Responsiveness-to-Intervention Symposium

December 4-5, 2003 • Kansas City, Missouri

The National Research Center on Learning Disabilities, a collaborative project of staff at Vanderbilt University and the University of Kansas, sponsored this two-day symposium focusing on responsiveness-to-intervention (RTI) issues.

The symposium was made possible by the support of the U.S. Department of Education Office of Special Education Programs. Renee Bradley, Project Officer. Opinions expressed herein are those of the authors and do not necessarily represent the position of the U.S. Department of Education.

Validity of Alternative Approaches to the Identification of L.D.: Operationalizing Unexpected Underachievement

Jack M. Fletcher

Department of Pediatrics

Center for Academic and Reading Skills

University of Texas Health Science Center at Houston

Carolyn Denton

Department of Special Education

University of Texas – Austin

When citing materials presented during the symposium, please use the following: "Fletcher, J. M. and Denton, C. (2003, December). *Validity of Alternative Approaches to the Identification of L.D.: Operationalizing Unexpected Underachievement.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO."

Validity of Alternative Approaches to the Identification of L.D.: Operationalizing Unexpected Underachievement

This paper addresses the validity of alternative models to the identification of learning disabilities (LD). In preparing this paper, we assumed that the "alternatives" involved any approach other than a discrepancy between a composite IQ test and achievement. Variations on four different approaches were considered: (1) models involving other forms of discrepancy; (2) low achievement models; (3) intra-individual differences models; and (4) models that incorporate response to intervention.

Two sets of assumptions were made in evaluating these models. The first assumptions were that the construct of LD operates as a latent variable that is imperfectly measured by these different models (Fletcher & Morris, 1986; Francis et al., in press). The key aspect of the construct is the notion of "unexpected underachievement," representing children and adults who should be able to learn, yet don't attain levels that would be expected based on the absence of obstacles to learning, including adequate opportunities to learn. Thus, a key aspect of evaluating validity is determining how well the approach to identification produces a unique group of poor achievers.

The second group of assumptions involves the assessment of validity. We assume that identification methods stem from an overarching classification that indicates how persons with LD are similar to and different from those with other childhood disorders, other groups of poor achievers, and typically achieving children (Fletcher et al., 1993). A valid classification will reflect definitions and measurements that operationalize the construct of unexpected underachievement by showing that the group identified as LD differ from other subgroups *on variables not used to create the classification* (Morris & Fletcher, 1988; Skinner, 1981). Obviously, people defined as LD will differ on variables used to create the classification, unless it is completely unreliable. The extent to which differences emerge may reflect the correlation of the external variables with those used to create the classification (Stuebing et al., 2002), which usually involve some form of cognitive assessment, so it is important to look beyond comparisons of cognitive skills whenever possible. In many instances, validity is limited because of problems with reliability (Francis et al., in press), but the focus of this paper is on the external validity of the classifications that lead to identification.

Other forms of discrepancy

The problems with models based on aptitude-achievement discrepancies are well known. Two major meta-analyses have shown that effect sizes on measures of achievement and cognitive functions are in the negligible to small range (at best) for the comparison of groups formed on the basis of discrepancies between a composite (Full Scale) IQ and reading achievement versus groups of poor readers without an IQ-discrepancy (Hoskyn & Swanson, 2000; Stuebing et al., 2002). Moreover, other

Fletcher, J. M. and Denton, C. (2003, December). *Validity of Alternative Approaches to the Identification of L.D.: Operationalizing Unexpected Underachievement.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO.

validity studies have not found that groups of poor readers formed on the basis of the presence and absence of composite IQ-achievement discrepancies differ in long-term prognosis (Francis et al., 1996; Silva et al., 1985), response to intervention (see Fletcher et al., 2001; Jiménez et al., 2003; Stage et al., 2003), or neuroimaging correlates (see Lyon, Fletcher, & Barnes, 2003; but also see Shaywitz et al., 2003, which shows differences in groups varying in IQ, but not due to IQ-discrepancy). Studies of genetic variability show negligible to small differences related to IQ-discrepancy models that may reflect regression to the mean (Pennington et al., 1992; Wadsworth et al., 2001). Despite the evidence showing weak validity when comparisons of underachievers based on IQ-discrepancy versus poor reading with no discrepancy (low achievers), alternatives based on discrepancy models continue to be proposed. Finally, similar empirical evidence can be cited for LD in math and language (Fletcher et al., 2002; Mazzocco & Myers, in press), which is not surprising given the problems with the underlying psychometric model.

The problems with the underlying psychometric model have been systematically outlined since IQ-discrepancy was first put into federal regulations in 1975 (Christensen, 1992). Most recently, Francis et al. (in press) showed that alternative models based on discrepancy are not viable, reflecting the underlying psychometric difficulties with any discrepancy model. However, there is also empirical evidence pertaining to validity. In the Stuebing et al. (2002) meta-analysis, 32 of the 46 studies had a clearly defined aptitude measure. Of these studies, 19 used Full Scale IQ, eight used Verbal IQ, four used Performance IQ, and one study used a discrepancy of listening comprehension and reading comprehension. In addition, Fletcher et al. (1994) and a study not eligible for inclusion in Stuebing et al. (Stanovich & Siegel, 1994) both evaluated the validity of different discrepancy models by systematically manipulating Full Scale IQ, Verbal IQ, and Performance IQ. Fletcher et al. (1994) also included a listening comprehension-reading comprehension discrepancy model, which continues to be proposed as an alternative identification model for children with reading disabilities (Joshi, in press).

Not surprisingly, these different operationalizations of discrepancy models did not yield results that were different from those apparent when a composite IQ measure is utilized. As Fletcher et al. (1994) showed, the effect of using different aptitude measures in a regression discrepancy model is to simply shift the slope upwards or downwards depending on the correlation of the aptitude and achievement measures. Thus, for example, using Verbal IQ produces a much steeper regression line than Performance IQ because Verbal IQ is much more highly correlated (.70) with reading outcomes than Performance IQ (.58). This has the effect of shifting individuals who are at the edges of the regression cutpoint on one aptitude measure to either a discrepancy or low achievement subgroup when a different aptitude measure is employed. In a sample where there is control for variation in the cutpoint used to identify people with LD, the overlap in terms of who is identified as an underachiever is substantial, usually exceeding 80% (Fletcher et al., 1994). However, exactly who becomes discrepant or low achieving will depend on the slope of the regression line. As the changes in identification reflect fluctuations around the regression cut off, it is not surprising that

effect sizes comparing poor readers with and without IQ-discrepancies are uniformly low across these different models. In Fletcher et al. (1994), the use of cognitive tests that were not part of the definition produced effect sizes that were generally below .10. Stanovich and Siegel (1994) found some differences on cognitive measures outside the language domain based on the use of Verbal IQ versus Performance IQ, although the magnitude of differences fluctuated. Neither Fletcher et al. (1994) nor Aaron et al. (1988) were able to demonstrate major differences between discrepancy and low achievement groups formed on the basis of listening comprehension-reading comprehension.

The real problem is the idea that a discrepancy model will produce differences between children with different forms of underachievement. A discrepancy model cannot possibly produce a clearly unique set of underachievers. None of the eight studies in Stuebing et al. (2002) that systematically explored Verbal IQ found any differences on measures closely related to reading, such as phonological awareness, findings also recently reported by Stage, Abbott, Jenkins, and Berninger (2003). As Francis et al. (in press) have shown, the reliability of classification models based on any form of discrepancy is not adequate to produce acceptable validity. Altogether, models based on aptitude-achievement discrepancies do not appear to identify a unique group of underachievers, and therefore do not adequately operationalize the construct of LD.

Low achievement models

Models based on the use of achievement markers can be shown to have a great deal of validity (see Fletcher et al., 2002; 2003; Siegel, 1992). In this respect, if groups are formed such that it is ensured that the participants do not meet criteria for mental retardation and have achievement scores that are below the 25th percentile, a variety of comparisons show that subgroups of underachievers emerge that can be validly differentiated on external variables and help demonstrate the viability of the construct of LD. For example, if children with reading and math disabilities are compared to typical achievers, it is possible to show that these three groups display different cognitive correlates. In addition, neurobiological studies show that these groups differ both in the neural correlates on reading and math tasks as well as the heritability of reading and math disorders (Lyon et al., 2002). These achievement subgroups, which by definition include children who meet either low achievement or IQ-discrepancy criteria, even differ in response to intervention, providing strong evidence for "aptitude by treatment" interactions: obviously we don't provide math interventions for children with only reading problems, and vice versa.

Despite this evidence for validity, many squirm when definitions based solely on achievement cutpoints are proposed, and for good reason. Simply utilizing a low achievement definition, even when different exclusionary criteria are applied, doesn't really operationalize the true meaning of unexpected underachievement. Although such an approach to identification is deceptively simple, it is arguable whether the subgroups that remain represent a unique group of underachievers. For example, how well are underachievers whose underachievement is attributed to LD differentiated from underachievers whose underachievement is attributed to emotional disturbance, economic disadvantage, or inadequate instruction (Lyon et al., 2001)? To use the example of word recognition, there

is little evidence that these subgroups vary in terms of phonological awareness or other language tasks, response to intervention, or even neuroimaging correlates. In this respect, the validity is weak. Finally, as Francis et al. (in press) demonstrated, there are still major issues with the reliability of low achievement identification models, particularly when they are based on a single assessment.

Intra-individual differences models

A commonly proposed alternative to models based on aptitude-achievement discrepancies and/or low achievement involves an examination of individual differences on measures of cognitive function. Thus, for example, a recent consensus paper from 10 major advocacy groups organized by the National Center for Learning Disabilities (NCLD, 2002) stated that "while IQ tests do not measure or predict a student's response to instruction, measures of neuropsychological functioning and information processing could be included in evaluation protocols in ways that document the areas of strength and vulnerability needed to make informed decisions about eligibility for services, or more importantly, what services are needed. An essential characteristic of SLD is failure to achieve at a level of expected performance based upon the student's other abilities (p. 4)."

This statement proposes intra-individual differences as a marker for "unexpected underachievement." As opposed to a single marker like IQ-discrepancy or low achievement, unexpectedness is operationalized as unevenness in scores across multiple tests. The person identified as LD (by definition) has strengths in many areas of cognitive or neuropsychological function, but weaknesses in core attributes that lead to underachievement. LD is unexpected because the weaknesses lead to selected and narrow difficulties with achievement and adaptive functions. Proponents of this view believe that such approaches identify children as LD based on profiles across tests that differentiate types of LD, and also differentiate LD from other childhood disorders, such as mental retardation and behavioral disorders like attention-deficit/hyperactivity disorder (ADHD). This approach leads to definitions based on *inclusionary* criteria in which children are identified as LD based on characteristics that relate to intra-individual differences (Lyon et al., 2002).

A major assumption of this model is that identification based on performance patterns will lead to enhanced treatment of children with LD. The weakness of the model is the absence of evidence that strengths and weaknesses in processing skills are related to intervention outcomes. It is well-established that training in underlying processes does not usually generalize into the related academic area. For example, training on phonological awareness skills without a letter component produces gains in phonological awareness, but not in reading (National Reading Panel, 2000). There is also the issue of perpetuating identification models that have not resulted in better outcomes for children with LD, which could be termed "test and treat" models. The result could be even more testing of children for eligibility purposes, reinforcing the model currently in place. The scaling issues for this model are significant. If administration of IQ and achievement tests as part of a search for a two-test discrepancy is not well implemented in many schools (MacMillan & Siperstein, 2002), how well can a multi-test discrepancy model be implemented? It is difficult to fathom teaching this type of clinic model to school psychologists and educational

diagnosticians across the country. Given the access to children in schools, why would a model often based on what is often a single exposure to a child in a lifetime be applied when the child could be seen many times within a school year as part of a longitudinal evaluation of progress? Finally, this approach to identification does not address the problem of children with relatively flat test profiles. Severity is correlated with unevenness due to the lack of independence of different tests that might be used to construct the profile (Morris et al., 1993). For example, children with either a severe reading or phonological awareness problem will show increasingly flat profiles in direct correspondence to severity. Thus, if the criterion is evidence of a discrepancy in neuropsychological or processing skills, such an approach may not exclude the most severely impaired children, irrespective of global measures like IQ, because more severely impaired children are less likely to show skill discrepancies due to the intercorrelation of the tests (Morris et al., 1993; 1998).

By definition, the intra- individual differences approach will produce unique subgroups of underachievers who vary in cognitive skills. Stanovich's (1988) phonological core- variable differences model, supported by the subtyping study of Morris et al. (1998), and Rourke's (1989) nonverbal learning disability model, are strong examples of the intra-individual differences model. One issue is independence on measures not used to form the classification. Keep in mind that Stanovich's model indicates that within the reading domain, language differences aren't expected. Rourke's model includes comparisons with children who have reading and math difficulties, showing characteristics of both NLD and verbal learning problems. The critical question is how such approaches lead to better outcomes for children with LD. The intra-individual differences model focuses on behaviors that are not directly related to intervention, such as processing skills (Torgesen, 2002). However, interactions beyond the interactions beyond the primary area of *academic* difficulty (word recognition, fluency, comprehension, and math) are hard to identify (Fletcher, Morris, & Lyon, 2003). Thus, the model has the most validity at the level of achievement markers, but simply collapses into a low achievement model in the absence of processing measures.

Models incorporating response to intervention

All of the approaches reviewed so far are based on assessments administered at a single timepoint. They become unwieldy and impractical if extended to multiple assessments. As Francis et al. (in press) showed, it would be difficult to implement this kind of model with any great reliability when a single time point is used. To certain extent, the intra-individual differences approach avoids some of these difficulties by using multiple tests at the same timepoint, looking for recurrent discrepancies that might make up a profile. However, the measures used in this approach typically have reliabilities that are much lower than those apparent for norm referenced IQ and achievement tests, magnifying the problem of reliably identifying profile variations. In contrast, models incorporating response to intervention typically involve identification based in part on multiple probe assessments of the same core area, such as reading or math. By tying multiple assessments to specific attempts to intervene with the child, the construct of unexpected underachievement can be operationalized in

part on the basis as non-responsiveness to instruction to which most other students respond (Fuchs & Fuchs, 1998; Gresham, 2002).

Such approaches do not obviate the measurement issues involved in the assessment of discrepancy (Fletcher et al., 2003). In fact, discrepancy is still part of the model, but is assessed relative to learning expectations based on multiple administrations of the same test over time as opposed to a comparison of two tests, or multiple different tests administered at the same timepoint. There are issues involved in the intervention component, estimation of slope and intercept effects, as well as decisions that have to be made about "cutpoints" that will differentiate responders and non-responders (Gresham, 2002). For these reasons alone, response to intervention cannot be the sole criterion for identification and flexibility in decision making will be required. At the same time, there appears to be considerable validity to this approach, implying that it is indeed possible to reliably identify non-responders as a group with "unexpected underachievement." Studies of children defined using different methods as responders and non-responders clearly show large differences in cognitive skills. For example, Stage et al. (2003), Vellutino et al. (1996) and Vaughn, Linan-Thompson, and Hickman-Davis (in press) found that non-responders to early intervention differed from responders in both pre-intervention achievement scores and pre-intervention cognitive tasks. Those who were non-responders were usually more severe. In our imaging studies involving both early intervention and remediation of older students (see Fletcher et al., in press), we have found that individuals who were non-responders showed more severe reading difficulties prior to intervention. More dramatic were the differences in neuroimaging correlates between those who responded to intervention and those who did not respond to intervention. We have found that non-responders persist with a brain activation pattern that generally demonstrated a failure to activate left hemisphere areas known to be involved in the development of reading skills. In fact, those who were non-responders showed predominant right hemisphere activity much like that observed in children and adults with identified reading disabilities (Fletcher et al., in press).

In addition to this evidence for validity (and the greater reliability of the underlying psychometric model), the model does not require the use of exclusionary criteria (especially emotional disturbance and economic disadvantage) to operationalize unexpected underachievement, thus capturing the construct of LD. This is an important consideration given the lack of evidence validating classifications that utilize these particular exclusions (Kavale, 1988; Lyon et al., 2001).

Conclusions

In evaluating these data, some will say that we should not accept the null hypothesis and propose continued investigations of discrepancy, low achievement, and intra-individual difference models. There is little possibility that discrepancy models can yield unique subgroups of underachievers. Low achievement models do not adequately assess the construct of unexpected underachievement. Intra-individual differences models are essentially an expansion of discrepancy models, maintaining the historic emphasis on the "LD test battery." It is time for change. Such approaches have not been effective, and there are psychometric

and logistical issues that make all these models difficult to implement.

Others will note that the amount of validity evidence that we are citing for response to intervention models is relatively sparse. We have not attempted to score the literature for validity studies. Nonetheless, models that incorporate response to intervention have considerable potential for identifying a unique group of underachievers. The studies that we can cite show *robust* differences in cognitive and achievement functions and in neural correlates. There is also evidence that different kinds of interventions are needed for non-responders, which is admittedly a somewhat circular argument. Most important, the potential for powerful demonstrations of validity and broadly understandable support for the construct of LD should be immediately apparent. In the end, measurement must assess the latent construct of LD. Previous classifications have been adopted without evidence that even the underlying measurement model was valid. No model will ever be perfect. Models that incorporate response to intervention also imperfectly assess the LD construct and we can debate the best procedures for operationalizing non-responsiveness: type of intervention, slope, intercept, number of assessment points, etc. (Fuchs, Mock, Morgan, & Young, in press). There will be no best procedure, but the potential validity is transparent and will likely transcend these measurement issues

It is useful to conceptualize models that incorporate response to intervention as extensions of low achievement models. The construct of LD begins with underachievement and the task is to define "unexpectedness." A hybrid model in which low achievement is used as one of several criteria (along with the exclusions) and response to intervention has considerable potential and was recommended by the LD Summit (Bradley et al., 2002). In this respect, no single assessment should ever be the sole the criterion for establishing eligibility for special education. It is one thing in a research study to specify identification criteria and still another to make a complex judgment about eligibility and services. However, previously utilized classification models do not have adequate reliability or strong validity. These models incorporating response to intervention at least have the promise of operationalizing the concept of unexpected underachievement in a way that is meaningful: as a demonstrable inability to learn. The children who are identified with this classification do appear to differ from children who have other forms of underachievement. Critics of these models worry that the result of implementation efforts will be an abandonment of historic concepts of discrepancy and a loss of integrity of the LD construct (Kavale, 2002; Scruggs & Mastropieri, 2002). Neither view is correct as the model retains the concepts of unexpected underachievement and discrepancy (Fletcher et al., 2003). In fact, if the field does not move away from historic concepts that were adopted with a severely limited research base, the construct of LD will perish under the weight of the research base not supporting the reliability or validity of alternative approaches to discrepancy and the lack of effectiveness of the services associated with current approaches to identification. More research is always needed, but the critical research issue is an evaluation of what happens when hybrid models are implemented on a wide scale. Without an attempt at implementation, it will be hard to determine the viability of this model.

References

Aaron, P.G., Kuchta, S., & Grapenthin, C.T. (1988). Is there a thing called dyslexia? *Annals of Dyslexia, 38*, 33-49.

Bradley, R., Danielson, R.L., & D.P. Hallahan, D.P. (2002). *Identification of learning disabilities: Research to practice*. Mahwah, NJ: Erlbaum.

Christensen, C.A. (1992). Discrepancy definitions of reading disability: Has the quest led us astray? *Reading Research Quarterly, 27*, 276-278.

Fletcher, J.M., & Morris, R. (1986). Classification of disabled learners: Beyond exclusionary definitions. In S. J. Cici (Ed.), *Handbook of cognitive, social, and neuropsychological aspects of learning disabilities* (pp. 55-80). Hillsdale, NJ: Erlbaum.

Fletcher, J.M., Shaywitz, S.E., Shankweiler, D.P., Katz, L., Liberman, I.Y., Stuebing, K.K., Francis, D.J., Fowler, A., & Shaywitz, B.A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 85*, 1-18.

Fletcher, J.M., Lyon, G.R., Barnes, M., Stuebing, K.K., Francis, D.J., Olson, R.K., Shaywitz, S.E., & Shaywitz, B.A. (2002). Classification of learning disabilities: An evidenced-based evaluation. In R. Bradley, L. Danielson, & D.P. Hallahan (Eds.). *Identification of learning disabilities: Research to practice* (pp. 185-250). Mahwah, NJ: Erlbaum.

Fletcher, J.M., Simos, P.G., Papanicolaou, A.C., & Denton, C. (in press). Neuroimaging in reading research. In N. Duke & M. Mallette (Eds.). *Literacy research methods*. New York: Guilford.

Fletcher, J.M., Morris, R.D., & Lyon, G.R. (2003). Classification and definition of learning disabilities: An integrative perspective. In H.L. Swanson, K.R. Harris, and S. Graham. *Handbook of learning disabilities* (pp. 30-56). Guilford Publications.

Fletcher, J.M., Francis, D.J., Rourke, B.P., Shaywitz, B.A., & Shaywitz, S.E. Classification of learning disabilities: Relationships with other childhood disorders. In G.R. Lyon, D. Gray, J. Kavanagh & N. Krasnegor (Eds.) *Better understanding learning disabilities.* New York: Paul H. Brookes, 1993, pp. 27-55.

Francis, D.J., Fletcher, J.M., Stuebing, K.K., Lyon, G.R., Shaywitz, B.A., & Shaywitz, S.E. (in press). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities.*

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*, 3-17.

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

Fuchs, L., & Fuchs, D. (1998). Treatment validity: A simplifying concept for reconceptualizing the identification of learning disabilities. Learning Disabilities Research and Practice, 4, 204-219

Fletcher, J. M. and Denton, C. (2003, December). *Validity of Alternative Approaches to the Identification of L.D.: Operationalizing Unexpected Underachievement.* Paper presented at the National Research Center on Learning Disabilities Responsiveness-to-Intervention Symposium, Kansas City, MO.

Gresham, F.M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D.P. Hallahan (Eds.). Identification of learning disabilities: Research to practice (pp 467-564). Mahwah, NJ: Erlbaum.

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. *The School Psychology Review, 29*, 102-119.

Jiménez, J.E., del Rosario Ortiz, M., Rodrigo, M., Hernández-Valle, I., Ramirez, G., Estévez, A., O'Shanahan, I., & de la Luz Trabaue, M. (2003). Do the effects of computer-assisted practice differ for children with reading disabilities with and without IQ-achievement discrepancy? *Journal of Learning Disabilities, 36,* 34-47.

Joshi, R.M. (in press). Diagnosis and remediation of dyslexia: A pragmatic solution. *Annals of Dyslexia.*

Kavale, K.A. (1988). Learning disability and cultural disadvantage: The case for a relationship. *Learning Disability Quarterly, 11*, 195- 210.

Kavale, K.A. (2002) Discrepancy models in the identification of learning disability. In R. Bradley, L. Danielson, & D.P. Hallahan (Eds.). Identification of learning disabilities: Research to practice (pp. 369-464). Mahwah, NJ: Erlbaum.

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. In C.E. Finn, Jr., RA.J. Rotherham, & C.R. Hokanson, Jr. (Eds.) *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.

Lyon, G.R., Fletcher, J.M., & Barnes, M.C. (2002). Learning disabilities. In E.J. Mash and R.A. Barkley (Eds.) *Child psychopathology* (pp. 520-586) (Second Edition). New York: Guilford.

MacMillan, D.L., & Siperstein, G.N. (2002). Learning disabilities as operationally defined by schools. In R. Bradley, L. Danielson, & D.P. Hallahan (Eds.). Identification of learning disabilities: Research to practice (pp. 287-367). Mahwah, NJ: Erlbaum.

Mazzocco, M.M.M., & Myers, G.F. (in press). Complexities in identifying and defining mathematics learning disability in the primary school age years. *Annals of Dyslexia.*

Morris, R., & Fletcher, J.M. (1988). Classification in neuropsychology: A theoretical framework and research paradigm. *Journal of Clinical and Experimental Neuropsychology, 10*, 640-658.

Morris, R.D., Fletcher, J.M., & Francis, D.J. Conceptual and psychometric issues in the neuropsychological assessment of children: Measurement of ability discrepancy and change. In I. Rapin & S. Segalovitz (eds.). Handbook of neuropsychology (Vol. 7). Amsterdam: Elsevier, 1993, 341-352.

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

National Center for Learning Disabilities (2002). *Achieving better outcomes – maintaining rights: An approach to*

identifying and serving students with specific learning disabilities. New York: Author.

National Reading Panel (2000). Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Washington, DC: National Institute of Child Health and Human Development.

Pennington, B.F., Gilger, J.W., Olson, R. K., & DeFries, J. C. (1992). The external validity of age- versus IQ-discrepancy definitions of reading disability: Lessons from a twin study. *Journal of Learning Disability, 25*, 562-573.

Rourke, B.P. (1989). *Nonverbal learning disabilities: The syndrome and the model*. New York: Guilford Press.

Scruggs, T.E., & Mastropieri, M.A. (2002). On babies and bathwater: Addressing the problems of identification of learning disabilities. *Learning Disability Quarterly*, *25*, 155- 168.

Silva, P.A., McGee, R., & Williams, S. (1985). Some characteristics of 9-year-old boys with general reading backwardness or specific reading retardation. *Journal of Child Psychology and Psychiatry, 26*, 407-421.

Shaywitz, S.E., Shaywitz, B.A., Fulbright, R.K., Skudlarski, P., Mencl, W.E., Constable,, R.T., Pugh, K.R., Holahan, J.M., Marchione, K.E., Fletcher, J.M., Lyon, G.R., & Gore, J.C. (2003). Neural systems for compensation and persistence: Young adult outcomes of childhood reading disability. *Biological Psychiatry, 54*, 25-33.

Skiner, H. (1981). Toward the integration of classification theory and methods. *Journal of Abnormal Psychology, 90*, 68-87.

Stage, S.A., Abbott, R.D., Jenkins, J.R., & Berninger, V.W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities, 36,* 24-33.

Stanovich, K.E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable difference model. *Journal of Learning Disabilities, 21*, 590-604.

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

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

Torgesen, J. K. (2002). The prevention of reading difficulties. *Journal of School Psychology, 40,* 7-26.

Torgesen, J.K. (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 565650).

Vaughn, S.R., Linan-Thompson, S., & Hickman-Davis, P. (in press). Response to treatment as a means of identifying students with reading/learning disabilities.

Vellutino, F.R., Scanlon, D. M., Sipay, E., Small, S., Pratt, A., Chen, R., and

Denckla, 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.

Wadsworth, S.J., Olson, R.K., Pennington, B.F., & DeFries, J.C. (2000). Differential genetic etiology of reading disability as a function of IQ. *Journal of Learning Disabilities, 33*, 192-199.