IS "LEARNING DISABILITIES" JUST A FANCY TERM FOR LOW ACHIEVEMENT? A META-ANALYSIS OF READING DIFFERENCES BETWEEN LOW ACHIEVERS WITH AND WITHOUT THE LABEL

Douglas Fuchs and Lynn S. Fuchs, Peabody College of Vanderbilt University; Patricia G. Mathes, University of Texas—Houston Health Science Center; Mark W. Lipsey and P. Holley Roberts, Peabody College of Vanderbilt University

This paper has two parts. The first part has three purposes: (a) to provide a brief history of how a formal definition of learning disabilities (LDs) was determined; (b) to explore how politics and research during the past two decades has influenced how people think about the validity of the LD construct; and (c) to explain why a quantitative synthesis, or meta-analysis, was necessary to determine whether low achievers with and without the LD label were more alike or different. The second part of this chapter describes method, results, and implications of the meta-analysis.

HISTORY, POLITICS, AND THE LD CONSTRUCT

In 1978, the first author of this paper went to Milwaukee to the annual conference of the Association of Children with Learning Disabilities (now Learning Disabilities Association) to present his just-completed dissertation research. Because it was his first professional meeting, he remembers it well. But it is memorable for an additional reason: Less than 3 years before, advocates had convinced Congress to include LD as a category of exceptionality in the Education for All Handicapped Children Act, and the celebratory greetings and congratulatory backslapping among advocates and practitioners, as well as their excitement and optimism, pulsed through the meeting rooms, hallways, and bars of the crowded conference hotel.

Contributing to the conference-goers' upbeat mood was their confidence in the validity of the LD construct. Throughout the 1970s, most practitioners, parents, and academics firmly believed that LD represented a discrete classification of exceptionality marked by two unique features: "unexpected" learning failure and "specific" learning failure (e.g., Kavale, 1987; Kavale & Forness, 1998). The child with "unexpected" learning failure was perceived by parents and teachers as generally competent. The learning difficulty was both surprising and puzzling. "Specific" learning failure suggested neurological dysfunction and processing deficits, which were presumed to cause severe problems in reading, writing, or math (e.g., Kavale & Forness).

There were at least two reasons to view "unexpected" and "specific" learning failure as a conceptual anchor and rallying cry for the field. First, as far back as the 1890s, physicians W. Pringle Morgan and John Hinshelwood separately described "the seemingly paradoxical inability of some children of average and superior intelligence to master academic concepts" (Lyon, 2001), a phenomenon documented more extensively by another physician, Samuel Orton, in the 1920s and 1930s (Hallahan & Mercer, 2001).

Second, in 1975, Rutter and Yule reported findings from an epidemiological study that seemed to buttress the clinical observations of Morgan, Hinshelwood, and Orton. Rutter and Yule measured the IQ and reading performance of all 9- and 14-year-olds on the Isle of Wight. The researchers regressed the children's IQ scores on their reading scores to produce a distribution of IQ-predicted reading performance. Scores above the mean represented overachievement (i.e., exceeding prediction); scores below the mean indicated underachievement (i.e., beneath prediction). Whereas such a distribution should resemble a Gaussian curve, with overachievement occurring as frequently as underachievement, Rutter and Yule reported a "hump" at the lower end of the distribution, which, they said, indicated that "extreme degrees" of reading underachievement occur at a greater rate than should be expected (Rutter & Yule, 1975, p. 185). When Rutter and Yule compared the "underachievers" to the children whose low reading performance was commensurate with their equally low IQ scores (i.e., "low achievers"), they found that the underachievers were different "in terms of sex distribution, neurological disorder, and pattern of neuro-developmental

deficit" (p. 194). Further, the underachieving readers had a worse prognosis for reading and spelling and a better prognosis for mathematics. These findings led Rutter and Yule to suggest that the group of underachievers, or children with "specific reading retardation," was distinctly different from the group of low achievers, or "generally backward readers." Findings appeared to confirm "unexpected" and "specific" learning failure as a valid marker of students with LD (Fletcher, 1995).

And yet, clinicians' and researchers' affirmation of the LD construct, and the general buoyancy of the advocates, belied longstanding concerns. For example, as described by Hallahan and Mercer (2001), the question of prevalence had been a point of contention since the early 20th century when Hinshelwood argued that fewer than 1 in 1,000 students might have "word blindness," or reading disabilities, and Orton countered that a more accurate ratio was 1 in 10. (As indicated below, some policymakers today suggest a 1-in-4 prevalence rate for reading disabilities.)

An obvious reason for such disparate estimates is that there has never been agreement on an LD definition. For a century, the field has tried unsuccessfully to invoke the central nervous system to explain the disorder. Hinshelwood, for example, required that a diagnosis of word blindness be associated with obvious pathology. Orton dismissed this criterion, noting the impossibility of distinguishing pathological from nonpathological cases (see Hallahan & Mercer, 2001). In the early 1960s, the federal government and Easter Seals cosponsored several task forces on LD, the first two of which focused on definitional issues. Task Force I, composed mostly of medical professionals, defined LD in terms of minimal brain dysfunction. The education professionals who constituted Task Force II rejected this definition "because special educators in the field of learning disabilities must base educational management and teaching strategies on functional diagnostic information" (Haring & Bateman, cited in Hallahan & Mercer, p. 34). This task force's substitute definition proposed in part that "Children with learning disabilities are those (1) who have educationally significant discrepancies among their sensory-motor, perceptual, cognitive, academic, or related developmental levels which interfere with the performance of educational tasks; (2) who may or may not show demonstrable deviation in central nervous system functioning; and (3) whose disabilities are not secondary to general mental retardation, sensory deprivation, or serious emotional disturbance" (Haring & Bateman, cited in Hallahan & Mercer). With minor modification, this language became part of the U.S. Office of Education definition in 1977.

Because the definition did not include criteria by which practitioners could identify children with LD, the federal government proposed regulations to operationalize it. The government's strategy was to suggest a "severe discrepancy" between intelligence and achievement as the primary criterion, or marker, for identification. In 1977, the government wrote that educators may determine that children have a specific learning disability if they receive appropriate learning experiences for their age and ability and still do not achieve commensurate with their age or ability levels in oral expression, listening comprehension, written expression, basic reading skill, reading comprehension, mathematics education, or mathematics reasoning.

This guideline, however, was viewed as insufficient by many state education agencies. Most adopted the federal government's severe discrepancy idea (Frankenberger & Fronzaglio, 1991; Mercer, King-Sears, & Mercer, 1990), but defined it in their own way. In 1983, the federal government tried again by convening the Work Group on Measurement Issues in the Assessment of LD. Its primary mission was to determine "[w]hat constitutes a severe discrepancy, from a statistical perspective between aptitude and achievement" (Mastropieri, 1987, p. 29). The Work Group found that (a) states had indeed adopted many varieties of measurement formulas for identifying a severe discrepancy, and (b) some of these formulas were not only excessively complex but mathematically incorrect. The Work Group recommended that practitioners regress an aptitude measure on an achievement measure to produce a predicted achievement score, and that "discrepancy" should be defined as the difference between actual and predicted achievement.

However, even this effort was criticized sharply. "The...discrepancy model," wrote Willson (1987), "is basically an atheoretical, psychologically uninformed solution to the problem of LD classification. For LD to move forward...statistical models...need to be replaced by constructs firmly grounded in psychological theories of learning" (p. 28; also see Lyon, 1987).

1980s: The Politicization of LD

These concerns percolated more or less quietly in the U.S. Office of Education, state education agencies, and academe until the 1980s. Then two things happened to cause a much greater number of educators and politicians to question the validity of the LD construct.

Special Education's Soaring Enrollments and Cost

Between 1977 and 1994, the number of students with disabilities increased from 3.7 million to 5.3 million "despite... [the fact] that overall public school enrollment [remained] roughly constant over this period" (Hanushek, Kain, & Rivkin, 2001, p. 7). These numbers represented an increase from 8.3 to 12.2% of the general student population. Virtually all of the growth came from increases in students classified as LD, a group that grew from 22 to 46% of all special-needs children over this period (Hanushek et al., p. 7). Hanushek and Rivkin (1997, cited in Hanushek et al.) suggested that "special education accounted for roughly 20% of the increase in per student spending during the 1980s, slightly less than double the share of special education students" (p. 7). These developments did not escape the attention of school boards, school superintendents, politicians, and other stakeholders in public education, some of whom began calling for an immediate downsizing of special education (e.g., Viadero, 1991).

The Regular Education Initiative

Another event dramatizing and deepening LD concerns was the Regular Education Initiative (REI), a reform movement bold and comprehensive in design. One of its founders was Madeleine Will who, in the 1980s, was assistant secretary of Education in the Office of Special Education and Rehabilitation Services. She was also the mother of a son with Down's syndrome and, more than anything else, she wanted to dramatically increase the number of children with disabilities in regular schools and classrooms. In 1986, she circulated a paper entitled "Educating children with learning problems: A shared responsibility," which became a manifesto of sorts of the movement. Will and other REI supporters (notably her friend, Margaret Wang, and Wang's colleagues, Maynard Reynolds and Herbert Walberg) were critical of what they perceived as special education's empire-building and profligate spending (e.g., Wang & Walberg, 1988); its use of putatively stigmatizing labels such as "mental retardation" and "behavior disorders" (e.g., Reynolds, Wang, & Walberg, 1987); and its separation of special-needs children from nondisabled peers, which they characterized as undemocratic (e.g., Wang & Walberg) and racist because of the overrepresentation of children of color in many resource rooms and self-contained classes (e.g., Lipsky & Gartner, 1989; Stainback & Stainback, 1988). Moreover, REI supporters claimed little was "special" (e.g., Sleeter, 1998; Spear-Swerling & Sternberg, 1998) or effective (e.g., Biklen & Zollers, 1986; Gartner & Lipsky, 1989; Wang & Walberg) about special education instruction.

The REI's distinctiveness, however, was not its litany of complaints against special education, or its goal of accelerating the mainstreaming of special-needs children. Instead, its noteworthiness was its broader aim of transforming general education into a more instructionally responsive system capable of accommodating a large majority of children with disabilities and thereby reducing the size and cost of special education. The Adaptive Learning Environments Model (ALEM; e.g., Wang & Birch, 1984), cooperative learning (e.g., Stevens, Madden, Slavin, & Farnish, 1987), reciprocal teaching (Palincsar & Brown, 1984), and other instructional programs designed for mainstream classrooms were advanced as proven means to such an ambitious end. REI advocates believed that the remaking of general education would require (a) massive professional development, which, they assumed, would be financed by the dollars saved from downsizing special education, and (b) a redefinition of the role of special educators, away from direct service and toward "collaborative consultation" and "coteaching" alongside classroom teachers. Both the professional development and the new roles for special educators would require a major reconfiguration of separate administrative systems (i.e., general education, special education, Title I, and English as a second language) into a "unified" system (see McLaughlin & Warren, 1992).

Will and many other REI backers viewed children with LD as most appropriate among all students with

disabilities for placement in transformed mainstream classrooms. There were at least two reasons for this. First, students with LD were understood by many to represent the mildest form of disability and, hence, they were seen as having the best chance of making it in the mainstream. Second, students with LD occupied the resource and self-contained classrooms coveted by advocates of children with mental retardation. In other words, some REI supporters reasoned that, if students with LD were mainstreamed, then many children with mental retardation could move from special schools to the more normal settings vacated by the children with LD. When LD advocates expressed skepticism about regular education's willingness and ability to accommodate the unique learning needs of many students with disabilities—a central assumption of REI supporters—a vigorous debate ensued (e.g., Bryan & Bryan, 1988; Fuchs & Fuchs, 1988a, 1988b; Hallahan, Keller, McKinney, Lloyd, & Bryan, 1988; Kauffman, 1989; Lloyd, Repp, & Singh, 1991; Reynolds, 1988; Wang & Walberg, 1988). This debate, in turn, further politicized the LD construct, as well as a good portion of LD research, which, deliberately or otherwise, contributed to the growing perception that LD was an invalid category of exceptionality.

LD Research

At least three lines of research in the 1980s addressed the LD construct. The first documented considerable variation of LD definitions and operationalizations across states (e.g., Gerber & Semmel, 1984; Mercer et al., 1990). Definitions differed in many ways: (a) the operationalization of discrepancy (e.g., standard scores for IQ minus standard scores for achievement vs. the regression of IQ on achievement); (b) the size of the discrepancy (e.g., 1.0 *SD* vs. 2.0 *SDs*); and (c) the choice of IQ and achievement tests. A popular and provocative way of expressing the findings from this work was to say something like "a child qualifying as LD in one state very well may have been excluded from the category in a neighboring state because of varying state regulations" (see Gerber & Semmel, 1984). In certain cases, some argued, the LD designation hinged more on the school district than the state in which one resided because of differences in regulations between districts within the same state (e.g., Peterson & Shinn, 1997).

Second, related work showed that many teachers purposely disregarded definitional rules and regulations to ensure special education for their students (e.g., Gottlieb, Alter, Gottlieb, & Wishner, 1994; MacMillan, Gresham, & Bocian, 1998; MacMillan, Gresham, Siperstein, & Bocian, 1996; Shepard & Smith, 1983). Gottlieb et al., for example, randomly selected 175 children with LD from six school districts and 165 elementary and middle schools in a large metropolitan area. Ninety percent of this group received some form of public assistance. The mean IQ of the sample was 81.4 (*SD* = 13.9). Students with LD in resource classes had higher IQ scores (*M* = 86.6) than those in self-contained classes (*M* = 75.0). Such scores, as well as teacher interviews, indicated that "children...classified as learning disabled...exhibit[ed] a generalized failure in their academic work rather than specific inefficiencies of cognitive processes or deficiencies in circumscribed academic subject matter" (p. 458). Only 15% met conventional identification criteria. Gottlieb et al. wrote, "Were the significant discrepancy feature of the learning disability definition observed, it would be extremely difficult to obtain with IQ scores so low" (p. 458).

"Why," asked Gottlieb et al. (1994), "is the severe discrepancy component of the definition so frequently ignored by school professionals?" They responded, "Our...discussions...with urban practitioners suggest that discrepancies are knowingly ignored...to marshall...resources for low-achieving (LA) students. Assessment staff and decision makers acknowledge that much of the school failure exhibited by children is more likely attributed to the effects of poverty...than to a 'learning disability' as defined in state regulations. Nevertheless...an educational fiction is agreed [on] to provide eligibility for special education services and programs. The current state of urban education, so woefully underfunded relative to its needs, provides students little access to intensive resources outside special education" (p. 459).

A third area of research in the 1980s reported considerable overlap in performance on various aptitude and educational tests between low achievers with and without the LD label (e.g., Ysseldyke, Algozzine, Shinn, & McGue, 1982) and between students with LD and Title I (high-poverty) students (e.g., Jenkins, Pious, & Peterson, 1988). Ysseldyke and colleagues (1982), and others, concluded from this work that virtually no important educational difference existed between students with LD and garden-variety poor achievers, that

LD was an "oversophistication" of the concept of low achievement (e.g., Algozzine, 1985; Algozzine & Ysseldyke, 1983). Such a claim motivated others to conduct similar research, a point to which we will return.

In aggregate, the research on the variability of state definitions of LD, teachers' disregard for these definitions, and the overlap in performance between low achievers with and without the LD label promoted a widely held view that an LD designation was essentially arbitrary (see Coles, 1987; Doris, 1993; Finlan, 1994; Klatt, 1991; Pugach, 1988; Reynolds, 1991; Reynolds, Wang, & Walberg, 1987; Skrtic, 1991; Sleeter, 1986; Wang, Reynolds, & Walberg, 1994/1995; Ysseldyke, Algozzine, & Epps, 1983; Ysseldyke, Algozzine, Richey, & Graden, 1982; Ysseldyke, Algozzine, Shinn, & McGue, 1982). Writing in 1983, following completion of a program of research associated with their federally funded, 5-year Institute for Research on Learning Disabilities, Ysseldyke and colleagues spoke for many when they wrote, "After five years of trying, we cannot describe, except with considerable lack of precision, students called LD. We think that LD can best be defined as 'whatever society wants it to be, needs it to be, or will let it be' at any point in time. We think [LD] researchers have compiled an interesting set of findings on a group of students who are experiencing academic difficulties, who bother their regular classroom teachers and who have been classified by societally sanctioned labelers in order to remove them, to the extent possible, from the regular education mainstream" (Ysseldyke et al., 1983, cited in Hallahan & Mercer, 2001, p. 50).

Summary

REI advocates, therefore, promoted the idea that all children with an LD label were simply low achievers. They further claimed that because so-called children with LD did not have unique learning needs, they, together with their nondisabled peers, could profit from the ALEM, cooperative learning, and other presumably proven instructional programs for the mainstream. To ensure all low-performing children's academic well-being, REI supporters argued that special educators should be retooled as consultants and coteachers. In these new roles, they would spend much of their time in the mainstream. The ALEM, cooperative learning, and similar programs, combined with in-class support, would in turn permit the responsible decertification of children with LD and a dramatic downsizing of special education. Despite such innovative ideas to strengthen mainstream classrooms and shrink special education, REI supporters could neither convince a critical mass of general educators to support their strategies (see Pugach & Sapon-Shevin, 1987) nor persuade important stakeholders in the disability community that general education would be willing and able to respond appropriately to the unique learning needs of students with disabilities.

1990S: THE NICHD GROUP

In the 1990s, an "NICHD group" became the most important voice expressing dissatisfaction with current LD definitions and encouraging fundamental change in our thinking about LD. The "NICHD group" refers in part to the principal investigators of learning disabilities centers funded by the National Institute of Child Health and Development (NICHD) and to Reid Lyon, the branch chief who supervises and coordinates their work. Our descriptor for this group is admittedly imprecise because we include researchers without NICHD funding who have conducted similar research, and we have little reason to believe everyone in the group thinks alike on all issues. Nevertheless, the group's work is sufficiently cohesive and important for us to regard it as an entity, and sufficiently tied to NICHD for us to use the acronym as an adjective. This group is different from the REI advocates in many ways. The NICHD group consists mostly of developmental, experimental, clinical, and neuropsychologists; REI advocates were largely special educators. The NICHD group focuses mostly on reading disabilities; REI supporters focused on the broad range of learning disabilities. Perhaps the most important difference between the groups is that the NICHD group claims to recognize the legitimacy of the LD construct. Lyon et al. (2001) have written, "Few would disagree that 5% or more of our school-age population experience difficulties with language and other skills that would be disruptive to academic achievement. The concept of LD is valid" (p. 7). What is invalid, says the NICHD group, are the definitions and operationalizations of the construct, which, they insist, must be reconceptualized.

The Argument Against IQ Discrepancy As a Valid LD Marker

Many concerns exist about IQ discrepancy as a definition or operationalization of LD, several of which already have been described. In addition, there are statistical problems (e.g., Willson, 1987) and the rejection by many of IQ as a meaningful estimate of overall intellectual potential (e.g., Spear-Swerling & Sternberg, 1998). The NICHD group's principal interest in the IQ discrepancy has been to explore the criterion validity of the "two-group hypothesis"; that is, the belief that qualitative differences exist between (a) children whose poor reading is discrepant from their IQ and (b) children whose poor reading is not discrepant from their IQ. This work has included a review of earlier studies conducted by others, secondary analyses of extant work, and the implementation of primary research.

"Earlier studies," wrote Fletcher (1995), "provided at best equivocal evidence for the validity of the two-group hypothesis. Many studies yielded null results, while other studies observed small but statistically significant differences between the two groups" (p. 16). Fletcher dismissed this earlier research with the claim that it suffers from methodological weaknesses that compromise its findings. In recent years, the NICHD group has conducted at least four studies on IQ discrepancy and the validity of the two-group hypothesis (Fletcher et al., 1994; Foorman, Francis, & Fletcher, 1995; Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Stanovich & Siegel, 1994). Two of these studies (Fletcher et al. and Francis et al.) used extant data from Shaywitz and colleagues' Connecticut Longitudinal Study (Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). The four studies are noteworthy for (a) their researcher-defined samples, (b) their broad selection of concurrent child measures, and (c) the systematic way in which the studies build on each other. Each of these studies fails to support the two-group hypothesis. Moreover, authors of these studies have argued that many related investigations have demonstrated that "IQ scores do not predict who is able to benefit from remediation" (Siegel, 1999, p. 312). Or, as Fletcher put it, "There is no evidence that low-IQ and high IQ poor readers respond differently to treatment" (p. 41). Hence, the NICHD group contends that predictive and concurrent validity studies indicate that poor readers characterized by an IQ-achievement discrepancy are no different from poor readers without this discrepancy in terms of most reading-related skills.

The Argument for Phonological Deficits As a Valid LD Marker

Share, McGee, and Silva (1991, cited in Fletcher, 1995) have written, "Professional preoccupation with IQ...is liable to obscure those significant advances achieved over the last 15 years in the field of reading research…to identify domain-specific factors...that are more potent than all-purpose measures, such as IQ. More importantly, these domain specific factors go much further than IQ in helping us understand and deal with reading failure" (p. 43). One important domain-specific factor, according to the NICHD group, is phonological processing. According to Fletcher (1995), Morris et al. (1998), Share et al. (1991), Siegel (1989, 1999), Stanovich (1999), Stanovich, Siegel, and Gottardo (1997), Torgesen, Morgan and Davis (1992), Vellutino et al. (1996), Wagner et al. (1997), and others, phonological processing figures prominently among the information-processing operations that are believed to underlie severe problems in word recognition. The NICHD group claims that phonological deficit should be recognized as a valid LD marker.

Further, they estimate about 25% of the student population demonstrates phonological deficits and argue that all of these children should be understood as reading disabled: "We have chosen to combine the [reading disabled] designation with children who (a) meet criteria for LD and typically receive services through special education; and (b) read below the 25th percentile but who do not qualify for the diagnosis of LD and often receive services through compensatory education [because] data [indicate] little difference between the two groups in the proximal causes of their reading difficulties" (Lyon et al., 2001, pp. 3–4).

Twenty-five percent of 50 million school-going children equals 12.5 million students. Together with approximately 3 million special-needs children who are not LD, this recommendation produces a population of students with disabilities of about 15.5 million, more than 2.5 times the number currently served under the Individuals with Disabilities Education Act (IDEA). With 12.5 million children identified

as LD, one might expect strong political pressure to reconceptualize this large group in terms of "nondisabled students in need of more effective general instruction."

The key to more effective instruction, says the NICHD group, is early identification and prevention. In his March 8, 2001, testimony before the U.S. House Subcommittee on Education and the Workforce, Lyon asked, "Can children with reading problems overcome their difficulties?" He answered, "Yes, the majority...can learn to read at average or above levels, but only if they are identified early and provided with systematic, explicit, and intensive instruction in phonemic awareness, phonics, reading fluency, vocabulary, and reading comprehension strategies" (p. 3). Moreover, he informed Congress, "Sufficient data exist to guide the development and implementation of early identification and prevention programs for children at-risk for LD" (p. 2). Early identification and prevention, says the NICHD group, should occur in general education: "Given that the underlying causes of most reading difficulties are similar for children regardless of whether they are currently served in special or compensatory education programs, we argue that the most valid and efficient way to deliver this early intervention in reading is as part of regular education" (Lyon et al., 2001, p. 20). However, a "major problem with such efforts is that special educators who typically provide instruction to children with LD have not been integrated into the early identification and prevention initiatives. It is important that both regular and special education embrace these efforts..." (Lyon et al., 2001, p. 36).

Similarities Between the NICHD and REI Groups

Many of the NICHD group's views about and policy recommendations for children with LD (and, more generally, for special education) are strikingly similar to those advanced by REI adherents in the 1980s. Both groups (a) are critical of special education effectiveness; (b) recommend that special education dollars should be combined with Title I dollars and possibly other funding streams to support the professional development necessary to strengthen general education's capacity to accommodate all low achievers; (c) argue that, with this accomplished, many special-needs children will be in mainstream classrooms, thereby permitting a reduction in the size and cost of special education nationwide; and (d) promote the notion that special educators' roles must change. REI adherents argued that special educators should become consultants and coteachers; the NICHD group recommends that they become heavily involved in early identification and prevention.

Most important, both groups view LA children with and without the LD label as the same children. For the NICHD group, "all low achievers are LD"; for the REI group, "all children with LD are low achievers." Although the NICHD group claims to believe in the validity of the LD construct, its critique of LD definitions and operationalizations seems to raise fundamental questions about the category. For example, Lyon et al. (2001) ask rhetorically, "Is the definition of LD that guides assessment and diagnostic practices too general and ambiguous to ensure accurate identification of younger students? Are the constructs and principles inherent in the definition of LD [invalid]? Are the diagnostic practices biased against the identification of younger, poor, or ethnically different children with LD?" (p. 7). To each question, they answer "yes." Similarly, Fletcher (1995) has written: "We have shown that the two-factor [poor readers with and without the LD label] classification implicit in the Federal Register definition lacks validity" (pp. 45–46). The NICHD group's dismissal of the distinction between low achievers with and without the LD label would seem to encourage a reconceptualization of children with LD—a subsumption of these children into a much larger, nondisabled group (i.e., the 12.5 million poor readers who, according to the NICHD group, require early identification and intensive prevention in general education). Hence, statements of support for the LD construct notwithstanding, the NICHD group, like the REI group before it, appears to be questioning whether the LD category deserves continued support.

META-ANALYSIS

Need for a Meta-Analysis

Since Morgan's and Hinshelwood's pioneering work at the turn of the last century, there has been

disagreement about the nature of LD. In the past two decades, as its prevalence and the associated costs of special education to local and state governments have escalated, these discussions have taken on a high-stakes tone. Many interested parties are now openly questioning the meaningfulness (and usefulness) of the LD construct. Researchers have played an important role in this discourse. Using *researcher-identified* samples, the NICHD group has repeatedly demonstrated that poor readers with and without an IQ-achievement discrepancy have more in common (e.g., phonological deficits) than not. On this basis, the NICHD group and others argue that the IQ-achievement discrepancy should *not* be a criterion in LD identification.

According to Gottlieb, MacMillan, Shepard, and Ysseldyke and their respective colleagues, however, many school districts deliberately disregard discrepancy information. In contrast to the NICHD group's research, Ysseldyke and his associates used *practitioner-identified* samples to explore whether low achievers with and without the label are different from each other. Across a series of studies, they reported no educationally important differences between the two groups. This provocative claim inspired many others to try to replicate their work. Findings have been inconsistent, and for good reason: Investigators have explored different performance domains (e.g., reading achievement vs. classroom behavior); chosen dissimilar measures within a given domain (e.g., reading comprehension vs. phonemic awareness); used contrasting definitions of LD (e.g., IQ greater than or equal to 90 vs. IQ greater than or equal to 70) and low achievement (e.g., teacher judgment vs. cutoff scores); involved demographically different student groups (e.g., low vs. middle vs. high socioeconomic status; urban vs. suburban vs. rural); and based their statistical comparisons on different metrics (e.g., degree of overlap vs. mean performance). Bottom line: There is no consensus as to whether the two groups of low achievers—those whom the schools have labeled and those who remain unlabeled—are distinguishable.

If a comprehensive review of the empirical evidence shows that students with the LD label cannot be distinguished from their LA, nonlabeled classmates, then it would seem only reasonable to support the abolition of this disability category. After all, the logical alternative would be to declare all LA students learning disabled, an assertion that we believe would make little economic, political, or legal sense. On the other hand, if a systematic review of research shows that the school-identified LD group performs more poorly, in both a statistically significant and educationally meaningful sense, then we can assume that the two groups represent different populations of students. Such a result may lend weight to the view that students with the LD label have different educational needs, in degree or kind, which might be addressed only within special education (e.g., Mather & Roberts, 1994; National Joint Committee on Learning Disabilities, 1994).

With these and other questions in mind, we have identified and quantitatively synthesized the extant literature in the domain of reading. We have chosen this domain for several reasons. First, a majority of studies comparing LA students with and without the LD label focus on reading. Second, most children with LD are identified as such because of chronic reading problems. Third, reading difficulty strongly affects overall school achievement (e.g., Stanovich, 1986).

In searching the scientific literature on reading, we coded each study that met our inclusion criteria and we analyzed the resulting data base. In the following sections, we summarize these methods and our results. We provide detailed information on the development of our coding system. For a thorough description of the literature search and data analysis, see Fuchs, Fuchs, Mathes, Lipsey, and Roberts (2001).

METHOD

Inclusion Criteria and Search Strategies

Our goal was to identify all published and unpublished studies in which the reading achievement of LD and LA nondisabled students could be compared. A study was defined by its participants: If two or more studies were conducted on the same students, the studies were counted as one. In a similar way, a single article could report more than one study if it included different samples of students with LD.

For inclusion, a study had to meet five criteria:

1. It had to present reading data.
2. Those data had to be reported separately for LD and LA groups.
3. Whenever the LD group included a mixture of students with high-incidence disabilities, students with LD had to constitute at least 85% of the group.
4. Participants had to be school age (i.e., kindergarten through grade 12).
5. The study had to report data necessary for calculating effect sizes (ESs).

To identify studies that met these criteria, we undertook a comprehensive search of journal articles, Educational Resources Information Center (ERIC) documents, and dissertations in Dissertations Abstracts International (DAI) produced between January 1975 and December 1996. This search comprised three phases: a manual search of journals, two computerized database searches (ERIC and DAI), and an ancestral search of titles in the references of identified investigations. Eighty-six studies met our inclusion criteria.

Coding the Studies

To systematically derive information from the studies, we developed a coding form in two phases. As we initially read the studies, it was unclear which study characteristics would eventually prove worthy of coding. Therefore, in Phase 1, we described many study features, knowing some would later be discarded. We began by reading a considerable portion of the research and becoming familiar with the typical range of study features described. We then developed a first-draft coding form with which we independently coded a sample study—Shinn, Ysseldyke, Deno, and Tindal (1986). After debriefing, we developed a second draft and accompanying code book. Then, we independently coded four studies, including Shinn et al. (1986) for a second time. After coding each study, we again discussed each item on the coding form. Throughout this process, definitions of codes were refined and decision rules about handling ambiguous situations were determined.

At this point, we began coding studies. However, within a couple of weeks, unacceptably low levels of interrater agreement indicated a need for more precise definitions, so the coding form was revised again. As a result, 30 articles that had already been coded with the second draft had to be recoded. A 16-page coding form emerged from Phase 1 (contact the first author for the final coding form). Using this iteration of the form, five studies were coded with interrater agreement of 90% or better on each study.

Then, the remaining journal articles and ERIC documents were coded independently. During this coding process, to check whether the raters were continuing to code in the same way, they completed independently the same set of 13 studies. Agreement on each exceeded 85%.

Recognizing the temptation to make reasonable inferences about information not clearly presented in studies, we instituted a no-guessing rule: If uncertainty arose about how to code an item, it was left blank. Later, an author determined the code. If questions still remained, the codes were discussed until consensus was achieved.

Approaching data entry, it became apparent that the 16-page coding form was too detailed; it contained codes inappropriate or irrelevant for many studies. Therefore, in Phase 2, the form was reduced to 45 codes that would be entered into the computer. During this scaling-down process, we added one code, "reading," which was redefined by various subdomains (e.g., phonological awareness, lexical retrieval, reading readiness).

The final coding form differed in appearance from the 16-page version because it was briefer and designed to match the computer spreadsheet. So, for example, both coding form and spreadsheet now displayed one line of data for every reading measure in a study.

Selected study codes were then transferred from the 16-page coding form to the briefer, final form. Before

beginning this process, two coders independently transferred the codes of five studies from one form to the other, immediately checking accuracy. One coder then transferred the codes of all previously coded studies to the final coding form. An independent coder then checked this transfer of codes for every study.

Codes for 86 studies were entered into an electronic spreadsheet. To ensure accuracy, two checkers examined the spreadsheet item by item. As one person read the data base entry, the second person checked the information on the coding form.

Computation of Individual ESs

Typically, ES was computed as the standardized mean difference (d index): the difference between the means of the comparison groups divided by the pooled standard deviation (Hedges & Olkin, 1985). This formula represents LD-LA differences scaled in the uniform metric of standard deviation units. A positive ES reflects higher performance by the LA group. As recommended by Hedges (1981), this ES formula was adapted to yield an unbiased estimate of the underlying population effect. Whereas a majority of studies presented the information necessary to compute ES using the basic formula, some studies presented other comparison statistics. In such cases, ES was estimated from those other statistics.

Aggregation of ES Within Studies

We aggregated two or more ESs in the same study, if those ESs were identical on eight variables: reading subdomain, research design, sample size for LD, sample size for LA, grade level, and IQ (Full Scale, Verbal, and Performance). Thus, any two ESs in the same study that did not match exactly on these eight dimensions were judged to be independent, with one important exception. In a few instances, subgroups of a sample differed in size, but were identical with respect to the remaining seven variables. In these cases, ESs associated with these subsamples were eliminated. Also eliminated at this point were seven studies in which LD and LA students were matched on reading achievement or reading achievement and IQ. ESs from the remaining 79 studies were included in the meta-analysis.

Preliminary Analyses

We undertook four preliminary analyses to formulate decisions about which data, in what form, should be incorporated into the major analyses. First, we examined the effect of four types of study designs: (a) descriptive/one point in time, (b) descriptive/change over time, (c) intervention/posttest only, and (d) intervention/change over time. We decided to conduct analyses on only one type, which had the vast majority of ESs: the descriptive/one-point-in-time studies ($n = 202$).

Second, we examined whether and if so how to consolidate data across the reading subdomains. We found that five reading domains (decoding isolated words, reading connected text, reading comprehension, overall reading, and vocabulary) yielded ES values sufficiently similar, as indexed by their central tendencies, to be considered comparable. However, the remaining domains (phonological awareness, rapid automatized naming, and reading readiness) were comparable neither with the other five domains listed previously nor with each other. The mean covariate-adjusted ESs for these three domains, respectively, were 0.05, 0.26, and –0.40. Thus, we did not combine these three domains with the remaining five domains or with each other. This left 172 ESs.

Third, with this smaller data base, we identified independent samples that contributed more than one ES. These records were aggregated by averaging all variables (except the reading subdomain). Because all other variable values in the averaged records were identical, a single record was produced for each independent sample. This resulted in a data file of 112 records, each representing an independent sample with an ES in one of the five reading subdomains or a mean ES averaged over two or more of the five subdomains.

Finally, the distribution of the 112 ESs revealed outliers at both ends. To reduce the possibly distorting effect of these outliers, we windsorized them. Two ESs less than –1.00 were increased to –1.00; five ESs

greater than 1.75 were reduced to 1.75. Doing so had a minimal effect on the overall mean ES.

RESULTS

Are the ESs Homogeneous?

Our first major analysis indicated considerable disagreement among the studies with respect to the magnitude of the differences between LD and LA groups in reading performance: The Q statistic indicated substantially greater variability among ESs than would be expected from sampling error alone, $Q(111) = 535.75$, $p < 0.001$. This finding led us to explore which study characteristics might be associated with variation in ES.

How Might We Consolidate the Large Number of Study Features?

Before examining the relation between study features and ESs, we consolidated some study features. First, based on analyses we conducted, we consolidated our definitions of LD and LA samples to five levels of LD/LA definitional pairings; this resulted in 109 ESs.

Second, we conducted several focused factor analyses on sets of variables that seemed to be related conceptually and were better represented as multivariate composites. A varimax-rotated solution seemed to fit these variables nicely. We thereby reduced the LD-LA student comparability data to three factors: achievement, which incorporated variables related to reading comparability; demographic characteristics, including age, race, and socioeconomic status (SES); and gender comparability, IQ, and SES comparability. We refer to these three factors as (a) achievement comparability, (b) demographic comparability, and (c) gender comparability, respectively.

Finally, we conducted another factor analysis to examine relations among variables describing the research method used for constructing the LD-LA samples. This analysis produced a sensible two-factor solution. The first factor showed a co-occurrence of the following: (a) lower IQ scores, (b) higher grade levels for the LD sample, and (c) referral of LA samples for special education testing. We called this factor "other sample features." The second factor cleanly combined the two variables describing whether the samples were district or researcher identified. We called this factor "identification source."

How Do the Clustered Study Features Relate to ESs? We used weighted least-squares regression, weighting each ES by the inverse of its variance. Our pool of predictor variables included the LD/LA definitional pairings, the five factor scores (reading comparability, demographic comparability, gender comparability, other sample features, and identification source), the three locale variables, technical adequacy, test format, study quality, and date of study. Predictors were entered simultaneously; then, the weakest was dropped and the model was refit. We repeated this process until all remaining variables were significant.

The regression model accounted for a statistically significant 41% of the variance among the ESs. The following variables made significant, independent contributions to the prediction of ES.

First, measurement format contributed to the prediction of ES, with a beta of 0.34. ESs were greater for the timed than the untimed measurement formats. For example, on tests requiring students to work in a fixed time (such as the Stanford Achievement Test or curriculum-based measurement), the difference between students with and without LD was larger than when tests permitted students as much time as they needed (e.g., Woodcock Reading Mastery Tests). This was true across reading domains.

Second, other sample features contributed to the prediction of ES, with a beta of 0.16. ESs were greater for LD samples with lower IQ and with higher LD grade; ESs were greater when LAs had been referred but had never qualified as appropriate for special education.

Third, LD/LA definitional pairings contributed to the prediction of ES. ESs were greater when LD samples

were defined by discrepancies and when LA samples were defined by teacher judgment; the associated beta was 0.51. ESs were smaller when LD samples were identified by multidisciplinary team judgment and when LA samples were defined by data-driven methods; the associated beta was –0.27.

Fourth, the three comparability factors contributed to the prediction of ES. ESs were greater when achievement and demographics were not comparable for LD and LA samples; the associated beta was 0.13. ESs were greater when gender and, to a lesser extent, IQ and SES were comparable; the corresponding beta was 0.08.

Finally, methodological study quality contributed to the prediction of ES, with somewhat greater ESs for lower quality studies. The associated beta was 0.12.

WHAT DOES THIS META-ANALYSIS TELL US?

Across the many substantive and methodological variables associated with studies in this meta-analysis, ESs demonstrated considerable heterogeneity. Analyses were, however, successful in identifying a large proportion of the variance among ESs. Ten variables operated independently to explain the variation. In particular, three variables maximized the degree of reading impairment associated with the LD label and, therefore, provide insight into the theoretical nature of the disability. They also may help practitioners and researchers develop more effective assessment and intervention procedures for students with LD, as well as more precise measures of treatment success.

On the basis of these meta-analytic findings, we offer three conclusions, which may guide future research and practice. First, across the many different ways in which students become identified as LD, results leave no doubt that these students' reading achievement differs dramatically from other LA, nondisabled students. Averaged across all the methodological and substantive variations in the studies, the mean effect size was 0.61 standard deviations units. This effect is sizable; it means than 72% of the LA population performs better in reading than the mean of the LD population. Moreover, regarding ESs for timed measurements, whereby students were required to perform (i.e., read aloud, read silently, answer questions, match words to meanings) within a fixed time, the ESs increased to well beyond one full standard deviation unit. And, in a similar way, when LD and LA samples had been identified using data-based methods, the overall ES of 0.61 rose to beyond a full standard deviation unit. Findings, therefore, suggest that researchers and school personnel in fact do identify as LD those children who have appreciably more severe reading problems compared to other low-performing students who go unidentified. As with any comparison of two populations, some overlap between these populations occurs; that overlap, however, is not sufficient to call the LD label into question. Consequently, in light of the more severe magnitude of LD students' reading problems, it seems reasonable and desirable that more intensive forms of reading instruction be directed at this group of students.

Second, the ESs associated with timed tests were larger than those associated with untimed tests. The beta associated with this effect was an impressive 0.34. This strong effect associated with timed measurement format suggests theoretical and practical implications. Failure at achieving automaticity may represent an important characteristic of students with LD, which may be associated with the low performance on rapid-naming tasks (Wolf, 1991) of many of these children. The possibility that difficulties in achieving automaticity may represent a key feature of students with LD warrants additional study. Methods of identifying LD children might incorporate timed reading assessments to focus on students' failure to achieve automatic word-reading performance. In addition, with respect to treatment, researchers should develop methods for helping students with LD transition from accurate to automatic word reading. Finally, results suggest that the effectiveness of interventions for students with LD should be evaluated at least in part by how they influence students' performance on timed reading measurements.

Finally, results underscore the importance of objective measurement of reading performance in the identification process. Larger differences between LD and LA students emerged when definitional and selection criteria for inclusion to studies relied on objective forms of reading measurement—that is, the

administration of tests. By contrast, when individual or team judgment was involved, differences between LD and LA samples on reading measures grew smaller. On one hand, this finding provides a basis for questioning human judgment in the identification process. On the other hand, it suggests that other considerations, such as a focus on social behavior, may play a viable role in the identification of children whose overall performance profiles warrant special treatment. Practitioners should be mindful of the advantages and disadvantages associated with reliance on nonobjective forms of input to the multidisciplinary team process. Future research should continue to identify which types of nonobjective data may be important in the identification process and should continue to examine the role of social behavior deficits and the possibility of comorbidity in children with LD.

REFERENCES

Algozzine, B. (1985). Low achiever differentiation: Where's the beef? *Exceptional Children, 52*, 72–75.

Algozzine, B., & Ysseldyke, J. E. (1983). Learning disabilities as a subset of school failure: The oversophistication of a concept. *Exceptional Children, 50*, 242–246.

Biklen, D., & Zollers, N. (1986). The focus of advocacy in the LD field. *Journal of Learning Disabilities, 19*, 579–586.

Bryan, J. H., & Bryan, T. H. (1988, April). *Where's the beef?* Paper presented at the annual meeting of The Council for Exceptional Children, Washington, DC.

Coles, G. (1987). *The learning mystique: A critical look at "learning disabilities."* New York: Pantheon.

Doris, J. L. (1993). Defining learning disabilities: A history of the search for consensus. In G. R. Lyon, D. B. Gray, J. F. Kavanagh, & N. A. Krasnegor (Ed.), *Better understanding learning disabilities: New views from research and their implications for education and public policies* (pp. 97–115). Baltimore: Brookes.

Finlan, T. G. (1994). *Learning disability: The imaginary disease*. Westport, CT: Bergin & Garvey.

Fletcher, J. M. (1995, May). *Diagnostic utility of intelligence testing and the discrepancy model for children with learning disabilities: Historical perspectives and current research.* Paper presented at the IQ Testing and Educational Decision Making Workshop, National Research Council, National Academy of Sciences.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., Fowler, A. E., & Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*(1), 6–23.

Foorman, B. R., Francis, D. J., & Fletcher, J. M. (1995, March). *Growth of phonological processing skill in beginning reading: The lag versus deficit model revisited.* Paper presented at the Society for Research on Child Development, Indianapolis, IN.

Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology, 88*(1), 3–17.

Frankenberger, W., & Fronzaglio, K. (1991). A review of states' criteria for identifying children with learning disabilities. *Journal of Learning Disabilities, 24*, 495–500.

Fuchs, D., & Fuchs, L.S. (1988a). Evaluation of the Adaptive Learning Environments Model. *Exceptional Children, 55*(2), 115–127.

Fuchs, D., & Fuchs, L.S. (1988b). Response to Wang and Walberg. *Exceptional Children, 55*(2), 138–146.

Fuchs, D., Fuchs, L. S., Mathes, P. G., Lipsey, M. W., & Roberts, P. H. (2001). Is "learning disabilities" just a fancy term for low achievement? A meta-analysis of reading differences between low achievers with and without the label. Paper presented at the Office of Special Education

Program's LD Initiative Conference, August 27–29, Washington, DC.

Gartner, A., & Lipsky, D. K. (1989). *The yoke of special education: How to break it*. Rochester, NY: National Center on Education and the Economy. (ERIC Document Reproduction Service No. ED 307 792)

Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*, 137–148.

Gottlieb, J., Alter, M., Gottlieb, B. W., & Wishner, J. (1994). Special education in urban American: It's not justifiable for many. *The Journal of Special Education, 27*, 453–465.

Hallahan, D. P., Keller, C. E., McKinney, J. D., Lloyd, J. W., & Bryan, T. (1988). Examining the research base of the regular education initiative: Efficacy studies and the ALEM. *Journal of Learning Disabilities, 21*, 29–35.

Hallahan, D. P., & Mercer, C. D. (2001). *Learning Disabilities: Historical Perspectives*. Washington, DC: Office of Special Education Programs, U.S. Department of Education.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2001, March). *Inferring program effects for specialized populations: Does special education raise achievement for students with disabilities?* Unpublished manuscript.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 490–499.

Hedges, L. V., & Olkin, I. (1985). *Methods of meta-analysis*. Newbury Park, CA: Sage.

Jenkins, J. R., Pious, C. G., & Peterson, D. L. (1988). Categorical programs for remedial and handicapped students: Issues of validity. *Exceptional Children, 55*, 147–158.

Kauffman, J. M. (1989). The Regular Education Initiative as Reagan-Bush education policy: A *trickle*-down theory of education of the hard to teach. *The Journal of Special Education, 23*, 256–278.

Kavale, K. A. (1987). Theoretical issues surrounding severe discrepancy. *Learning Disabilities Research, 3*(1), 12–20.

Kavale, K. A., & Forness, S. R. (1998). The politics of learning disabilities. *Learning Disability Quarterly, 21*(4), 245–273.

Klatt, H. J. (1991). Learning disabilities: A questionable construct. *Educational Theory, 41*, 47–60.

Lipsky, D. K., & Gartner, A. (1989). *Beyond separate education: Quality education for all*. Baltimore: Paul Brookes.

Lloyd, J. W., Repp, A. C., & Singh, N. N. (Eds.), (1991). *The regular education initiative: Alternative perspectives on concepts, issues, and models*. Sycamore, IL: Sycamore.

Lyon, G. R. (1987). Severe discrepancy: Theoretical, psychometric, developmental, and educational issues. *Learning Disabilities Research, 3*(1), 10–11.

Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., Schulte, A., & Olson, R. (2001). *Rethinking Learning Disabilities*. Hudson Institute.

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

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

Mastropieri, M. A. (1987). Statistical and psychometric issues surrounding severe discrepancy: A discussion. *Learning Disabilities Research, 3*(1), 29–31.

Mather, N., & Roberts, R. (1994). Learning disabilities: A field in danger of extinction? *Learning Disabilities Research and Practice, 9*, 49–58.

McLaughlin, M., & Warren, S. (1992, September). *Issues and options in restructuring schools and special education programs*. College Park, MD: The Center for Policy Options in Special Education.

(ERIC Document Reproduction Service No. ED 350 774)

Mercer, C. D., King-Sears, P., & Mercer, A. R. (1990). Learning disabilities definitions and criteria used by state education departments. *Learning Disability Quarterly, 13*, 141–152.

Morris, R. D., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Stuebing, K. K., Fletcher, J. M., Lyon, R. G., Francis, D. J., & Shaywitz, B. A. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*, 347–373.

National Joint Committee on Learning Disabilities. (1994). A reaction to "full inclusion": A reaffirmation of the right of students with learning disabilities to a continuum of services. In *Collective Perspectives on Issues Affecting Learning Disabilities: Position Papers and Statements.* Austin, TX: Pro-Ed.

Palincsar, A. M., & Brown, A. L. (1984). Reciprocal teaching of comprehension monitoring activities. *Cognition and Instruction, 2*, 117–175.

Peterson, K., & Shinn, M. R. (1997). *An examination of competing models of learning disabilities identification practice: The role of achievement context in school-based decision making.* Unpublished paper.

Pugach, M. (1988). Special education categories as constraints on the reform of teacher education. *Journal of Teacher Education, 39*(3), 52–59.

Pugach, M., & Sapon-Shevin, M. (1987). New agendas for special education policy: What the national reports haven't said. *Exceptional Children, 53*, 295–299.

Reynolds, M. C. (1988). Reaction to the JLD special series on the Regular Education Initiative. *Journal of Learning Disabilities, 21*, 352–356.

Reynolds, M. C. (1991). Classification and labeling. In J. W. Lloyd, A. C. Repp, & N. N. Singh (Eds.), *The regular education initiative: Alternative perspectives on concepts, issues, and models* (pp. 29–41). Sycamore, IL: Sycamore.

Reynolds, M. C., Wang, M. C., & Walberg, H. J. (1987). The necessary restructuring of special and regular education. *Exceptional Children, 53*, 391–398.

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

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading disability. *The New England Journal of Medicine, 326*, 145–150.

Shepard, L., & Smith, M. L. (1983). An evaluation of the identification of learning disabled students in Colorado. *Learning Disability Quarterly, 6*, 115–127.

Shinn, M. R., Ysseldyke, J. E., Deno, S. L., & Tindal, G. A. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of classroom performance. *Journal of Learning Disabilities, 19*, 545–552.

Siegel, L. S. (1989). I.Q. is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469–478.

Siegel, L. S. (1999). Issues in the definition and diagnosis of learning disabilities: A perspective on Guckenberger v. Boston University. *Journal of Learning Disabilities, 32*, 304–319.

Skrtic, T. M. (1991). *Behind special education: A critical analysis of professional culture and school organization.* Denver: Love.

Sleeter, C. E. (1986). Learning disabilities: The social construction of a special education category. *Exceptional Children, 53*, 46–54.

Sleeter, C. E. (1998). Yes, learning disabilities is political; what isn't? *Learning Disability Quarterly, 21*, 289–296.

Spear-Swerling, L., & Sternberg, R. J. (1998, January). Curing our "epidemic" of learning disabilities. *Phi Delta Kappan*, 397–401.

Stainback, S., & Stainback, W. (1988). Letter to the editor. *Journal of Learning Disabilities, 21*,

452–453.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360–406.

Stanovich, K. E. (1999). The sociopsychometrics of learning disabilities. *Journal of Learning Disabilities, 32*, 350–361.

Stanovich, K. E., & Siegel, L. S. (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.

Stanovich, K. E., Siegel, L. S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89*, 114–127.

Stevens, R. J., Madden, N. A., Slavin, R. E., & Farnish, A. M. (1987). Cooperative integrated reading and composition: Two field experiments. *Reading Research Quarterly, 22*, 433–454.

Torgesen, J. K., Morgan, S. T., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology, 84*, 364–370.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, S., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*(4), 601–638.

Viadero, D. (1991, June 12). States turn to special education programs for budget cuts. *Education Week*, p. 16.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donohue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A five-year longitudinal study. *Developmental Psychology, 33*, 468–479.

Wang, M. C., & Birch, J. W. (1984). Comparison of a full-time mainstreaming program and a resource room approach. *Exceptional Children, 51*, 33–40.

Wang, M. C., Reynolds, M. C., & Walberg, H. J. (1994/1995). Serving students at the margins. *Educational Leadership, 52*, 12–17.

Wang, M. C., & Walberg, H. J. (1988). Four fallacies of segregationism. *Exceptional Children, 55*, 128–137.

Willson, V. L. (1987). Statistical and psychometric issues surrounding severe discrepancy. *Learning Disabilities Research, 3*(1), 24–28.

Wolf, M. (1991). Naming speed and reading: The contribution of the cognitive neurosciences. *Reading Research Quarterly, 26*, 123–141.

Ysseldyke, J. E., Algozzine, B., Richey, L., & Graden, J. (1982). Declaring students eligible for disability services: Why bother with the data? *Learning Disability Quarterly, 5*(1), 37–43.

Ysseldyke, J. E., Algozzine, B., Shinn, M. R., & McGue, M. (1982). Similarities and differences between low achievers and students classified learning disabled. *The Journal of Special Education, 16*, 73–85.