing paradigms model in which information is based on different criteria (local norms versus national norms) and is differentially weighted in a team decision regarding placement. We believe that a more fundamental reason for the difficulties in consistently applying the LD label rests with the fact that it is a category reflecting differences in *degree* rather than differences in *kind* of low achievement.

Table 1 depicts several categories of child disabilities showing differences in degree and kind. For example, take the category of mental retardation that historically had four levels (mild, moderate, severe, and profound). The average IQ scores of children with mild mental retardation are around 55-70 and the average IQ scores of children with profound mental retardation are around 25 or below. Few, if any, people would suggest that these two groups of children do not

differ on a number of variables such as identification prior to or after school entry, whether on not the child has severe deficits in independent functioning, and presence or absence of comorbid biomedical conditions (MacMillan, Gresham, & Siperstein, 1993). It is quite clear that children with mild mental retardation and profound mental retardation differ in *kind*. The same can be said for hard of hearing children versus profoundly deaf children and visually impaired children versus blind children.

Where the field of disabilities has much greater difficulty in diagnosing the so-called mild or high-incidence disabilities. For example, at what point or level do attention problems and overactivity become an Attention Deficit-Hyperactivity Disorder (ADHD) or when do emotional and behavioral difficulties become Emotional Disturbance? Most germane to this current paper, at what point can one differentiate low academic achievement from LD? Are there particular marker variables that reliably distinguish low achievers from children with LD? It is clear in all of the above cases that we are talking about groups that differ in *degree* rather than *kind*. As we shall see in the following section, research contrasting low achieving groups from LD groups has not been able to make this differentiation reliably.

The issue of distinguishing low achievement from LD creates even greater consternation for child study teams charged with the task of making eligibility determination decisions for children referred to them for assessment and classification. As discussed earlier, these child study (multidisciplinary) teams are permitted to exercise judgment in making an eligibility determination. This judgment, however, is a *team judgment* that takes into consideration local/classroom norms (upon which the referral is based), national norms (upon which the testing information is based) and sociocultural factors (e.g., ethnicity, gender, socioeconomic status) (Bocian et al., 1999; MacMillan & Siperstein, 2002).

MacMillan and colleagues have studied this team judgment process of child study teams extensively over the past several years (Gresham, MacMillan, & Bocian, 1996; MacMillan et al. (1998); MacMillan, Gresham, Bocian, & Siperstein, 1997; MacMillan, Gresham, Lopez, & Bocian, 1996; MacMillan, Gresham, Siperstein, & Bocian, 1996; MacMillan, Siperstein, & Gresham, 1996). One of these studies (conducted in California) provides a clear example of how schools use the LD category as a blanket designation for eligibility for special education. At the time of this study, the state of California reported to Office of Special Education (OSEP) that 5.93% of its students were classified as LD. MacMillan et al. classified the referred children they studies using research diagnostic criteria for mild mental retardation (MMR), LD, ADHD, and emotional/behavioral disorder (EBD). Of the 150 referred children studies, schools classified 61 of them as LD.

Applying the research diagnostic criteria to these school identified LD cases reveals the large amount of heterogeneity in this sample. About one third of the cases met the LD criteria only, but the same number meet diagnostic criteria as MMR yet were classified erroneously as LD. A total of 10 cases were classified as LD by schools, but none of these cases met any of the diagnostic criteria for any group (MMR, LD, ADHD, or EBD). Of the 61 school identified LD cases, only 29 cases met research diagnostic criteria for LD classification reflecting less than chance agreement (47.5%). MacMillan and Siperstein (2002) interpreted these findings as follows:

> The evidence clearly documents that the public schools similarly disregarded the "exclusionary criteria" specified in the authoritative definition of LD. By focusing on absolute low achievement and forgoing the requisite discrepancy, schools knowingly

include children with subaverage general intellectual functioning in eligibility classifications of students with LD. (p. 304).

Summary. It is clear from the research conducted over the past 15 years that there is a lack of a consistent definition in policy or practice in identifying children as LD that Lyon (1996) suggested has hampered effective research and practice. Various reactions to this state of affairs have ranged from concluding that the category of LD is not valid or instructionally relevant to impugning child study teams for not implementing classification criteria correctly (Gresham, 2002). However, some researchers have noted that school desire an opportunity to use professional judgment rather than being held accountable to rigid (and irrelevant) formulae for determining who is or is not LD based on IQ-achievement discrepancy (Keogh & Speece, 1996; McLeskey & Waldron, 1991).

It seems that the most serious flaw in the current process by which students are deemed eligible or ineligible for special education services under the category of LD is the almost complete absence of a *direct link* between assessment procedures used for identification and subsequent intervention strategies (Gresham, 2002). It is incontrovertible that most reading difficulties are caused by poor or otherwise inadequate early literacy experiences, inadequate instruction, or some combination of the two (Clay, 1987; Torgesen, Rose, Lindamood, Conway, & Garvan, 1999; Vellutino, Scanlon, & Jacccard, 2003; Vellutino, Scanlon, & Tanzman, 1998). As we shall discuss later, this state of affairs prompts us to find an alternative approach to the identification of LD students that is more accurate and instructionally relevant than the current approach based on IQ-achievement discrepancy.

Assumptions Regarding LD

There are a number of assumptions upon which professionals based their practices in the identification of LD. These assumptions (described below) greatly influence how LD is viewed guide assessment practices in the identification of LD.

Assumption 1: Achievement Difficulties of RD Children are Intrinsic to the Individual

The definition of specific learning disabilities specified in IDEA (1977) states that it is:

> . . . a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, . . . The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. (USOE, 1977; p. 65083).

Clearly, the above definition connotes that LD is a within-child problem due to intrinsic factors. The National Joint Committee on Learning Disabilities (NJCLD; McLoughlin & Netick, 1983) stated, "These disorders are intrinsic to the individual and presumed to be due to central nervous system dysfunction." (p. 22) Some have argued, however, that acceptance of the assumption of intrinsic causes of LD has lead to a virtual ignorance or disregard of the contextual factors that may explain the presence of LD (see Keogh & Speece, 1996; MacMillan & Siperstein, 2002; Speece, 1993).

A fundamental principle underlying the identification of RD using a psychometric approach is to identify those children having specific cognitive deficits (intrinsic factors) and learning problems that are caused by deficits in general intellectual functioning or experiential or instructional deficits (Vellutino, Scanlon, & Tanzman, 1998). Ruling out mental retardation and experiential/instructional deficits are known as the exclusionary criteria in the identification of LD in IDEA (1997). Unfortunately, the current approach to identifying RD using a psychometric

battery of tests including measures of general intelligence, reading ability, and processing does not allow one to rule out experiential deficits as the cause of reading difficulties.

In fact, Clay (1987) argued that the overwhelming majority of reading problems are caused by inadequate or inappropriate preliteracy experiences, reading instruction, or both. Clay's (1987) further suggested that the psychometric approach to identifying causes of reading problems does not take into consideration the cumulative effects of limited prereading experiences and/or inadequate instruction on reading acquisition. Moreover, these limited experiences are often mistaken for basic cognitive deficits (i.e., a specific learning disability) when, in fact, these deficits can be directly linked to experiential deficits. Clay further maintained that most struggling readers can read at grade level if they receive early, individualized, and intense reading instruction focusing on remediation of reading deficits (see Clay, 1985). In short, Clay (1985, 1987) forcefully argues that the overwhelming majority of reading difficulties are due to poor preliteracy experiences and/or poor instruction rather than being due to constitutional or neurologically-based factors.

Vellutino et al. (1998) further suggested that there is strong converging evidence that supports the reading difficulties of impaired readers having at least average intelligence are due to deficits in phonological coding skills such as phoneme awareness, name encoding, name retrieval, and letter-sound correspondence. All of these skills, in turn, greatly affect the acquisition of reading subskills of word identification, phonological decoding, and spelling (Blachman, 1997; Liberman, Shankweiler, & Liberman, 1989; Vellutino, 1979, 1987; Vellutino, Scanlon, & Lyon, 2000; Vellutino, Scanlon, Sipay, Small, Pratt, Chen, & Denkla, 1996).

Lastly, there is now an irrefutable body of empirical evidence showing that visual-perceptual deficits such as visual-motor integration difficulties and visual-spatial reasoning

problems do not reliably differentiate poor readers from normal readers (Olson, Kleigl, & Davidson, 1983; Vellutino, 1979, 1987; Vellutino et al., 1998, Vellutino et al., 1996). Nationally representative surveys of practicing school psychologists, however, show that tests of visual-motor integration and visual-spatial reasoning such as the Bender-Gestalt (Koppitz, 1975) and the Beery-Buktenica Developmental Test of Visual-Motor Integration (Beery, 1997) are among the most frequently used in LD assessments by school psychologists (Wilson & Reschly, 1996). This practice continues unabated despite the virtual absence of any empirical evidence supporting a relationship between visual-spatial reasoning or visual-motor integration abilities and reading.

In summary, the research reviewed above strongly suggests that many children who are or might be identified as RD do not have so-called intrinsic, constitutional deficits that explain their reading difficulties. The majority of children having reading difficulties and are labeled LD are more accurately characterized as "instructional causalities" (Vaughn, 2003). As Clay (1987) maintained, these children "learn to be learning disabled" because they are not exposed to early fundamental literacy skills (e.g., phoneme awareness, print concepts, letter-sound correspondence). Given this unfortunate history of instructional deficiency, poor readers fall further and further behind their peers and masquerade as children with LD (Vellutino et al., 1998).

Assumption 2: IQ-Achievement Discrepancy is a Valid Marker for RD

The most fundamental assumption in the identification of RD is that the 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 than students not showing such discrepancies (nondiscrepant

underachievers). Four methods have been used to operationalize RD based on a discrepancy-based conceptualization: (a) grade level deviations, (b) expectancy formulae, (c) simple differences in standard scores, and (d) regressed standard score analysis (Berringer & Abbott, 1994). All of these methods have serious measurement difficulties including unreliability, additivity of error components, regression artifacts, and scaling difficulties.

Discrepancy based on *grade-level deviation* is based on the inaccurate assumption that all students should be functioning at grade level (Gresham, 2002). This assumption ignores the most basic principle of standardized, norm-referenced achievement testing that mandates that half of the students will be above grade level and half will be below grade level. This problem in using this approach in defining LD is the extent of the deviation below grade level one might use to label a child as LD. This, in turn, is influenced by level of intelligence, socioeconomic status of the school or district, and the myriad of measurement problems associated with grade equivalent scores (Crocker & Algina, 1986; Gresham, 2002).

Deviation from *expected grade level* in an academic area is another approach taking into account the child's IQ is another approach used to identify RD. The *Diagnostic and Statistical Manual* (DSM-IV) used this approach to identify reading and mathematics disorders (American Psychiatric Association, 1994). Berringer and Abbott (1994) have criticized this approach because is uses grade-equivalent scores that vary across grade levels in terms of raw scores and the lack of difficulties in comparing grade equivalents across different achievement tests.

The *simple difference method* is a third approach used to operationalize discrepancy in the identification of LD. This method computes simple standard scores differences between IQ and achievement measures and uses this difference to identify a child as LD. This method suffers

from a host of measurement problems including unreliability of difference scores and the corresponding effects of statistical regression (Gresham, 2002).

Perhaps the most popular method of calculating discrepancy is the *regression discrepancy approach* which controls for unreliability and attendant statistical regression effects by taking into account the measurement errors of the aptitude, achievement, and difference scores (see Reynolds, 1984). The major conceptual and empirical difficulty with this method is that it makes the assumption that IQ is the exclusive and self-limiting cause of achievement (Gresham, 2002). As we describe in a subsequent section, this assumption is simply not supported by empirical research.

Is discrepancy, however calculated, a valid marker for the presence of LD? Several converging lines of research strongly suggest that it is not. Independent researchers, however, have not replicated the previously mentioned Rutter and Yule (1975) Isle of Wight studies (Fletcher, Shaywitz, Shankweiler, Katz, Liberman, Steubing, et al., 1994; Rogers, 1983; Share, McGee, McKenzie, Williams, & Silva, 1987; Stanovich & Siegel, 1994). This line of research provides a direct test of the hypothesis that discrepant low achievers (RD) are different from nondiscrepant low achievers ("garden variety" low achievers). The Fletcher et al. (1994) and Stanovich and Siegel (1994) studies showed that discrepant learners performed no differently than nondiscrepant learners on measures reading achievement, cognitive abilities, phoneme awareness, orthographic awareness, short-term verbal memory, visual analysis, and word retrieval (see Vellutino et al., 1998 for a review).

Siegel (1988) investigated whether or not poor readers and normal readers differed on reading-related cognitive skills independent of IQ level. She divided a large sample of 250 poor readers and 719 average readers into 1 of 4 IQ bands: <80, 81-90, 91-109, and >110 and

contrasted the poor and average readers within each IQ band on a variety of cognitive measures related to reading (e.g., pseudoword decoding, verbal short-term memory, and phoneme awareness). Siegel's results showed that a large number of poor and average readers fell within each of the four IQ bands suggesting that IQ was not a significant predictor of reading-related cognitive skills. She concluded that measures of language achievement were better predictors of reading than IQ.

Vellutino et al. (2000) calculated IQ-achievement discrepancy scores in a sample of 74 poor readers and 28 average IQ and 37 above-average IQ normal readers. The hypothesis Vellutino et al. tested was that IQ-achievement discrepancy should reliably distinguish between children with and without reading disability. These authors found that for poor readers, the size of the IQ-achievement discrepancy was dictated primarily by the magnitude of the child's reading deficit whereas for normal readers, the size of the discrepancy was determined largely by the child's IQ level. In fact, many of Vellutino et al.'s normal readers would have been erroneously classified as reading disabled. Similar findings reported by Fletcher et al. (1994) lead that group to conclude that use of the IQ-achievement discrepancy to define RD essentially sorts children by IQ given that higher IQ children are more likely to show a discrepancy than lower IQ children.

Several meta-analyses have been conducted in which low achieving (LA) and RD groups have been contrasted on a variety of measures (Fuchs, Fuchs, Mathes, & Lipsey, 2002; Hoskyn & Swanson, 2000; Steubing, Fletcher, Ledoux, Lyon, Shaywitz, & Shaywitz, in press). Hoskyn and Swanson (2000) selected 19 studies that contrasted LA (nondiscrepant) and 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 processing deficit, a finding consistent with other findings in reading disability research (Perfitti, Beck, Bell, & Hughes, 1987; Stanovich & Siegel, 1994; Torgesen, 1999;Torgesen & Burgess, 1998; Torgesen, Burgess, & Rashotte, 1996).

More recently, Steubing et al. (in press) conducted a meta-analysis of 46 studies that investigated the validity of the IQ-achievement discrepancy for children with reading disabilities in the domains of achievement, behavior, and cognitive skills. These studies contrasted discrepant (RD) and nondiscrepant (LA) children and included most of the 19 studies found in the Hoskyn and Swanson (2000) meta-analysis. Steubing et al. reported statistically insignificant effect sizes in between RD and LA groups in the domains of achievement (-.12) and behavior (-.05). These authors found a small (.30) effect size for measures 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 surmised that defining RD on the basis of IQ-achievement discrepancy has little evidence for validity.

Finally, Fuchs et al. (2002) conducted a meta-analysis of 79 studies contrasting LD and LA groups on a variety of measures. These authors sought to address the issue of whether LD and LA groups differed in the *degree* of underachievement or were these groups different in the *kind* of underachievement. Unlike the Steubing et al. (in press) meta-analysis, Fuchs et al. assumed their LD group (designated as such based on a LD label) had a severe discrepancy rather than requiring independent verification of whether or not they did have such a

discrepancy. Research reviewed earlier in this paper showed the flaws of using school-identified LD samples versus LD samples using research diagnostic criteria (see MacMillan & Siperstein, 2002 for a review).

An analysis of 112 effect sizes showed a weighted mean effect size of .61 (95% CI: .56 to .65). There was, however, considerable hetereogeneity among studies regarding the size of differences between the LD and LA groups. Fuchs et al. interpreted this .61 effect size as large suggesting that the LD and LA groups could be differentiated. A .61 effect size is considered moderate and a large effect size should be .80 or greater (Cohen, Cohen, West, & Aiken, 2003). This .61 effect size translates into only a nine- point (*M*=100; *SD*=15) standard score difference between LD and LA groups. If one assumes a liberal median reliability coefficient of .90 for reading domain measures used to calculate the effect sizes and a standard error of measurement of 4.74 (*SD*=15), a 95% confidence interval around this effect size would be +9.48 standard score points. One cannot convincingly make the case that an effect size that is within the range of measurement error represents substantial differences in *kind* and thereby validates the LD construct as being different from LA.

Assumption 3: IQ-Achievement Discrepant Learners Are More Resistant to Intervention

A fundamental assumption held by many LD professionals is that discrepant learners somehow are more resistant to intervention than nondiscrepant learners. There is now a large body of evidence that overwhelmingly disputes the assumption regarding the resistance of discrepant learners to intervention (Clay, 1987; Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998; Iversen & Tunmer, 1993; Siegel, 1989; Torgesen et al., 2001; Vellutino et al., 2000; Vellutino et al., 1996; Wasik & Slavin, 1993). Results from the Connecticut Longitudinal Study comparing discrepant (LD) and nondiscrepant (LA) groups showed no differences

between these groups on a battery of child, teacher, and parent measures and no differences in reading growth from 1st to 12th grade (Francis, Shaywitz, Steubing, Shaywitz, & Fletcher, 1996; Shaywitz, Fletcher, Holahan, & Shaywitz, 1992).

Vellutino et al. (2000) directly tested the resistance hypothesis in a sample of 74 poor readers and 65 average readers in first grade (*n*=28 with average IQ and *n*=37 with above-average IQ). The poor readers received 30 minutes of one-to-one tutoring per day five days a week for a total of 15 weeks (see Vellutino et al., 1996 for a complete description of the reading instruction). After 15 weeks of instruction, Vellutino et al. classified the poor readers into 1 of 4 groups based on growth rates in reading: Very Limited Growth, Limited Growth, Good Growth, and Very Good Growth. To test the hypothesis that IQ-achievement discrepancy is a significant predictor of response to instruction, Vellutino calculated IQ-achievement discrepancy scores for each child in both the average reader and poor reader groups. Subsequently, slope measures in reading (reflecting growth) were calculated for each child based using IQ-achievement discrepancy as the predictor variable. This analysis showed a nonsignificant relationship between the magnitude of IQ-achievement discrepancy and the rate of growth in reading resulting from the remediation intervention. Vellutino et al., concluded:

> . . . the IQ-achievement discrepancy does not reliably distinguish between disabled and nondisabled readers . . . Neither does it distinguish between children who were found to be difficult to remediate and those who were readily remediated, prior to institution of remediation, and it does not predict response to remediation. (p. 235)

The research literature clearly shows that IQ-achievement discrepancy is neither a reliable nor valid predictor of response to intervention for reading difficulties. In fact, IQ is not a strong predictor of basic reading skills (word identification, phoneme segmentation,

phonological decoding, verbal memory, rapid naming) (Fletcher et al., 1994; Share et al., 1987; Stanovich & Siegel, 1994; Vellutino et al., 2000; Vellutino, 2001). Continued use of the discrepancy approach to identify which children will and will not respond to remedial efforts is based on the fallacy that discrepant learners are in most need of special education relative to nondiscrepant learners.

Assessment in RD Students Reveals Valid Aptitude X Treatment Interactions

Lee Cronbach (1957) described the concept of aptitude X treatment interaction (ATI) in his presidential address to the American Psychological Association. He argued that the scientific disciplines of correlational (individual differences research) and experimental (treatment) psychology should be integrated using ATI research designs. ATI research concentrates on the measurement of valid aptitudes (characteristics or traits) and how these aptitudes might interact with various treatments (instructional methods or types of therapy). One can discover interactions when instructional methods (treatments) differentially affect persons who differ in measured aptitudes or characteristics (Gresham, 2002; Reschly & Ysseldyke, 2002).

Fundamentally, ATI studies require the measurement of a minimum of two aptitudes and two treatments thereby creating four mean data points. An example might be to use the Verbal Comprehension and Perceptual Organization factors from the Wechsler Intelligence Scale for Children (WISC-III, Wechsler, 1991) to operationalize two aptitudes. The two treatments might be a phonics-based approach to reading and a whole-word approach to reading. The research hypotheses tested in this study would be that "verbal learners" will respond better to phonics-based instruction and "visual learners" will respond best to whole-word instruction thereby revealing a disordinal aptitude X treatment interaction. School psychologists frequently use this logic to make instructional recommendations to teachers based on cognitive abilities tests (see Gresham & Witt, 1997; Reschly & Ysseldyke, 2002).

Although a comprehensive review of the ATI literature exceeds the scope of the current paper, a number of literature reviews do not support the feasibility of matching aptitudes to treatments for children with RD or other learning problems (Gresham, 2002). Reviews of the modality matching literature do not support the practice of matching aptitudes to treatments in

remediating learning difficulties (Arter & Jenkins, 1979; Kavale & Forness, 1987, 1995; Ysseldyke & Mirkin, 1982). Research conducted in the cognitive style/processing literatures do not consistently show independently replicated ATIs based on representative samples of RD children (Ayres & Cooley, 1986; Ayers, Cooley, & Severson, 1988; Das, 1995; Das, Naglieri, & Kirby, 1995; Good, Vollmer, Creek, Katz, & Chowdhri, 1993).

Neuropsychological models also use ATI logic in making remedial recommendations based on inferred brain strengths or functioning. A child with left hemispheric strength might be assumed to learn more effectively or efficiently using phonics or other verbally delivered instruction. In contrast, a child with right hemispheric strengths might be assumed to perform better when instructed using holistic or visually presented instruction.

Despite the face valid appeal of ATI logic in neuropsychological assessment (D'Amato, Rothlisberg, & Work, 1999; Hynd, 1989; Reynolds & Fletcher-Janzen, 1989), sound empirical studies demonstrating significant ATIs based on neuropsychological assessment, interpretation, and treatment recommendations for children with RD are nonexistent (Gresham, 2002). In short, the whole practice of applying ATI logic in neuropsychological assessment to make treatment recommendations for students with RD should be summarily abandoned based on the lack of scientific support (Gresham, 2002; Reschly & Gresham, 1989; Teeter, 1987).

The disappointing results of ATI research based on modality matching, cognitive style/processing, and neuropsychological assessment provide little empirical support for matching aptitudes to treatments in hopes of discovering an ATI. Expressing his frustration with ATI research, Cronbach (1975) stated: "Once we attend to interactions, we enter a hall of mirrors that extends to infinity" (p. 119). Cronbach argued that we should give up our search for ATIs and suggests that we engage in context-specific evaluation and short-run empiricism: "One

monitors responses to treatment and adjusts it" (p. 126). Cronbach's (1975) recommendations are the basis for responsiveness to treatment as the criterion in making LD eligibility determinations. The following section provides the conceptual basis for responsiveness to intervention.

Responsiveness to Intervention

Response to intervention (RTI) is based on a very simple premise: interventions should take place within general education using multi-tiered interventions to determine which children need what services delivered with how much intensity. If the amount of services needed is beyond the scope of general education, then the child is made eligible to receive additional resources available through special education. This practice is currently occurring on a limited basis in the United States (Reschly, Tilly, & Grimes, 1999). Moreover, both the House of Representatives (HB 1350) and Senate (SB 1248) versions of the reauthorization of IDEA explicitly state the following in identical language:

> Notwithstanding section 602 of the Act or any other provision of law, when determining a child has a specific learning disability as defined in 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 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)

Quite clearly, the Congress was convinced by testimony before its respective House and Senate Education Subcommittees as well as testimony presented to the President's Commission on Excellence in Special Education (2002) of the problems with IQ-achievement discrepancy and the benefits of using a RTI approach to LD eligibility determination. RTI captures a fluid concept (intervention responsiveness) and dichotomizes two possible outcomes for decision-

making purposes. When presented with a evidenced-based intervention, the student either *responds adequately* to the intervention and the problem improves or the students *does not respond adequately* and the problem stays the same or gets worse.

RTI Defined

RTI can be defined as the change in behavior or performance as a function of intervention (Gresham, 1991, 2002). RTI is uses the concept of discrepancy, however this discrepancy is between pre- and post-intervention levels of performance rather than IQ-achievement discrepancy. In a problem-solving model, a "problem" is defined as current and expected levels of performance (Bergan & Kratochwill, 1990; Upah & Tilly, 2002). The failure to demonstrate a discrepancy between pre-intervention and post-intervention levels of performance can be taken as evidence for the presence of LD (Gresham, 2002; Steubing et al., 2002; Vellutino et al., 1996).

Gresham (2002) provided the following example in the area of reading to demonstrate the logic of RTI. A reading intervention whose goal is to produce oral reading fluency would be considered effective if it produced reading fluency rapidly and reliably during reading instruction and if reading fluency maintained after the intervention was withdrawn. However, if oral reading fluency deteriorated after the intervention is withdrawn, teachers would probably not be satisfied with the rate of oral reading fluency no matter how well the student read during the intervention. Moreover, if oral reading occurred at low rates with numerous errors during the intervention, teachers most likely would conclude that the intervention was ineffective and would seek to extend, intensify, or change the reading intervention.

Advantages of RTI

The advantages of using a RTI are many and compelling. We highlight three major advantages of this approach in the following sections.

Quicker intervention services provided to students and teachers. Perhaps the most compelling idea for researchers and practitioners alike is the possibility of providing help to struggling children immediately. Traditional refer-test-place models, operating on prescribed timelines with the requisite avalanche of paperwork, can take several months to return a verdict regarding eligibility for special education. With RTI, there is no need to "wait for the child to fail." Universal screening can be combined with RTI allowing for early, proactive identification of at-risk children and establishing baseline scores for measuring growth in performance over time.

Fuchs and Fuchs (1998) advocated a RTI approach to identification of LD because it focuses on maximizing general education's potential effectiveness for all students, not just the referred child. Decisions about the need for special education are reserved until the effects of the instructional intervention or adaptations have been assessed in the general education classroom and data *verify* that a special education program would enhance learning. RTI approaches occurring in a general education classroom can utilize teachers trained in effective remedial strategies that increase the likelihood the referred child's peers will benefit from the intervention.

More accurate decision-making. The use of universal screening and RTI provides an objective basis for making decisions regarding special education eligibility and placement decisions. Research suggests that when teachers are asked to delineate the causes of students' problems prompting their initial referral for evaluation, they overwhelmingly attributed the causes of students' problems to within-child causes and the home situation (Ysseldyke, Pianta,

Christenson, Wang, & Algozzine, 1983). Also, teachers indicated that they hoped that the outcome of the referral would be assessment and ultimately placement in special education, as opposed to obtaining strategies to help the child remain in their classroom. It is possible, perhaps even likely, that teachers form these expectations based on experiences with school-based professionals (e.g., school psychologists) who have failed to provide data that are useful to teachers and helpful to students (Gresham & Witt, 1997; Severson, Pickett, & Hetrick, 1985).

The expectations and philosophies of the decision-making team may introduce error and bias into the decision-making process for special education. RTI decreases the need to rely solely on teacher referral to evaluate the severity of the referral problem, the appropriateness of the intervention strategy, and the success of intervention implementation. As such, RTI produces more accurate decision-making and more effective problem solving (MacMillan et al., 1998; Shinn, Tindal, & Spira, 1987). Because RTI increases accurate identification, overidentification errors are reduced. A major cause of overidentification of LD is that professionals have not been able to effectively distinguish between children with LD from "garden variety" low achievers (Fletcher et al., 1994; Gresham, 2002; Gresham & Witt, 1997).

Many have argued that the primary problem with many low achievers is the lack of exposure to good instruction, or what Vaughn (2003) calls "instructional causalities" (see Clay, 1987; Donovan & Cross, 2002; Velluntino et al., 1998). Implementation of scientifically validated interventions in general education classrooms can dramatically reduce the rather large false positive identification errors in the field of LD (Gresham, 2002). The philosophical shift from the refer-test-place approach to refer-intervene-evaluate (RTI) needs to make its way into schools and classrooms on a much larger scale.

Assessments have treatment validity. A final advantage of using a RTI model is that is utilizes assessments that have treatment validity. Treatment validity refers to the degree to which any assessment procedure informs or otherwise contributes to beneficial outcomes for individuals (Cone, 1989; Hayes, Nelson, & Jarrett, 1987). Treatment validity contains the notion of incremental validity because it requires assessments to improve prediction over and above existing procedures (Sechrest, 1963). Treatment validity also includes the principles of utility and cost-benefit analysis that is used extensively in the personnel selection literature (Mischel, 1968; Wiggins, 1973). Finally, treatment validity involves Messick's (1995) idea of evidential basis for test interpretation and use, specifically as it relates to construct validity, relevance/utility, and social consequences.

Gresham (2002) pointed out that it is entirely possible for a particular test interpretation to have construct validity, but have little or no relevance or utility for a particular use of a test such as recommendations for intervention based on the test. This is the major problem with using intelligence tests and attendant profile analyses in futile attempts to make instructionally relevant intervention recommendations to teachers (Gresham & Witt, 1997). Witt and Gresham, 1985) stated in their review of the Wechsler Intelligence Test for Children-Revised (WISC-R):

> "The WISC-R lacks treatment validity in that its use does not enhance remedial interventions for children who show specific academic skill deficiencies . . . For a test to have treatment validity, it must lead to better treatments (i.e., better educational programs, teaching strategies, etc.)" (p. 1717).

Fuchs and Fuchs (1998) described the value of a treatment validity approach for the LD field by suggesting that this approach enhances regular education's potential effectiveness for all students. Decisions regarding the need for special education are withheld until the effects of

instructional adaptations have been assessed in the regular education classroom and data verify that a special education program is needed to enhance learning. One approach to assessment, curriculum-based measurement (CBM), meets the treatment validity criterion and can be used in making eligibility determinations for special education (Fuchs & Fuchs, 1997; Reschly & Tilly, 1999; Shinn, Good, & Parker, 1999).

Using a treatment validity approach to identify students with LD must meet several technical requirements. These requirements are: (a) ability of measures to model academic growth over time (Burchinal, Bailey, & Synder, 1994; Fuchs & Fuchs, 1997, 1998; Velluntino et al., 1998; Velluntino et al., 1996), (b) availability of validated treatment protocols (Berninger & Abbott, 1994; Torgesen et al., 2001), (c) capability of distinguishing between ineffective instruction and unacceptable student learning (Fuchs & Fuchs, 1997, 1998; Velluntino et al., 2000), (d) ability to inform instructional decisions (Fuchs & Fuchs, 1997, 1998; Velluntino et al., 1996, 1998), and (e) sensitivity in detecting treatment effects (Fuchs & Fuchs, 1997; Marston, Fuchs, & Deno, 1986; Marston, 1987-88; Velluntino et al., 1996, 1998).

Interventions are designed in such a way that changes in the target behavior (the dependent variable) can be attributed to systematic and controlled changes in an intervention (the independent variable). Typically, this change is assessed using a pretest/posttest design in which experimental (intervention) and control groups are measured before and after the intervention. Changes from pretest to posttest are assessed either using a repeated measures ANOVA or ANCOVA (using pretest scores as the covariate) or by computing simple mean differences between groups. These types of analyses can reveal whether or not an intervention produced mean differences on a dependent variable, these analyses do not provide sufficient data to model individual change or growth over time (Burchinal et al., 1994).

One alternative to pretest/posttest comparisons is growth curve analysis (GCA) that uses hierarchical linear models (HLM) to model academic growth over time (Burchinal et al., 1994; Raudenbush & Bryk, 2002). GCA analysis addresses three questions: (a) What are the patterns of change for both individuals and groups? (b) Do different groups (intervention and control or two intervention groups) show different patterns of change over time?, and (c) What are the moderators of change over time for groups and individuals? Fuchs and Fuchs (1998) have used CBM as a means of modeling academic growth within the special education eligibility determination process. CBM meets many of the assumptions of GCA (normality, interval/ratio measurement, ability to model growth) and can be used in making LD eligibility determinations (Gresham, 2002).

Another requirement in adopting a treatment validity approach is the availability of validated treatment protocols for students with learning difficulties. Many students classified as LD fail to acquire basic academic skills not because of some underlying processing disorder or some other within-child problem, but because they have not been exposed to adequate instruction (Clay, 1987; Gresham, 2002; Velluntino et al., 1998). Most reading difficulties of children labeled as LD are due to inadequate preliteracy experiences, inadequate instruction, or both (see Velluntino et al., 1996, 1998, 2000).

A number of validated treatment protocols exist that have been shown to be effective in remediating reading difficulties (Clay, 1985; Perfetti et al., 1987; Torgesen et al., 2002; Velluntino et al., 1996, 1998; Vaughn, 2003). These will not be reviewed here, however, these approaches to reading share a number of similarities such as focus on phonemic awareness, phonological decoding, letter-sound correspondence, and writing The meta-analysis by Swanson

and Hoskyn, 1999 provides more detail regarding the components of effective remedial instruction for students with reading difficulties.

Conclusions

We have argued in this paper that a RTI approach to defining RD should be adopted because of the theoretical and psychometric problems with using the traditional IQ-achievement discrepancy to identify this population. There is now substantial evidence that the use of a discrepancy approach to define RD is not valid and lacks treatment validity.

Some detractors of the RTI method of identifying students with RD may argue that medical diagnoses are not ruled in or ruled out based on how a patient responds to intense treatment efforts. We acknowledge this, however, we also would maintain that medical diagnoses often have direct treatment implications (e.g., cancer, diabetes, high cholesterol) and the causes of medical problems are known (unlike LD). Additionally, the intensity of medical treatments is matched to the severity of the medical diagnosis and increases in the "dose" of treatment are calibrated to the degree of unresponsiveness to that treatment. Our point is that not all students will require the most intense form of academic remediation and the strength, intensity, and duration of treatment are increased in proportion to the student's unresponsiveness to that treatment (Gresham, 2002).

Part of the appeal of RTI as a decision-making tool is that it might control for inadequate instruction and allow practitioners to make accurate discriminations where little variation in level of performance (across peers) exists and allow for decision-making within an ecologically valid setting (i.e., the general education classroom). RTI as a decision-making tool has the potential to quantify the dynamic nature of multiple contextual variables (molar and molecular) that may impact child performance at any given moment in time and program interventions designed to

maximize the degree of "fit" between the child and the environment (Sheridan & Gutkin, 2000). Such an approach might provide some protection against disproportionate identification of children as needing special education services in low-achieving schools and "raise the bar" concerning intervention goals. That is, the RTI model provides for ongoing progress monitoring and therefore increases program accountability. Intervention goals then may be broadly conceived as immediate improvement in performance *and* growth toward long-term functional competence (Torgesen, 2002).

A RTI approach offers practitioners an opportunity to "aim directly at the target" in that it uses a child's inadequate response to an evidence-based intervention as an index of the presence of RD. This model 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 render decisions having little relevance for consumers (Brooks & Baumeister, 1977; Gresham & Witt, 1997; Macmann et al., 1989; Wolf, 1978). The RTI approach will require practitioners (e.g., school psychologists) to go where children must perform (the classroom), directly measure the behaviors necessary for successful performance using low inference assessment tools, and judge the outcomes of interventions as a teacher might (i.e., Did the intervention produce an acceptable level of performance?).

Pending Issues in RTI

There are several questions that merit research attention in the adoption of the RTI approach in identifying RD. These include: (a) what constitutes an evidence-based intervention? (b) What is the optimal length and intensity of an intervention?, and (c) How will the integrity of interventions be guaranteed? (Gresham, 2002).

What constitutes an evidence-based intervention? There is now a consensus concerning the most important skills to target in reading interventions and which instructional methods have been shown to be the most effective in remediating reading difficulties. Over the past 20 years, reading research demonstrates that 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). Growth in reading is best predicted initial levels of phonological skill and not by verbal ability or discrepancy between IQ and reading achievement (Torgesen et al., 2001; Vellutino et al., 1996, 1998). Weaknesses in phonological skills require reading instruction that is more phonemically explicit and systematic that might be provided to other children and there are numerous ways in which this sort of instruction might be delivered (Torgesen et al., 2001).

A meta-analysis by Swanson and Hoskyn (1999) informs the field about the most effective remedial interventions for reading problems of children with LD. Results from this meta-analysis showed that interventions using a combination of direct instruction and strategy training produced the largest effect sizes with 80% of the treatment group participants having average reading scores higher than or equal to the control group participants. This meta-analysis showed that the average intervention was about 13 hours spread over 10 weeks. Vellutino et al.'s (1996) study was comparable consisting of 35 hours over 15 weeks and Torgesen et al.'s (2001) study was much more intense with approximately 68 hours of instruction over 8-9 weeks. The large variability among research studies makes comparisons across studies difficult. However, there are numerous evidence-based intervention options available in reading (see Foorman et al., 1997) and decisions concerning which of these should be selected must be made at the local level.

What is the optimal length and intensity of an intervention? The fundamental principle underlying an RTI approach it that the length and intensity of an intervention will depend on the degree of a student's responsiveness to that intervention. A useful model is one that has been used by the Heartland Area Education Agency (AEA) in Iowa to make special education entitlement decisions (see Grimes, 2002; Reschly & Tilly, 1999; Reschly & Ysseldyke, 2002). The Heartland AEA model (Heartland AEA, 2002) is a multiple gating, problem solving process that is based on five assumptions:

(1) The intensity of intervention (and thus costs) is matched to the degree of unresponsiveness to that intervention.

(2) Progression through various levels of intervention intensity is based on inadequate response to interventions that have been implemented with integrity.

(3) Decision concerning movement though levels of intervention intensity are based on continuous progress monitoring of data collected from multiple sources.

(4) As a student moves through different intensity levels, an increasing body of data are collected to inform decision-making.

(5) Special education and subsequent IEP determinations are only considered after the student shows inadequate responsiveness to intervention at earlier levels of intervention intensity.

Grimes (2002) indicates that RTI models similar to Heartland AEA (2001) model have been used in a number of states including Florida, Illinois, Kansas, Minnesota, Ohio, South Carolina, and Wisconsin. Although there are variations, RTI approaches in these various states share five common features: (1) emphasis on *direct assessment* of behavior, (2) linking assessment information to intervention services, (3) progress monitoring of student

responsiveness to interventions, (4) outcome judgments based on student data, and (5) parental involvement in the decision-making process (Grimes, 2002).

How will the integrity of interventions be guaranteed? One of the most important characteristics of a RTI model to defining RD is the establishment and maintenance of the integrity of interventions. Treatment integrity or fidelity refers to the degree to which an intervention is implemented or delivered as intended (Gresham, 1989; Yeaton & Sechrest, 1981). In all likelihood, the inadequate response to many instructional interventions is due to the poor integrity with which these interventions were implemented. Treatment integrity is a central feature in any RTI model because the entire concept of "responsiveness" is predicated upon the reliable and accurate implementation of treatment procedures.

What do we know about the treatment integrity of interventions with the LD population? Gresham and colleagues analyzed articles from three major learning disability journals between 1995 to 1999 (*Learning Disability Quarterly, Journal of Learning Disabilities,* and *Learning Disabilities: Research & Practice*). Of the 479 articles published, 65 articles were intervention research studies. Of these 65 articles, 12 articles or 18.5% reported actual measures and data on treatment integrity (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). Earlier, Swanson and colleagues showed that less than 2% of the studies they reviewed reported data on treatment integrity (Swanson, Carson, & Saches-Lee, 1996). Subsequently, in their review of single-case design 24 studies, Swanson and Saches-Lee (2000) showed that only eight studies reported specific data on treatment integrity.

Clearly, any use of a RTI model must be concerned with the integrity issue. Some years ago, Peterson, Wonderlich, and Homer (1982) argued that there is a "curious double standard" in the research literature with respect to the measurement and assessment of the reliability for

independent and dependent variables. In short, reliability data are almost always reported for the *dependent variable* in treatment-outcome research, but the reliability of treatment implementation (treatment integrity) is rarely reported for the independent (treatment) variable. It is incumbent upon the field to assess, monitor, and ensure the integrity of interventions in adopting a RTI model.

It is abundantly clear that the "business as usual" approach to defining RD using IQ-achievement discrepancy is indefensible considering is lack of validity and its failure to inform intelligent treatment decisions. Detractors may argue that a RTI model does not improve upon a discrepancy model in identifying students as RD, that it has serious measurement problems, and that it will be extremely expensive. These arguments may or may not be true, however, there are enough empirical data at this point to suggest that a RTI model can be effectively utilized in identifying children as RD. Many arguments against change in the RD field to a RTI model are based more on political, financial, or personal reasons than on sound empirical evidence. It is time to move beyond this type of thinking to an approach that improves outcomes for children.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Arter, J., & Jenkins, J. (1979). Differential-diagnosis-prescriptive teaching: A critical appraisal. *Review of Educational Research, 49,* 517-555.

Ayres, R., & Cooley, E. (1986). Sequential versus simultaneous processing on the K-ABC: Validity in predicting learning success. *Journal of Psychoeducational Assesment, 4,* 211-220.

Ayres, R., Cooley, E., & Severson, H. (1988). Educational translation of the Kaufman Assessment Battery for Children: A construct validity study. *Journal of Psychoeducational Assessment, 4,* 113-124.

Bateman, B. (1965). An educational view of a diagnostic approach to learning disorders. In J. Hellmuth (Ed.), *Learning disorders: Vol. 1* (pp. 219-239). Seattle, WA: Special Child Publications.

Beery, K. (1997). *Developmental Test of Visual-Motor Integration-Fourth Edition.* Cleveland, OH: Modern Curriculum.

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

Berringer, V., & Abbott, R. (1994). Redefining learning disabilities: Moving beyond apitude-achievement discrepancies to failure to respond to validated treatment protocols. In G. R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities* (pp. 163-183). Baltimore: Paul H. Brookes.

Blachman, B. (Ed.) (1997). *Foundations of reading acquisition and dyslexia: Implications for early intervention*. Mahwah, NJ: Lawrence Erlbaum.

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.

Burchinal, M., Bailey, D., & Synder, P. (1994). Using growth curve analysis to evaluate child change in longitudinal investigations. *Journal of Early Intervention, 18,* 403-423.

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.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Cone, J.D. (1989). Is there utility for treatment utility? *American Psychologist, 44*, 1241-1242.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory.* New York: Holt, Rinehart, & Winston.

Cronbach, L. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671-684.

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

D'Amato, R.C., Rothlisberg, B., & Work, P. (1999). Neuropsychological assessment for intervention. In C. Reynolds & T. Gutkin (Eds.), *Handbook of school psychology* (3rd ed.), pp. 452-475). New York: Wiley.

Das, J.P. (1995). Neurocognitive approach to remediation: The PREP Model. *Canadian Journal of School Psychology, 9,* 157-173.

Das, J.P., Naglieri, J., & Kirby, J. (1995). *Assessment of cognitive processes.* Needham, MA: Allyn & Bacon.

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

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.

Franzen, R. (1920). The accomplishment quotient: A school mark in terms of individual capacity. *Teachers College Record, 21,* 432-440.

Fuchs, D., Fuchs, L., Mathes, P., Lipsey, M., & Roberts, H. (2002). "Learning disabilities": Just a fancy term for low achievement? A meta-analysis of reading differences between low achievers with and without the label. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 737-762). Mahwah, NJ: Lawrence Erlbaum.

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.

Gerber, M., & Semmel, M. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 14,* 137-146.

Good, R., Vollmer, M., Creek, R., Katz, L., & Chowdhri, S. (1993). Treatment utility of the Kaufman Assessment Battery for Children: Effects of matching instruction and student processing strength. *School Psychology Review, 22,* 8-26.

Gresham, F.M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18,* 37-50.

Gresham, F.M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. *School Psychology Review, 20,* 23-36.

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.

Gresham, F.M., MacMillan, D., & Bocian, K. (1996). Learning disabilities, low achievement, and mild mental retardation: More alike than different? *Journal of Learning Disabilities, 29*, 570-581.

Gresham, F.M., MacMillan, D., & Bocian, K. (1997). Teachers as "tests": Differential validity of teacher judgments in identifying students at-risk for learning difficulties. *School Psychology Review, 26*, 47-60.

Gresham, F.M., MacMillan, D., Beebe-Frankenberger, M., & Bocian, K. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice, 15*, 198-205.

Gresham, F.M., Reschly, D., & Carey, M. (1987). Teachers as "tests": Classification accuracy and concurrent validation in the identification of learning disabled children. *School Psychology Review, 16*, 543-563.

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.

Hallahan, D., & Mercer, C. (2002). Learning disabilities: Historical perspectives. In R. Bradely, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (1-67). Mahwah, NJ: Lawrence Erlbaum.

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

Heartland Area Education Agency (2001). *Procedures manual for special education.* Johnston, IA: Author.

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.

Hynd, G. (1989). Learning disabilities and neuropsychological correlates: Relationship to neurobiological theory. In D. Bakker & H Van der Vlugt (Eds.), *Learning disabilities: Neuropsychological correlates and treatment* (Vol. 1, pp 123-147). Amsterdam: Swets & Zeitlinger.

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

Iversen, S., & Tunmer, W. (1993). Phonological processing skills and the Reading Recovery program. *Journal of Educational Psychology, 85,* 112-126.

Kavale, K. (2002). Discrepancy models in the identification of learning disability. In R. Bradley, L. Danielson, & D. Hallahan (Eds.), *Identification of learning disability: Research to practice* (pp. 369-426). Mahwah, NJ: Lawrence Erlbaum.

Kavale, K., & Forness, K. (1987). How not to specify learning disability: A rejoinder. *Remedial and Special Education, 8,* 60-62.

Kavale, K., & Forness, S. (1995). *The science of learning disabilities.* San Diego: College-Hill Press.

Keogh, B., & Speece, D. (1996). Learning disabilities within the context of schooling. In D. Speece & B. Keogh (Eds.), *Research on classroom ecologies: Implications for inclusion of children with learning disabilities* (pp. 1-14). Mahwah, NJ: Lawrence Erlbaum.

Kirk, S. (1962). *Education exceptional children.* Boston: Houghton Mifflin.

Koppitz, E. (1975). *The Bender-Gestalt Test for young children: Vol. 2: Research and application, 1963-1973.* New York: Grune & Stratton.

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.

Lyon, G.R. (1985). Learning disabilities research: False starts and broken promises. In S. Vaughn & C. Bos (Eds.), *Research in learning disabilities: Issues and future directions* (pp. 69-83). Boston: Little Brown.

Lyon, G.R. (1996). The future of children: Special education for students with disabilities. *Learning Disabilities, 6,* 54-76.

Lyon, G.R. (1999, December, 12). Special education in state is failing on many fronts. *Los Angeles Times,* p. A1.

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., Gresham, F.M., & Bocian, K. (1998). Discrepancy between definitions of learning disabilities and what schools use: An empirical investigation. *Journal of Learning Disabilities, 31,* 314-326.

MacMillan, D., Gresham, F.M., & Siperstein, G. (1993). Conceptual and psychometric concerns over the 1992 AAMR definition of mental retardation. *American Journal on Mental Retardation, 98,* 324-335.

MacMillan, D., Gresham, F.M., Bocian, K., & Lambros, K. (1998). Current plight of borderline students: Where do they belong? *Education and Treatment of Children, 33,* 83-94.

MacMillan, D., Gresham, F.M., Bocian, K., & Siperstein, G. (1997). The role of assessment in qualifying students as eligible for special education: What is and what's supposed to be. *Focus on Exceptional Children, 33,* 83-94.

MacMillan, D., Gresham, F.M., Lopez, M., & Bocian, K. (1996). Comparison of students nominated for prereferral interventions by ethnicity and gender. *Journal of Special Education, 30,* 133-151.

MacMillan, D., Gresham, F.M., Siperstein, G., & Bocian, K. (1996). The labyrinth of IDEA: School decisions of referred students with subaverage general intelligence. *American Journal on Mental Retardation, 101,* 161-174.

MacMillan, D., Siperstein, G., & Gresham, F.M. (1996). Mild mental retardation: A challenge to its viability as a diagnostic category. *Exceptional Children, 62,* 356-371.

Marston, D. (1987-88). The effectiveness of special education: A time series analysis of reading performance in regular and special education settings. *Journal of Special Education, 21,* 13-26.

Marston, D., Fuchs, L., & Deno, S. (1986). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-based measures. *Diagnostique, 11,* 71-90.

McLeskey, J., & Waldron, N. (1991). Identifying students with learning disabilities: The effect of implementing state guidelines. *Journal of Learning Disabilities, 24,* 501-506.

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.

Mischel, W. (1968). *Personality and assessment.* New York: Wiley.

Monroe, M. (1932). *Children who cannot read.* Chicago: The University of Chicago Press.

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

Olson, R., Kleigl, R., & Davidson, B. (1983). Dyslexic and normal readers' eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 816-825.

Perfetti, C., Beck, I., 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.

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis, 15,* 477-492.

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.

Raudenbush, S., & Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods* 2nd ed.). Thousand Oaks, CA: Sage Publications.

Reschly, D.J., & Gresham, F.M. (1989). Current neuropsychological diagnosis of learning problems: A leap of faith. In C.R. Reynolds & E. Fletcher-Janzen (Eds.), *Handbook of clinical child neuropsychology* (pp. 503-519). New York: Plenum.

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.

Reynolds, C.R., & Fletcher-Janzen, E. (Eds.). *Handbook of clinical child neuropsychology*. New York: Plenum.

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.

Sechrest, L. (1963). Incremental validity: A recommendation. *Educational and Psychological Measurement, 23,* 153-158.

Severson, H., Pickett, M., & Hetrick, D. (1985). Comparing pre-service, elementary, and junior high teachers' perceptions of school psychologists: Two decades later. *Psychology in the Schools, 22,* 179-186.

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.

Sheridan, S., & Gutkin, T. (2000). The ecology of school psychology: Examining and changing our paradigm for the 21st century. *School Psychology Review, 29,* 485-501.

Shinn, M., Good, R., & Parker, C. (1999). Noncategorical special education services with students with severe achievement deficits. In D. Reschly, W.D. Tilly, & J. Grimes (Eds.), *Special education in transition: Functional assessment and noncategorical programming* (pp. 81-106). Longmont, CO: Sopris West.

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

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. (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.

Siegel, L. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22,* 469-478.

Speece, D. (1993). Broadening the scope of classification research: Conceptual and ecological perspectives. In G.R. Lyon, D. Gray, J. Kavanaugh, & N. Krasnegor (Eds.), *Better understanding learning disabilities; New views from research and their implications for public policies* (pp. 57-72). Baltimore: Brookes.

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. (in press). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal.*

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.

Teeter, P. (1987). Review of neuropsychological assessment and intervention with children and adolescents. *School Psychology Review, 16,* 582-593.

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.

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

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

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

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

Vellutino, F., 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.

Wasik, B., & Slavin, R. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quarterly, 28,* 179-200.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-III.* San Antonio, TX: Psychological Corporation.

Wiggins, J. (1973). *Personality and prediction: Principles of personality assessment.* Reading, MA: Addison-Wesley.

Wilson, M., & Reschly, D.J. (1996). Assessment in school psychology training and practice. *School Psychology Review, 25,* 9-23.

Witt, J.C., & Gresham, F.M. (1985). Review of the Wechsler Intelligence Scale for Children-Revised. In J. Mitchell (Ed.), *Ninth mental measurements yearbook* (pp. 1716-1719). Lincoln, NE: Buros Institute.

Wolf, M.M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 2030214.

Yeaton, W., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: Strength, integrity, and effectiveness. *Journal of Consulting and Clinical Psychology, 49,* 156-167.

Ysseldyke, J., Pianta, R., Christenson, S., Wang, J., 7 Algozzine, B. (1983). An analysis of prereferral interventions. *Psychology in the Schools, 20,* 184-190.

Table 1. *Differences in Degree versus Differences in Kind of Disabilities*

Mild Mental Retardation	Profound Mental Retardation
Mild Hearing Loss	Profound Deafness
Overactivity	Attention Deficit/Hyperactivity Disorder
Behavior Problem	Emotional Disturbance
Low Achievement	Specific Learning Disability