Fletcher, J. M., & Miciak, J. (2019). *The Identification of Specific Learning Disabilities: A Summary of Research on Best Practices*. Austin, TX: Meadows Center for Preventing Educational Risk.

Abstract

In this report, we summarize research on the identification of SLD and make recommendations for practice. The report begins with a summary of the legal requirements for SLD identification and what constitutes a comprehensive evaluation. It then discusses the attributes of SLD according to different conceptual frameworks and reviews research on the reliability and validity of different methods for SLD identification that emanate from these frameworks. The report concludes with recommendations for best practice, regardless of the specific identification methods employed.

This research was supported by grant P50 HD052117, Texas Center for Learning Disabilities, from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and grant R324B170012 from the National Center for Special Education Research (NCSER).

The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, the National Institutes of Health, NCSER, the Institute of Education Sciences, or the Texas Education Agency.

The Identification of Specific Learning Disabilities:

A Summary of Research on Best Practices

Jack M. Fletcher and Jeremy Miciak
University of Houston

TEXAS CENTER
for
LEARNING
DISABILITIES

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit
http://creativecommons.org/licenses/by-nc-nd/4.0

For inquiries about using this product outside the scope of this license, contact
licensing@meadowscenter.org

About the Texas Center for Learning Disabilities

The Texas Center for Learning Disabilities (TCLD) is a research center that investigates the classification, early intervention, and remediation of learning disabilities. Research activities are conducted at the University of Houston, The University of Texas at Austin, and The University of Texas Health Science Center at Houston. To learn more about TCLD, visit our website at **www.texasldcenter.org**

This research was supported by grant P50 HD052117, Texas Center for Learning Disabilities, from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and grant R324B170012 from the National Center for Special Education Research (NCSER).

The content is solely the responsibility of the authors and does not necessarily represent the official views of NICHD, the National Institutes of Health, NCSER, the Institute of Education Sciences, or the Texas Education Agency.

Address all correspondence to:

Jack M. Fletcher
University of Houston
4849 Calhoun Rd., Ste. 273
Houston, TX 77204-6022

Contents

Legal Requirements for SLD Identification 1

Overview of SLD 6

Frameworks for Understanding SLD 10

Validity Research on Cognitive Discrepancy Methods 12

Instructional Discrepancy (RTI) Approaches 19

Recommendations for SLD Identification 21

References 30

The Identification of Specific Learning Disabilities: A Summary of Research on Best Practices

Specific learning disability (SLD) is the most common eligibility category through which students receive special education services under the Individuals with Disabilities Education Act (IDEA). In Texas alone, more than 150,000 students received special education services in the 2016–2017 school year due to an identified SLD (Texas Education Agency, 2017). The rules and procedures by which students are identified with SLD affect an even greater number of students. As a result, the validity of these rules and procedures for identification must be considered as not just a legal and scientific question, but also as a question of fairness and access.

In this report, we summarize research on the identification of SLD and make recommendations for practice. The report begins with a summary of the legal requirements for SLD identification and what constitutes a comprehensive evaluation. It then discusses the attributes of SLD according to different conceptual frameworks and reviews research on the reliability and validity of different methods for SLD identification that emanate from these frameworks. The report concludes with recommendations for best practice, regardless of the specific identification methods employed.

Legal Requirements for SLD Identification

SLD in IDEA 2004

IDEA 2004 defines the term "specific learning disability" as "a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations." This definition has been unchanged since the initial legislation that is now IDEA 2004 was passed in 1975, Public Law 94-142. However, this consistency in the federal statute belies a significant shift in how SLD is conceptualized in federal regulatory guidance. When regulations for Public Law 94-142 were released in 1977, SLD was identified by a significant discrepancy between an intelligence quotient (IQ) and achievement (IQ-achievement discrepancy method). Additional criteria were designed to ensure that low achievement was not primarily due to another factor: sensory or motor disabilities, intellectual disabilities, emotional or behavioral disorders, economic disadvantage, cultural factors, or limited English proficiency. Each state then defined its own specific criteria for SLD identification following this guidance.

In Texas, a significant discrepancy was defined as a score in one of the seven academic domains of SLD (derived from the original statutory definition) that was 16 points lower than the student's IQ score. A comprehensive evaluation under this method included the administration of IQ and achievement tests, as well as data gathered by other procedures, such as through observations of the child and evaluation of the exclusionary factors. Consistent with current standards, the assessment procedures were determined by the Admission, Review, and Dismissal (ARD) team to be necessary for a determination of eligibility for special education services as a student with SLD.

The special education legislation has been reauthorized several times since 1975. Through the last revision, in IDEA 2004, the statutory definition of SLD has not changed. However, with each revision, there were changes in the regulations and guidance, although the use of IQ-achievement discrepancy criteria remained constant until the 2004 revision. The 2004 committee recommended movement away from

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

IQ-achievement discrepancy criteria because of a lack of evidence for the validity of such procedures. In addition, the commission recommended that states permit use of methods based on response to intervention (RTI). In IDEA 2004, Congress indicated that states (1) could not require districts to use IQ tests for the identification of students for special education in the SLD category and (2) had to permit districts to implement identification models that incorporated response to scientifically based instruction (IDEA 2004, 34 CFR §300.309). In addition, the statute indicated that children could not be identified for special education if poor achievement was due to lack of appropriate instruction in reading or math or due to limited proficiency in English.

IDEA 2004 Revised SLD Eligibility Criteria

A state must adopt…criteria for determining whether a child has a specific learning disability … In addition, the criteria adopted by the state:

- Must not require the use of a severe discrepancy between intellectual ability and achievement for determining whether a child has a specific learning disability …
- Must permit the use of a process based on the child's response to scientific, research-based intervention; and
- May permit the use of other alternative research-based procedures for determining whether a child has a specific learning disability…

—U.S. Department of Education, 2006, p. 46786

In its 2006 regulations, the Department of Education indicated that states must allow districts to make choices about procedures for SLD identification if the decisions were consistent with state rule-making processes. States had to permit the use of RTI criteria but did not have to make rules that permitted other methods of identification.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

2006 Regulations for IDEA 2004

(1) The child does not achieve adequately for the child's age or to meet State-approved grade-level standards in one or more of the following areas when provided with learning experiences and instruction appropriate for the child's age or State-approved grade-level standards:

(i) Oral expression.

(ii) Listening comprehension.

(iii) Written expression.

(iv) Basic reading skill.

(v) Reading fluency skills.

(vi) Reading comprehension.

(vii) Mathematics calculation.

(viii) Mathematics problem solving.

(2)

(i) The child does not make sufficient progress to meet age or State-approved grade-level standards in one or more of the areas identified in paragraph (a)(1) of this section when using a process based on the child's response to scientific, research-based intervention; or

(ii) The child exhibits a pattern of strengths and weaknesses in performance, achievement, or both, relative to age, State-approved grade-level standards, or intellectual development, that is determined by the group to be relevant to the identification of a specific learning disability, using appropriate assessments, consistent with §§ 300.304 and 300.305; and

(3) The group determines that its findings under paragraphs (a)(1) and (2) of this section are not primarily the result of

(i) A visual, hearing, or motor disability;

(ii) Mental retardation;

(iii) Emotional disturbance;

(iv) Cultural factors;

(v) Environmental or economic disadvantage; or

(vi) Limited English proficiency.

Texas Guidelines for SLD

Texas essentially adopted these criteria for SLD eligibility when it wrote its rules for special education, with some modifications. Consistent with the federal guidelines, the Texas guidelines require formal documentation of the provision of adequate general education instruction. The Texas guidelines are also flexible, permitting either a process based on RTI or one based on a discrepancy model. Under the latter, additional clarifications were provided on what constituted inadequate achievement and how to establish a discrepancy based on a pattern of strengths and weaknesses.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Texas Guidelines for SLD Identification

(9) Learning disability.

(A) Prior to and as part of the evaluation described in subparagraph (B) of this paragraph and 34 CFR, §§300.307-300.311, and in order to ensure that underachievement in a student suspected of having a specific learning disability is not due to lack of appropriate instruction in reading or mathematics, the following must be considered:

(i) data that demonstrates the student was provided appropriate instruction in reading (as described in 20 United States Code (USC), §6368(3)), and/or mathematics within general education settings delivered by qualified personnel; and

(ii) data-based documentation of repeated assessments of achievement at reasonable intervals, reflecting formal evaluation of student progress during instruction. Data-based documentation of repeated assessments may include, but is not limited to, response to intervention progress monitoring results, in-class tests on grade-level curriculum, or other regularly administered assessments. Intervals are considered reasonable if consistent with the assessment requirements of a student's specific instructional program.

(B) A student with a learning disability is one who:

(i) has been determined through a variety of assessment tools and strategies to meet the criteria for a specific learning disability as stated in 34 CFR, §300.8(c)(10), in accordance with the provisions in 34 CFR, §§300.307-300.311; and

(ii) does not achieve adequately for the student's age or meet state-approved grade-level standards in oral expression, listening comprehension, written expression, basic reading skill, reading fluency skills, reading comprehension, mathematics calculation, or mathematics problem solving when provided appropriate instruction, as indicated by performance on multiple measures such as in-class tests; grade average over time (e.g., six weeks, semester); norm- or criterion-referenced tests; statewide assessments; or a process based on the student's response to scientific, research-based intervention; and

(I) does not make sufficient progress when provided a process based on the student's response to scientific, research-based intervention (as defined in 20 USC, §7801(37)), as indicated by the student's performance relative to the performance of the student's peers on repeated, curriculum-based assessments of achievement at reasonable intervals, reflecting student progress during classroom instruction; or

(II) exhibits a pattern of strengths and weaknesses in performance, achievement, or both relative to age, grade-level standards, or intellectual ability, as indicated by significant variance among specific areas of cognitive function, such as working memory and verbal comprehension, or between specific areas of cognitive function and academic achievement.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

The Texas criteria parallel the federal regulations in providing a number of options for determining inadequate achievement. The criteria also specify ways of determining patterns of strengths and weaknesses, which can include "significant variance among specific areas of cognitive function, such as working memory and verbal comprehension, or between specific areas of cognitive function and academic achievement." In the federal regulations, like the Texas regulations, districts adopt either a framework based on an RTI service delivery model or a framework based on a discrepancy of achievement with age, grade-level standards, IQ, or a pattern of strengths and weaknesses in cognitive skills. Generally, both criteria should not be required because the frameworks are different and the application of both would result in more testing than is needed for SLD identification. The regulations explicitly indicate that a choice should be made between RTI and discrepancy methods, but both frameworks have common requirements, including ensuring the adequacy of general education instruction, requirements for a comprehensive evaluation, and interdisciplinary team decisions.

Comprehensive Evaluation

Regardless of the SLD identification framework a district chooses to implement, a comprehensive evaluation is required. IDEA 2004 specified eight criteria for a comprehensive evaluation. These criteria help to ensure that the evaluation addresses all possible explanations for the student's academic difficulties and that multiple criteria are met for identification. The regulations for IDEA 2004 defined a comprehensive evaluation as a "data-gathering process." The regulations also indicated that eligibility could not be established based on a single criterion, reflecting concern about some states' rigid use of formulae for SLD as the primary eligibility criterion and to reduce concerns that some districts would use only RTI as the primary criterion. The sidebar "Eight Components of a Comprehensive Evaluation" is a summary of these requirements; the specific federal guidelines should be consulted as authoritative. The criteria reviewed above for SLD identification are in addition to these criteria. Additionally, note that a comprehensive evaluation does not require standardized testing and that the use of a formula as the primary requirement for eligibility, such as a specific index of inadequate response to instruction, a fixed discrepancy threshold, or a fixed low-achievement threshold, would represent the use of a single criterion and would not meet requirements for a comprehensive evaluation.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Eight Components of a Comprehensive Evaluation

1. Use a variety of assessment tools and strategies to gather relevant functional, developmental, and academic information about the child, including information provided by the parent (comprehensive data-gathering process)
2. May not use any single measure or assessment as the sole criterion
3. Must use technically sound instruments that are
   - racially and culturally fair, administered in native language;
   - used for purposes for which they are reliable and valid;
   - administered as designed by trained and knowledgeable personnel; and
   - tailored to area of educational need, adapted to physical and sensory disabilities
4. The child is assessed in all areas related to the suspected disability (i.e., it's a data-gathering process)
5. The evaluation is coordinated with assessments of other [local education agencies] (e.g., when the student comes to a new school district with a previous evaluation and [individualized education program], these data must be considered)
6. The evaluation is sufficiently comprehensive to identify the child's special education and related service needs, whether or not commonly linked to the identified disability category (i.e., interventions may be provided that reflect the child's individual needs regardless of the eligibility category)
7. Assessment data directly assist persons in determining the educational needs of the child (e.g., IQ scores are composites and not indicators of intervention goals)
8. Additional requirements: Review existing relevant evaluations and data and determine what additional data are needed (e.g., formal testing may not be needed)

Overview of SLD

Classification Versus Identification

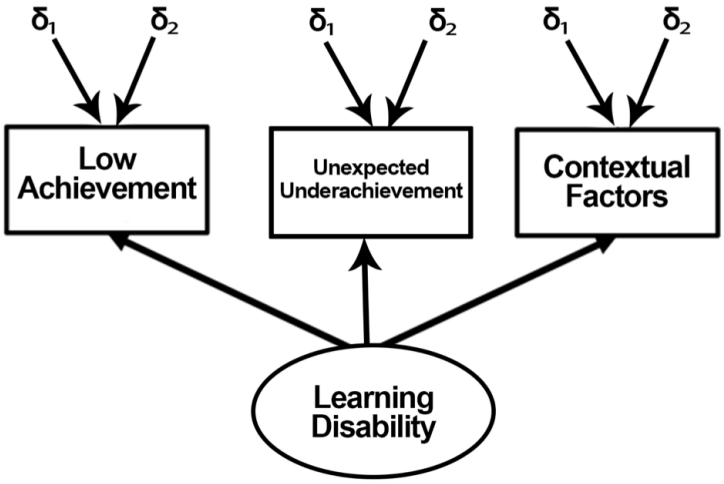
The methods used to implement a comprehensive evaluation often involve standardized tests, and the specific tests vary depending on the method chosen. The tests also vary in reliability and validity, but contemporary tests generally are well-constructed and strong in reliability. However, a small amount of unreliability inherent to a test can dramatically affect the reliability of SLD identification decisions for individual students. There are also varying degrees of validity for the identification methods themselves. There is no litmus test for SLD, a construct that can be known only by virtue of how it is measured. In this respect, SLD is like other constructs, such as IQ. IQ is known through measurement, and different theories lead to different approaches to measurement, yielding different assessments of IQ in an individual.

This issue is illustrated in Figure 1, which shows SLD as a construct indicated by how it is measured (δ

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

= a measured data point). As Figure 1 shows, we can implement different methods, which may include an evaluation of achievement, cognitive skills, and the exclusionary factors. The latter are contextual factors, like emotional functioning or economic disadvantage, or other disorders that may explain low achievement, like an intellectual disability or sensory disorder. For these indicators of SLD, our measures are inherently imperfect indicators not only because of the small degree of unreliability of the tests and observations, but also because of the validity of the underlying conceptual model of SLD.

Figure 1. The Measurement of SLD



To understand how to assess the reliability and validity of different methods for SLD identification, it is important to consider the relation of classification and identification. When different methods for identification are used, a set of attributes are proposed that are considered essential characteristics of the disorder. Among the attributes we discussed above are IQ, low achievement, cognitive discrepancies, and instructional response. These attributes are combined based on a theory of what does and does not constitute a particular disorder, like SLD. They then form a classification that specifies subgroups with and without the proposed characteristics of SLD. When we test a person for SLD, we give tests and consider other factors to identify members of the group with SLD. This process, **identification**, is a decision made about whether the individual meets criteria for SLD. The underlying classification must be valid for identification to be valid. In addition, the method must have adequate coverage. For example, if the prevalence of LD is 5% to 10%, as is widely believed, the method should identify 5% to 10% of the population as having SLD. The validity of classifications is evaluated by comparing the subgroups (e.g., low achievers vs. SLD) on variables like other cognitive skills, brain activation profiles, intervention response, or prognosis. If the classification is valid, the subgroups should show meaningful differences and the coverage should be adequate.

Identification is a reliability issue. If identification is reliable, different measures of the same attribute should show agreement that the person meets criteria for SLD. For IQ-achievement discrepancy, for example, two sets of IQ and achievement tests should agree that the person has SLD or two measures of word decoding should lead to the same decision about which individuals are impaired in basic reading skills.

Attributes of SLD

Regardless of the framework by which we wish to classify SLD, inherent characteristics of SLD make reliable identification difficult, create uncertainty, and possibly foment scientific disagreement.

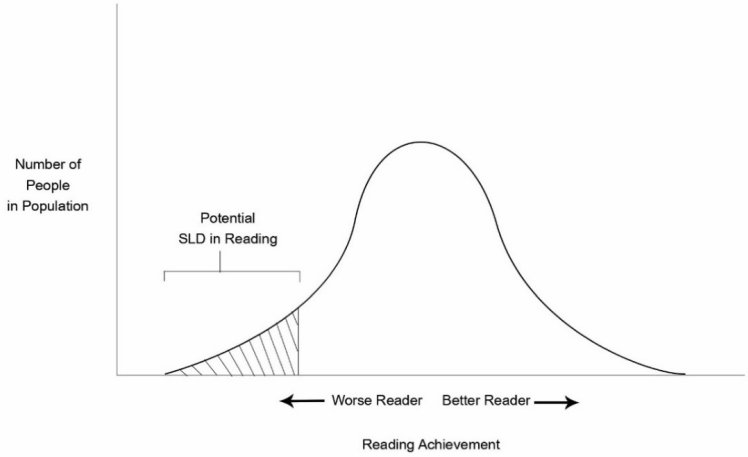
© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

The Attributes of SLD Are Dimensional

Generally speaking, there are two types of disorders: categorical and dimensional. Categorical disorders are binary conditions. An individual has the condition or doesn't. For example, pregnancy is a binary condition. You are either pregnant or not. Most cancers and viruses are categorical disorders. However, not all disorders are categorical; some represent a division of a continuous distribution with no natural demarcation, like high blood pressure. There is no natural threshold that separates an individual with hypertension from one without. Instead, clinicians draw artificial thresholds to identify individuals with hypertension based on empirical evidence of outcomes for individuals with elevated blood pressure. Interestingly, these thresholds are usually expressed as ranges, not single values, and take into account other factors, such as family history, dietary factors that increase risk for cardiovascular events, and the individual's history of blood pressure assessments.

The attributes of LD are similar. Although researchers once thought that a bimodal distribution was indicative of a categorical disorder, more rigorous evaluations find that the attributes of LD are continuously distributed, as illustrated for one dimension (reading achievement) in Figure 2. For example, no natural demarcation in reading comprehension separates students with SLD in reading from students without SLD. As a result, rigid cut points (e.g., performance greater than one grade level below grade expectation) are inherently problematic. A student who scores .9 grade levels below grade-level expectations in reading and a student who scores 1.1 grade levels below expectations are very similar, and the difference is likely measurement error if based on a single assessment. Further, it is likely that the educational needs of these students are highly similar. As a result, it is important that school psychologists and diagnosticians think about a range of scores, as expressed by confidence intervals, and do not apply rigid cut points.

Figure 2. A Continuous, Normal Distribution of Reading Achievement



SLD Is a Latent, or Unobservable, Construct

A latent construct is theoretical and cannot be directly observed. SLD is not directly observable outside of attempts to measure hypothesized attributes. However, no test perfectly measures the latent construct of interest; no reading test perfectly measures a child's ability to understand what she reads and no math test perfectly measures math ability. Thus, it is important to remember that all test scores, observations, or rating scales include uncertainty and error. This fact has important implications for the reliability of SLD identification decisions at the individual level.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

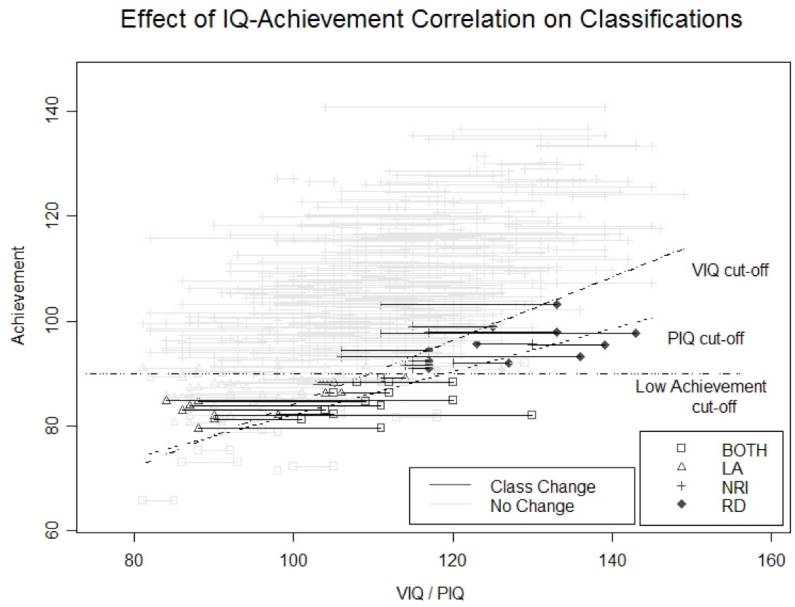
The Reliability of Identification Is a Universal Problem

All SLD identification methods have problems with reliability. If a formula or firm threshold is used, a student identified with one method may not be identified with SLD using another method or even another set of tests. The issue of low agreement is a universal concern when identifying learning disabilities using psychometric tests with fixed cut points. In addition, different measures are correlated, and the measures themselves are slightly unreliable, so it becomes difficult to assess exactly where an individual resides relative to a fixed cut point. This is true whether the cut point is the score on an achievement test, such as everyone who reads below the 15th percentile has SLD. It is also true if we use a 16-point IQ-achievement discrepancy or a threshold of 60 words read correctly on an oral reading probe in an RTI method. Our ability to assess precisely where the student's true score is relative to this firm threshold is not reliable. Even with the same student, different tests or the same tests on different measurement occasions will generate a range of scores around the 15th percentile, a 16-point discrepancy, or reading 60 words per minute. If we have multiple tests of the same type of achievement and they are consistently below these thresholds, we can be more confident that the student's true score is below the threshold. Even better, we could express the unreliability of the test as the standard error of measurement and specify a confidence interval, so that a range of scores could indicate the presence of SLD. We could also incorporate other data supports that might inform the judgment of the ARD team, such as previous academic and classroom performance, grades, observations of the child, and the parents' and teacher's perceptions of the student's performance.

Examples of the Reliability Issue

Figure 3 illustrates the effects of rigid cut points and correlated variables on the reliability of identification decisions using an IQ-achievement discrepancy method. In the figure, two sloped lines differentiate those with SLD based on an IQ-achievement discrepancy measured with a verbal IQ score or a performance IQ score. The line is steeper for verbal IQ because it is more highly correlated with reading (r = .69) than performance IQ (r = .40) Changing the IQ measure shifts the individuals' IQ score right or left on the x-axis but does not move their achievement score up or down because the achievement measure is the same. That simple difference in IQ measures shifts individuals at the edges of the regression cut point on one IQ measure to either a discrepant SLD or low-achieving subgroup when the other IQ measure is used.

Figure 3. Illustration of the Reliability of Identification for SLD (Francis et al., 2005)



© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

These shifts are displayed by a horizontal line that connects pairs of observations. An observation that does not change in the identified group has the same symbol connected by a faint horizontal line; observations that change groups have two different symbols (reflecting the different identification decisions) connected by a dark horizontal line. As Figure 3 illustrates, observations with IQ scores that are most different and near the cut point are most likely to shift in their SLD identification decision, reflecting both measurement error and differences between how the construct of aptitude is assessed by verbal and performance IQ. Recall, though, that each of the dark lines in the figure denotes a single student who would change SLD category based on the test or criteria applied—not because the child or her educational needs have changed, but due to the inherent limitations of methods that rely on firm thresholds or use different methods or criteria. Further, one may observe in the figure that the distribution of students who shift categories results in a sloped area of uncertainty, in which identification decisions are most unstable. Students whose scores fall close to or within this area of uncertainty are likely very similar to those highlighted in the figure as shifters. What is different about these students, even if they are somewhat more stable in our illustration? Their educational needs and assessment results will be highly similar.

We will discuss how to deal with these reliability issues in the last section. But these examples show why firm thresholds and formulae are problematic for SLD identification. Methods are not interchangeable and different tests will give different answers, especially if identification is based on a firm threshold.

Frameworks for Understanding SLD

Embedded in the IDEA 2004 statutes and 2006 regulations are controversies about the most valid framework for understanding classification for SLD and the methods for identification that would emanate from those frameworks. These classifications are inherent in the regulations, but many districts have sought guidance in making decisions about these frameworks. Two general frameworks are relevant for IDEA 2004: cognitive discrepancy frameworks and instructional frameworks that emanate from RTI. Because these regulations were published more than 10 years ago, additional research has emerged on the reliability and validity of these different identification frameworks. In the sections that follow, we discuss the evidence base for different implementations of these frameworks and the identification methods embedded in each framework to provide guidelines and best practices.

Cognitive Discrepancy Frameworks

Most school psychologists and special educators are familiar with cognitive discrepancy frameworks for understanding SLD because of the 1977 regulations that introduced the IQ-achievement discrepancy method as a necessary inclusionary criterion for operationally defining SLD. Cognitive discrepancy frameworks for SLD hold that academic deficits among children with SLD are unexpected because of the presence of average or strong cognitive abilities or due to the presence of specific cognitive strengths and weaknesses. Within this framework, the discrepancy between cognitive processing ability and academic achievement represents one of the defining features of SLD and differentiates it from "garden variety" low achievement, which is low achievement that is commensurate with low overall ability and is therefore not unexpected.

Over the ensuing decades, concerns about the reliability and validity of IQ-achievement discrepancy methods emerged. As a result, new proposals for cognitive discrepancy frameworks began to be put forth, which we refer to as patterns of processing strength and weaknesses (PSW) methods. These

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

methods hypothesize that low academic achievement is unexpected because of the presence of cognitive processing strengths, in combination with specific cognitive weaknesses that provide a potential explanation for specific academic weaknesses. These methods draw a distinction between expected underachievement, which can be attributed to commensurate aptitude and achievement, and unexpected underachievement, which is by marked an intraindividual pattern of strengths and weaknesses, the latter representing an inclusionary criterion. Thus, methods based on this framework feature a comprehensive assessment that includes an extensive evaluation of achievement and cognitive processes.

Instructional Frameworks

An alternative hypothesis is put forth by advocates for an instructional framework for understanding SLD. Instructional approaches represent a different classification hypothesis concerning the intrinsic nature of SLD. These approaches hypothesize that unexpected underachievement is due to inadequate RTI interventions that are generally effective with most students. Thus, operational definitions and measurement models within an instructional framework attempt to document inadequate instructional response as a key inclusionary criterion, a major contrast with frameworks in which cognitive discrepancy is the major inclusionary criterion. Instructional approaches do not recognize a cognitive discrepancy as an intrinsic attribute of students with SLD. Therefore, the comprehensive assessment need not incorporate assessment of cognitive functioning, except to rule out other disabilities or disorders (e.g., intellectual disabilities).

How Can Competing Frameworks Be Compared?

Empirical Research

School psychology and special education are wide-ranging fields with foundations in empirical research. Among other topics, this research addresses (1) the nature of cognition and learning; (2) effective intervention practices for all learners, but particularly those who struggle to master basic foundational skills; (3) the relations between cognition, attention, and executive control processes, and academic achievement; and (4) the cognitive profiles of students who experience academic difficulties. Such studies are important and can identify potential intervention targets and inform proposals for SLD identification methods. However, prior to widespread adoption, proposed practices in school psychology and special education should be directly investigated in empirical research. SLD identification methods should be evaluated by applying the proposed criteria to form subgroups and comparing those resulting subgroups. Reliability for the proposed method should be evaluated by comparing decisions across differences in measures, measurement occasion, cut points, or specific criteria to determine agreement for identification decisions. The reliability of different methods is expressed at the level of individual decisions. For validity, groups that meet and don't meet the method-generated criteria can be compared on measures that are not used for group formation, such as other assessments of academic achievement, subsequent intervention response, cognitive functioning, or neuroimaging. If the proposed groups differ on these external dimensions, the underlying classification accrues validity. This process of validation through empirical research defines evidence-based practice in education and school psychology.

Data Simulation

Data simulation can also help inform valid decision-making from psychoeducational assessments. In this context, data simulation refers to a procedure in which datasets are constructed and used to evaluate the reliability and/or validity of specific methods or statistical analyses or to evaluate the effects on outcomes of changes in some of the relevant variables. When evaluating methods for SLD identification, data simulations are particularly valuable. First, data simulations allow researchers to investigate complex

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

methods that rely on multiple data points and would be expensive to investigate with real children. Second, data simulation allows researchers to evaluate how well the application of proposed methods parallel a "known" underlying structure. For example, a researcher could create a "true" SLD status, based on a priori criteria and evaluate whether unreliable test data would accurately identify children as having SLD or not, an impossible task in the real world, in which all data are unreliable. Third, data simulation permits manipulation of specific variables (e.g., test reliability, different cut points) to evaluate the effect of these changes on identification decisions. Finally, data simulation allows researchers to draw conclusions about the universality of underlying psychometric properties across a full range of possible correlations between tests and test reliabilities. For these reasons, a careful evaluation of the results of data simulation studies can help address important questions about whether proposed methods to identify SLD can ever achieve high enough reliability to warrant consideration for widespread adoption.

Fairness and Coverage

Finally, it is important to note the consequential nature of SLD identification procedures. The SLD identification decision may result in the provision or denial of procedural safeguards, as well as accommodations and modifications to instruction and assessment. Due to these significant consequences, it is critical that the identification process demonstrate an underlying fairness and that it be rooted in a response to educational need, rather than theoretical definitions of what SLD is. In addition, the method should identify a reasonable number of children with SLD. Macmann, Barnett, Lombard, Belton-Kocher, and Sharpe (1989) summarized this imperative as a need to root our classification systems in a "coherent psychology of helping" (p. 145). We echo this call in a humble assertion that the best method for the identification of SLD is the one that helps the most children learn to read, write, and do math well.

Validity Research on Cognitive Discrepancy Methods

IQ-Achievement Discrepancy

The key attributes of an IQ-achievement discrepancy method are aptitude and achievement, as illustrated in Figure 4. Aptitude is a **capacity** measure that indicates how much children should be able to learn. Although many have questioned this idea, it is important to recognize the meaning of aptitude as an indicator of the capacity to achieve. We have referred to these methods throughout this report as IQ-achievement methods because IQ scores are the most common measure of aptitude. However, IQ is not the only potential indicator of aptitude. For example, in the past, some have proposed that tests of listening comprehension or language might be used, but there has been little research into the reliability or validity of this use. Additionally, these methods may rely on different IQ tests or scores. Thus, one school psychologist may use a full-scale composite IQ, another might use performance IQ, and a third might rely on verbal IQ.

Figure 4. Illustration of IQ-Achievement Discrepancy

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Fletcher, Lyon, Fuchs, and Barnes (2019) reviewed the evidence for the validity of aptitude-achievement methods using the most commonly employed approach: IQ-achievement discrepancy. Within the school psychology and special education research communities, there is broad consensus that these methods lack validity because it is hard to find meaningful differences between students with low achievement who meet and don't meet IQ-achievement discrepancy. This research is summarized below.

What's Wrong With IQ-Achievement Discrepancy

1. The classification lacks validity. IQ-discrepant students and students with low achievement consistent with IQ who do not meet criteria for an intellectual disability do not differ practically in behavior, achievement, cognitive skills, response to instruction, and neurobiological correlates once definitional variability is controlled (Stuebing et al., 2002).
2. Although it is not an intuitive finding, IQ is not a strong predictor of intervention response when the initial level of academic development is included (Stuebing, Barth, Molfese, Weiss, & Fletcher, 2009).
3. Brain activation profiles of these students are not meaningfully different (Simos, Fletcher, Rezaie, & Papanicolaou, 2014; Tanaka et al., 2011).

At the time when IQ-achievement discrepancy methods were adopted, initial research seemed to support the validity of the discrepancy criteria based on a "hump" in the distribution of reading achievement, which the authors hypothesized represented a subgroup with SLD, although they did not use this term at the time (Rutter & Yule, 1975). However, subsequent studies have questioned these findings because no exclusionary criteria were applied. Many of the poor readers who represented the subgroup without an IQ-achievement discrepancy were brain-injured and/or had intellectual disabilities (Fletcher et al., 1994). Since this study, most efforts to validate these differences in IQ-discrepant and low-achieving poor readers have yielded no meaningful differences.

Figure 3 is based on a data simulation by Francis et al. (2005) that found that about 30% of students tested twice would change status from SLD to not-SLD and another 30% would change from not-SLD to SLD. This type of finding has been demonstrated in other data simulations (Macmann & Barnett, 1985; Shepard, 1980).

Two empirical syntheses evaluating the differences between poor readers who demonstrate an IQ-achievement discrepancy and poor readers who do not demonstrate a discrepancy have been completed. Empirical syntheses are systematic summaries of all of the research in a specific area. Often, the results are combined statistically in a process called meta-analysis to evaluate a specific research question across many research samples. In the first of these empirical syntheses, Hoskyn and Swanson (2000) identified 69 studies conducted between 1975 and 1996 with results that could be used to address the validity of the IQ-achievement discrepancies. They found negligible to small differences on several measures of reading and phonological processing but larger differences (IQ-discrepant > low achievement) on measures of vocabulary and syntax. Importantly, there was no evidence of practical differences in reading and measures most closely related to reading, like phonological awareness.

In a second empirical synthesis, Stuebing et al. (2002) evaluated 46 studies from a sample of more than 300 from 1973 to 1998. The synthesis sought to compare poor readers who were discrepant in IQ and achievement and poor readers who were not on external dimensions, like achievement, behavior, or cognitive functioning. The two groups were clearly not different in behavior or achievement. Cognitive

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

abilities closely related to reading, such as phonological awareness, rapid naming, verbal memory, and vocabulary, also demonstrated no differences. Nonverbal measures of spatial processing and concept formation indicated small, but higher scores for the IQ-discrepant group, reflecting their relation with the IQ test. The average difference in cognitive ability in these areas was about 0.3 standard deviations, despite a difference of 1 standard deviation in IQ (which had been used to form the groups). This finding means that the application of IQ-achievement discrepancy criteria required an IQ-achievement difference of 15 points, but the difference between the groups in other cognitive abilities was 4 to 5 standard score points.

Other studies found little evidence that validity evidence would emerge if researchers used different or better aptitude and achievement measures, although these differences in measures would affect who is identified with SLD at the individual level. Further, no differences were found in the long-term development of reading skills in poor readers with and without IQ-achievement discrepancies (Flowers, Meyer, Lovato, Wood, & Felton, 2001; Vellutino, Scanlon, & Lyon, 2000). Over time, these two groups develop reading skills similarly. This is true even when evaluating how groups of poor readers with and without an IQ-achievement discrepancy respond to specific reading interventions. To address this question, several studies have examined whether IQ alone or the existence of an IQ-achievement discrepancy predicted how well a student would respond to an intensive reading intervention. In an empirical synthesis summarizing research addressing these questions, Stuebing et al. (2009) found that IQ accounted for almost no unique variance in intervention outcomes. This finding means that although some students showed good response and others demonstrated little response to intensive reading interventions, IQ (and therefore IQ-achievement discrepancy) was not meaningfully related to these differences and could not be used to predict who would respond and who would not. Finally, in functional neuroimaging studies, Tanaka et al. (2011) and Simos et al. (2014) found no differences in the activation patterns associated with word reading in groups of poor readers divided by the presence or absence of an IQ-achievement discrepancy.

What Do These Findings Mean?

When multiple studies show weak validity, indicated by no practical differences between groups produced by the identification criteria, the validity of the underlying approach to identification must be questioned. At a superficial level, it may appear that aptitude-achievement methods represent a valid identification process because the methods identify the lowest-achieving students at each level of IQ. However, that does not represent the most critical comparison; aptitude-achievement methods are not valid because there is not sufficient evidence to demonstrate that low-achieving students with and without an aptitude-achievement discrepancy are different in any educationally meaningful way. This creates an issue of fairness because their educational needs are similar. Coverage is not an issue because that would depend on the threshold for low achievement that was set; the method will generally identify the number of children at the threshold.

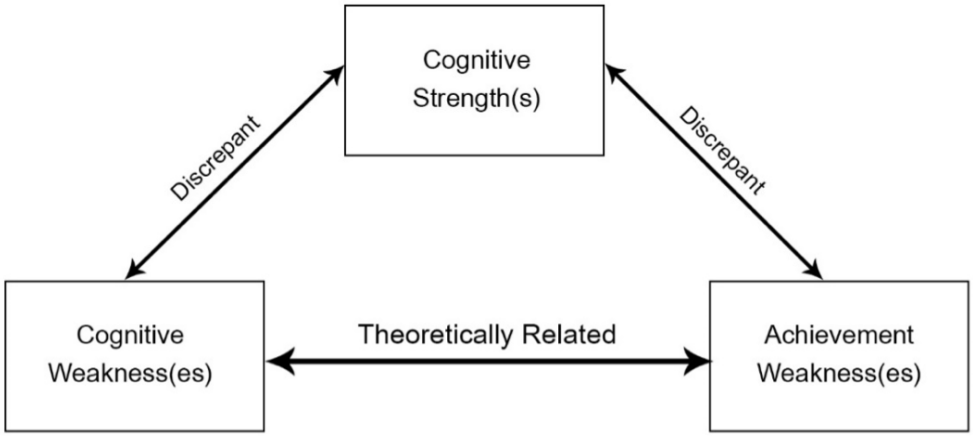
Processing Strengths and Weaknesses

Identification methods based on PSW are commonly proposed and discussed in the professional school psychology literature and are frequently implemented in Texas and many other states. However, there is considerable controversy about the reliability and validity of these models. As Figure 5 illustrates, PSW models generally require the identification of a cognitive strength and weakness, in addition to an academic weakness. The cognitive weakness must be theoretically related to the achievement weakness. Proponents argue that the related academic and cognitive weaknesses can identify the cause of academic difficulties, but such conclusions are not possible from limited test data at one time point. Described with terms such as "concordance-discordance" and "cross-battery," PSW methods are often

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

treated as interchangeable, independent of the tests used to operationalize the methods, and facilitating of intervention. While appealing logically, there is little evidence for the reliability and validity of these approaches.

Figure 5. Illustration of PSW for SLD Identification Conceptual Model



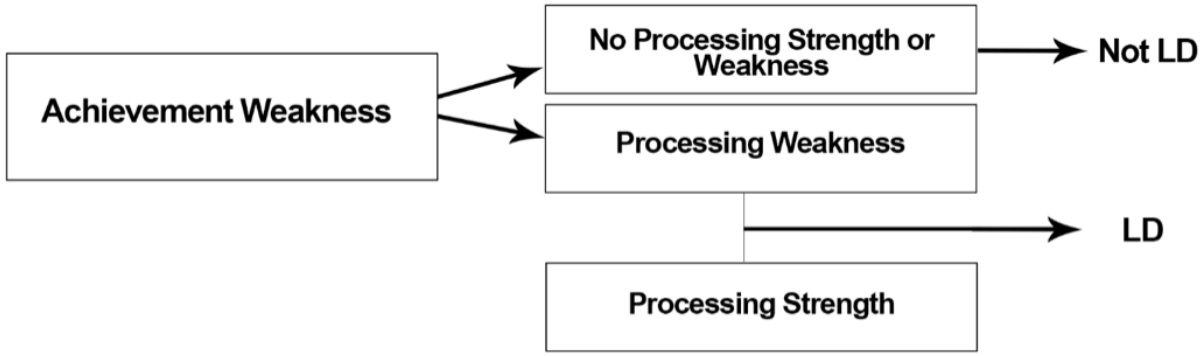
Multiple approaches have been proposed to identify SLD with a PSW approach and are in use in Texas: (1) the concordance/discordance method (C/DM; Hale & Fiorello, 2004); (2) the dual discrepancy/consistency criteria (DD/C; Flanagan, Ortiz, & Alfonso, 2013), a method that emanates from the cross battery assessment (XBA) approach to comprehensive evaluation; (3) the discrepancy/consistency method (D/CM; Naglieri, 1999); (4) the core-selective evaluation process (C-SEP; Schultz & Stephens, 2015, 2017); and (5) the psychological processing analyzer (PPA) method (Dehn, 2013). Figure 6 illustrates the inclusionary criteria that must be documented across these PSW methods.

However, these methods differ in several ways, including how low achievement is established, how a profile of PSW is defined, and how exclusionary factors are considered. For example, in C/DM, cognitive scores are used to identify a within-person PSW without consideration of normative expectations. In contrast, DD/C, C-SEP, and D/CM use both normative and within-person comparisons. The methods also differ in how they apply cognitive theory and what tests are specified. For example, DD/C and C-SEP are closely linked to XBA and use the Cattell-Horn-Carroll theory of intelligence. D/CM uses the planning, attention, simultaneous, and successive factors of intelligence measured by the Cognitive Assessment System only (Naglieri & Das, 1997). In contrast, C/DM emphasizes flexibility across different tests and theoretical orientation. Finally, all PSW methods are used in a broader context that permits application of exclusionary criteria and clinical judgement. Our focus is on the reliability and validity of the inclusionary criteria (i.e., the complex statistical algorithms used to identify the patterns outlined in Figure 6).

These methods share common shortcomings, including problems with the algorithms and formulae that have the same unreliability problems outlined for IQ-achievement discrepancy methods because of the use of firm cut points and discrepancy scores. None of the methods has convincing evidence of validity (Beaujean, Benson, McGill, & Dombroski, 2018; Benson, Beaujean, McGill, & Dombroski, 2018; Fletcher & Miciak, 2017). All have been understudied (Schneider & Kaufman, 2017). They are not interchangeable, and the same individual might be identified as having SLD by one method and not having SLD by a different method (Fletcher et al., 2019).

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Figure 6. Illustration of the Identification Criteria of PSW Methods



Validity of PSW Methods

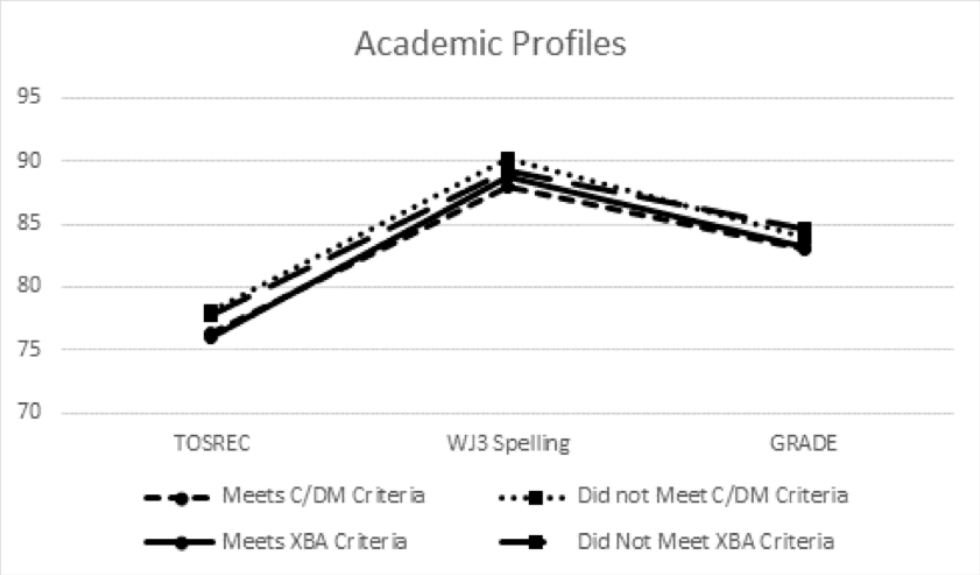
Some proponents of PSW methods argue that IDEA 2004 requires cognitive assessments for identification. In fact, the statute defines SLD as "a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations." This definition indicates that the manifestations of these cognitive difficulties in academic achievement are critical, not the psychological processes themselves. Academic difficulty is the defining feature of SLD, and the statute explicitly defines the areas in which these difficulties may occur. This conclusion is clearly supported by the regulations accompanying IDEA 2004: "The Department does not believe that an assessment of psychological or cognitive processing should be required in determining whether a child has an SLD. There is no current evidence that such assessments are necessary or sufficient for identifying SLD. Further, in many cases, these assessments have not been used to make appropriate intervention decisions" (Individuals with Disabilities Education Act Regulations, 2006, p. 46651).

A decade later, there remains very little evidence that assessments of psychological or cognitive processing are necessary or sufficient for the identification of SLD (Beaujean et al., 2018). Few empirical or simulation studies have investigated the validity of PSW methods for SLD identification. Proponents of PSW methods generally cite four types of supporting evidence, including studies that investigate (1) the structure of cognitive functioning, (2) the relations between cognitive functioning and academic achievement, (3) whether neuropsychological profiles can be established from test data, and (4) potential aptitude by treatment interactions (e.g., Dehn, 2013). However, such studies do not represent direct evidence of the validity of SLD identification decisions that require the documentation of a PSW profile. Comparatively few peer-reviewed studies have directly compared groups formed on the basis of PSW profiles on external dimensions. Thus, we cannot rely on meta-analyses and must look instead to simulations and the few empirical studies that have been completed.

Miciak, Fletcher, Stuebing, Vaughn, and Tolar (2014) compared low-achieving children with reading problems identified with SLD using either C/DM or DD/C (called XBA in Figure 7) and those identified as slow learners who were not identified with SLD based on these PSW criteria. The groups were compared on achievement tests that were not used to define the groups. As Figure 7 shows, there was little difference in the shape or elevation of the achievement profiles generated by four different operationalizations of PSW methods.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Figure 7. Achievement Profiles for Students Who Meet and Do Not Meet PSW Criteria



In another study, Miciak et al. (2016) examined a large intervention database with extensive assessments of cognitive functions to determine SLD status of a large group of struggling readers using procedures recommended by C/DM or DD/C. The results demonstrated little evidence for value-added increments relative to pretest assessments of reading skills. The individual cognitive assessments and application of PSW SLD identification methods did not help predict who would respond and who would not respond to the intensive interventions.

PSW methods continue to be proposed and recommended, despite the fact that the basic psychometric issues and shortcomings are well understood and have been documented for many years. In studies of profile analysis based on the Wechsler intelligence scales, little evidence has emerged linking SLD to specific cognitive PSW. In a simulation, Macmann and Barnett (1994) evaluated differences in verbal IQ and performance IQ factor index scores and ipsative profile patterns on the Wechsler Intelligence Scale for Children, reporting that the reliability was poor and that practitioners should not use the results for making identification decisions. The measurement issues make any method based on cognitive discrepancies unlikely to achieve reasonable levels of reliability (Beaujean et al., 2018).

Finally, advocates of PSW methods for SLD identification argue that academic interventions can be tailored to specific cognitive profiles (revealed through the PSW assessment process) and that a full understanding of the student's cognitive profile is necessary to design an effective intervention for the student. This argument rests on the assertion that cognitively tailored interventions will improve academic outcomes. However, little evidence supporting the effectiveness of interventions based on cognitive-process profiles has emerged, much less interactions of cognitive processes and interventions to influence treatment outcomes (Fletcher et al., 2019). In a recent meta-analysis, Burns et al. (2016) examined the role of cognitive tests in relation to intervention. Across different uses (screening, intervention design), the effect of cognitive tests and outcomes ($g = .17$) was much smaller than the effect of reading fluency ($g = .43$) and phonological awareness ($g = .48$). To the extent that academic interventions should be tailored to individual students' strengths and weaknesses, the results of Burns et al. suggest that it is considerably more effective to tailor interventions based on patterns of academic strengths and weaknesses, rather than more weakly related cognitive processes.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

PSW Methods Are Not Interchangeable

Proponents of PSW methods have implied that the four primary methods are research based and interchangeable. This is not the case. In two nonoverlapping samples, Miciak and colleagues (Miciak et al., 2014; Miciak, Taylor, Stuebing, & Fletcher, 2018) compared the SLD identification decisions emerging from two PSW methods: the C/DM and the DD/C approaches. When similar cut points were employed, agreement for SLD identification decisions did not exceed what would be expected by chance, raising significant questions about the comparability of the methods.

Even when the method used to identify SLD is held constant, identification decisions fluctuate due to differences in test selection. In a third nonoverlapping sample, Miciak, Taylor, Denton, and Fletcher (2015) compared the identification decisions emerging through an application of the C/DM criteria using two test batteries that were theoretically equivalent but used distinct academic measures. Results demonstrated low agreement, particularly for positive SLD identification decisions (i.e., who is identified as having SLD). Taylor, Miciak, Fletcher, and Francis (2017) completed a simulated replication of Miciak et al. (2015) to evaluate whether those findings could be generalized beyond the sample and measures used. Across a broad range of potential relations between academic weakness, cognitive weakness, and cognitive strength, percent positive agreement was low to moderate, ranging from .33 to .59 across all scenarios, meaning that in the most optimal scenarios, slightly more than half of all students identified as having SLD with one test battery would be identified by both batteries. These results suggest that within complex PSW methods that rely on specific patterns of difference scores between tests, changes in test selection will have significant, deleterious effects on agreement for SLD identification.

The low reliability for SLD identification decisions highlighted above are not likely to be ameliorated with more or better tests. Miciak et al. (2018) conducted a simulation to investigate whether classification accuracy was improved within C/DM through the use of multiple measures, rather than a single indicator within each domain. A second test of each construct (e.g., reading, verbal knowledge) was incorporated under two scenarios: (1) a recursive test-retest procedure in which a positive SLD identification decision was confirmed through a second assessment and (2) a mean score procedure. When the test-retest procedure was used, accuracy in positive SLD identification decisions was diminished, reflecting a trade-off in error types—fewer students were identified as having SLD, and correspondingly more students who "truly" had SLD were identified as not having SLD. When mean scores were used, modest improvements in classification accuracy were observed. However, these improvements in accuracy must be understood in a context of limited resources and time. In the most favorable scenario, a 3% increase in correct classifications was observed. To achieve this 3% increase, the total amount of testing had to be doubled (two tests were given within each domain instead of one). These results demonstrate that the inherent psychometric limitations of PSW methods are not likely to be corrected through the addition of tests to more accurately measure the construct. Instead, the results of this simulation illustrate how the greater complexity of PSW methods premised on finding an intraindividual PSW across multiple domains exacerbate the reliability issues highlighted in the prior sections.

Fairness and Coverage

A related issue concerns the fairness and coverage of PSW methods for SLD identification. SLD identification is a relatively high-stakes decision affecting educational programming and civil rights protections. It is therefore important that the application of SLD identification methods allow for the provision of these programs and protections for students who need them. Recent research suggests that PSW methods fail this critical test. Stuebing, Fletcher, Branum-Martin, and Francis (2012) used simulated data to evaluate the identification rates and classification accuracy of the C/DM, DD/C, and DC/M. Across methods, identification rates for SLD were quite low (1% to 2%). Further, results indicated that the

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

methods were particularly unreliable in identifying "true" SLD cases, with many false-positive identification decisions. Kranzler, Floyd, Benson, Zaboski, and Thibodaux (2016) obtained similar results for the DD/C methods used in XBA in a study using normative data from the Woodcock-Johnson III cognitive and achievement tests. SLD identification rates were low across multiple scenarios. Thus, DD/C was relatively good at detecting "true" cases of students without SLD. However, the method was less accurate for detecting "true" SLD. On average, the method detected only 21% of the "true" SLD cases and only 34% of the participants identified as SLD were "true" SLD cases. These results highlight fundamental questions of coverage and fairness for PSW methods.

Problems With PSW Approaches to SLD Identification

1. Federal statutes do not mandate that cognitive skills be assessed—just their manifestations.
2. There is little empirical research on how PSW methods actually work and how these methods may inform instruction.
3. Arguments for PSW methods are frequently premised on a straw-person critique of other SLD identification methods, particularly RTI. There is no standalone RTI SLD identification method, and a comprehensive evaluation is always required, regardless of the identification method.
4. Reliability issues associated with the use of discrepancy scores of any kind are well known, especially the use of rigid cut points and profile interpretation of difference scores.
5. Simulation and empirical studies suggest that PSW methods identify very few students with SLD and are unreliable at detecting "true" SLD cases, raising issues of coverage and fairness.

What Do These Findings Mean?

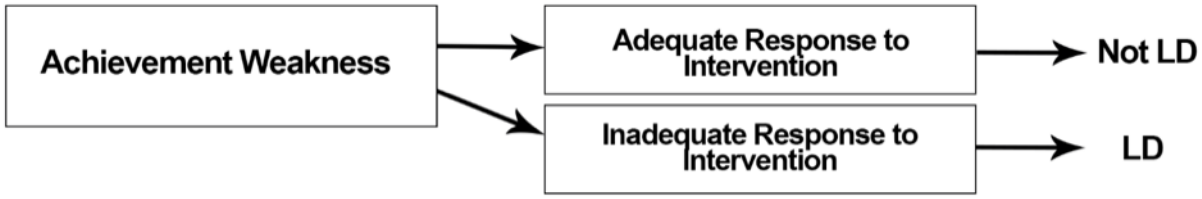
As with IQ-achievement discrepancy, PSW methods fail to show meaningful differences between low-achieving groups that meet and don't meet the identification criteria. Little evidence supports PSW methods or many of the assumptions on which the methods are premised. Like all methods to identify SLD, PSW methods have inherent problems with reliability for individual identification decisions. The complexity of PSW methods and use of difference scores on multiple, correlated tests exacerbate these problems. Of particular concern is the fairness and coverage of PSW methods, where very low numbers of individuals are identified with SLD, and evidence suggests even those decisions are unreliable.

Instructional Discrepancy (RTI) Approaches

In a method based on instructional response, inadequate response to instruction is the key attribute of the classification. Most methods based on instructional response have three components: low achievement, inadequate instructional response, and consideration of other disorders and contextual factors that indicate the absence of SLD (see Figure 8). Intervention response, like all attributes of SLD, is unobservable outside of attempts to measure it. Additionally, research suggests that intervention response lies on a continuum, with no naturally occurring demarcations that would separate adequate from inadequate responders, similar to other attributes of SLD. As a result, students whose measured performance lies near the cut point(s) for inadequate response (however it is defined) will be similar and individual identification decisions will demonstrate some level of unreliability.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Figure 8. Illustration of the Inclusionary Criteria for Instructional Discrepancy Approaches



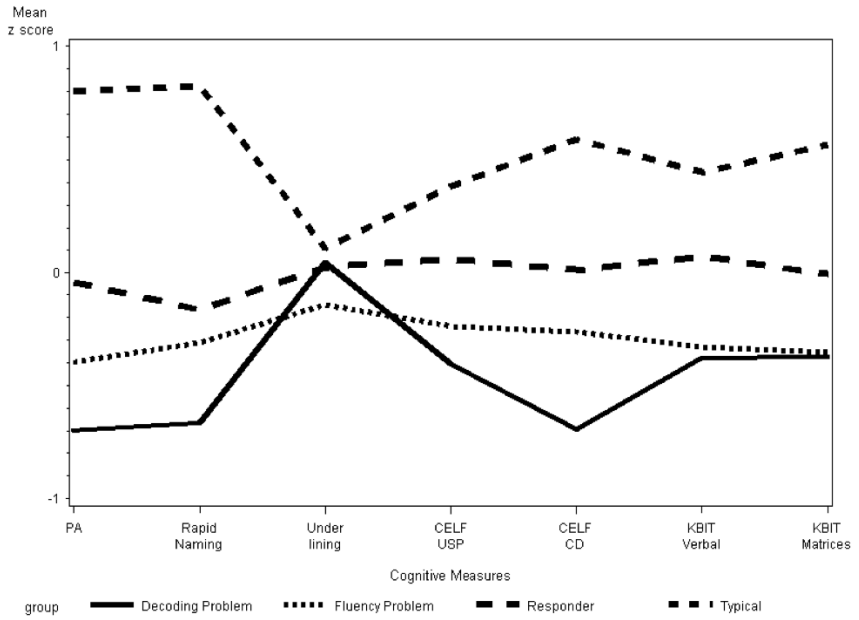
There is no universally agreed-upon criterion for operationalizing inadequate instructional response. How it is operationalized varies across states and districts, as well in research investigating RTI. In broad terms, inadequate response can be based on three types of data: (1) student growth over time (slope), (2) postintervention performance (final status), or (3) both (dual discrepancy). However, none of these methods overcomes the inherent unreliability that emerges when we apply arbitrary cut points on imperfect, continuous data that are normally distributed. Many critics of RTI methods focus on the use of curriculum-based methods for assessing progress, but many proposed operationalizations of RTI methods also use traditional norm-referenced tests of achievement as a final status method. A debate on the relative merits of the many different ways in which inadequate instructional response has been defined is beyond the scope of this practice guide, except to note that all attempts to operationalize inadequate instructional response are subject to the same challenges confronting all psychometric approaches for the identification of SLD.

To evaluate the validity of classifications based on intervention response, inadequate and adequate responders should be compared on domains not used to identify responder status, similar to how other methods for SLD identification have been evaluated. These domains could be highly related attributes, such as reading or writing, or could include other domains like behavior, attention, or neuroimaging. It is in these comparisons that key differences in the research support for methods based on instructional discrepancies and cognitive discrepancies emerge. Unlike for methods based on cognitive discrepancies, empirical research consistently demonstrates that classifications based on how students respond to intensive interventions create subgroups who differ on a number of theoretically related attributes, including academic level, cognitive characteristics, behavior, and even brain-activation patterns (Fletcher et al., 2019).

These data are evidence for the validity of intervention response as a classification attribute because they suggest that subgroups demonstrate different characteristics beyond those differences on which the groups were formed. These comparisons are the critical validity test for any proposed classification. However, these differences between groups are not evidence for a categorical disorder. Instead, these differences lie on a continuum reflecting the severity of academic difficulty. Students who are most severely impaired in reading are most likely to demonstrate cognitive deficits or abnormal brain-activation patterns. For example, Fletcher et al. (2011) investigated the academic and cognitive profiles of adequate and inadequate responders to a reading intervention. Inadequate responders were identified using fluency or decoding measures at posttest and example of the application of final status response criteria. This method resulted in four groups, including students with (1) deficits in decoding **and** fluency, (2) deficits in fluency only, (3) adequate response to the intervention, and (4) no reading difficulties (typically achieving children). Across a range of cognitive measures, a clear stepwise progression was evident, with the typically achieving and adequate responder groups outperforming both inadequate responder groups. Comparisons of both inadequate responder groups with the typically achieving and adequate responder groups were statistically significant, providing evidence for the difference between these groups formed on the basis of intervention response.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Figure 9. Illustration of Cognitive Differences Based on Responder Status



What Do These Findings Mean?

Methods that emanate from RTI show validity because groups formed on the basis of intervention response are different on a number of theoretically related attributes. Importantly, this evidence for validity is the same as what would accrue for a definition based on simple low achievement. If an achievement distribution is subdivided and groups of students in one part of the distribution (e.g., below 20th percentile) are compared to groups selected from another part (e.g., above the 20th percentile), differences will be seen on other external variables because of the correlation with the achievement variable used to create groups.

Recommendations for SLD Identification

Summary of Reliability and Validity Issues

All identification methods for SLD show low reliability, especially in terms of agreement across methods, for individual decisions if implemented as formulae with firm thresholds. These problems are inherent in IQ-achievement discrepancy methods. At least in an IQ-achievement discrepancy method, the lowest achievers are selected at each point in the distribution of IQ scores. Although the classification is not valid and the decisions may lack fairness and potentially reduce coverage, most students identified with this method will meet reasonable criteria for SLD as long as there is a low achievement threshold. The same cannot be said for PSW methods, which under-identify students with SLD, thus demonstrating coverage and fairness problems. There are formulae and thresholds, but the user is often encouraged to go through a multistep process that allows the clinician to make decisions that override the algorithms.

Both of these methods require testing that is not clearly related to achievement or instructional planning. IQ scores are weak predictors of intervention response in the absence of an intellectual disability. And there is little evidence that training cognitive processes leads to better academic achievement or that having IQ or cognitive assessments improves the reliability of individual identification. Because

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

discrepancy scores are involved between two correlated domains, cognitive discrepancy methods actually reduce the reliability for individual decisions. There is little validity evidence for both cognitive discrepancy frameworks, and different methods are clearly not interchangeable. In fact, specific achievement tests from different batteries are not interchangeable because of differences in how they are constructed, the items, the reliability, and the normative basis. Even in statistical simulations, different decisions about where a student scores relative to a threshold will vary with the selected test and its normative basis. It is for this reason that IDEA 2004 defined the comprehensive evaluation as a data-gathering process. Cognitive testing is not required as part of this comprehensive assessment. Additionally, federal regulations mandate that the ARD team use multiple criteria and consider the exclusionary factors.

Methods that include RTI have promise, as demonstrated by strong evidence for the validity of classifications based on intervention response. This body of research clearly establishes that intervention response is educationally meaningful and is strongly related to a number of educationally relevant domains (e.g., achievement, cognitive functioning, behavior). However, these methods are not a panacea and successful implementation requires a fully implemented multitier system of supports. These methods do not solve the reliability problem if the primary criterion is an assessment of intervention response that is set to a strict threshold (e.g., curriculum-based measurement [CBM] with a firm grade-level benchmark). It is inherently difficult to reliably assess the position of a student relative to a firm threshold. Additionally, methods based on RTI must adhere to IDEA 2004 requirements for a comprehensive evaluation. This evaluation must include multiple criteria, exclusionary factors, and observations of the student to meet special education eligibility requirements and to plan effective treatment plans.

Recommendations for Improving the Reliability of Identification

With these issues in mind, we first address approaches to dealing with the reliability problem, regardless of the identification framework employed.

1. **Use multiple data points.** SLD identification decisions should never be based on a single data point, whether a single test or the documentation of a single criterion. The use of multiple indicators meets statutory requirements for a comprehensive assessment. Additionally, the use of multiple data points allows for (1) greater sensitivity in identifying potential academic problems, (2) assessment of a full range of component academic skills, and (3) the collection of data that might inform future academic interventions.

2. **Avoid fixed cut points.** Any cut point applied to test data that is continuous and normally distributed is inherently arbitrary. There is no natural demarcation between scores indicative of SLD and scores not indicative of SLD; students close to the cut point will be very similar. Thus, diagnosticians should avoid applying methods that rely on the strict application of cut points. Districts should not use formulae or strict thresholds as a gatekeeper for SLD.

3. **Use confidence intervals.** Confidence intervals reflect the degree of uncertainty associated with a test score by providing a range of possible values. Instead of adopting a firm threshold, the standard error of measurement should be applied to the threshold to generate an interval in which there is a high probability that the true score will reside. Diagnostic decisions in other disability categories have moved toward the incorporation of confidence intervals and clinical judgment. For example, determining levels of IQ for an intellectual disability typically requires application of the standard error of measurement to create a 95% confidence interval. Because IQ scores 2 standard deviations below the average of 100 are usually required (i.e., 70 or below), this is expressed as a score between 65 and 75.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

4. **Employ high thresholds for treatment planning.** When possible, ARD teams should error in the direction of providing necessary academic interventions, particularly in the context of early intervention efforts in basic academic skills in which small amounts of intervention can have a relatively large impact on academic outcomes. This recommendation reflects the relatively low cost of treatment and the highly consequential impact of untreated academic difficulties. Misidentifying a student in need of services (false-negative error) is a costlier error than identifying a student as needing services who does not have SLD (false-positive error). In the latter example, the child can be pulled out of any interventions, which is why continuous monitoring of students in an intervention is important. With a false-negative error, the student continues to languish when it is well-known that early intervention is critical for the successful treatment of SLD.
5. **Use tests with the same normative basis.** One source of unreliability across different tests is due to norm scores based on comparisons to different norming populations. This misalignment can be controlled by using tests that were normed on the same population, thereby reducing a significant source of variability in score comparisons and allowing for more reliable decision-making.

A Treatment-Based Approach to the Comprehensive Evaluation

Regardless of the method for SLD identification, practitioners should adhere to the recommendations above for implementing a more reliable process. However, maximizing reliability does not necessarily result in a valid process. To ensure a valid identification process, we must consider aspects beyond test selection and decision-making. Modern validity theory posits that validity is not an inherent attribute of a test or procedure. Instead, validity must be considered holistically as an evaluation of the procedures and decisions we make, as well as the consequences of those decisions (Messick, 1986). With this in mind, the discussion must move beyond classification and eligibility and toward processes that improve the lives of children. These recommendations often accompany discussions of best practices for methods based on RTI but in fact represent a best-practice approach to SLD identification overall. This is because the primary goal of identification is not simply eligibility. The primary goal is improved treatment of persistent academic difficulties. Funds spent on eligibility subtract from funds available for intervention. Therefore, the amount of testing administered as part of the identification process should be limited to only those that inform future intervention (i.e., the formulation of an effective individualized education program). This approach to identification lends itself to a comprehensive evaluation that is less time consuming and gives priority to intervention because of its focus on the assessment of academic skills and instructional response.

A consensus group convened by the Office of Special Education Programs in the Department of Education recommended three essential criteria for SLD (Bradley, Danielson, & Hallahan, 2002). The comprehensive evaluation must document the following:

1. The student demonstrates low achievement.
2. There is insufficient response to effective research-based interventions.
3. The team considered and ruled out exclusionary factors, including intellectual disabilities, sensory deficits, serious emotional disturbances, a lack of opportunity to learn, and language-minority status where low achievement is due to lack of proficiency in English.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Role of Norm-Referenced Achievement Tests

A comprehensive assessment of current academic functioning in all areas of suspected difficulties with a norm-referenced assessment provides valuable information not only for eligibility, but also for treatment. Current achievement levels, as well as individual strengths and weaknesses in reading, math, and writing, can help instructors individualize an intervention plan and determine the necessary level of intensity. Proponents of cognitive assessments often argue that achievement tests are not helpful because the student has already been assessed on numerous occasions. However, the comprehensive evaluation is often the first formal evaluation of achievement for the student across all domains and subdomains. Proponents of methods based on RTI often argue that CBMs are sufficient. In fact, CBMs may be slightly less reliable than many norm-referenced tests, represent only a single criterion, and do not provide a comprehensive picture of the student's skills within a specific domain (e.g., basic reading, reading fluency, comprehension). Even in methods based on RTI, more reliable identification result from the use of both CBM and norm-referenced test data, which is required in many states.

Norm-referenced achievement tests should be targeted to the academic domains of potential SLD identified in IDEA. These assessments should include a brief assessment of foundational skills involved in basic reading, math calculations, and basic writing, such as spelling and the higher-order skills of reading comprehension, math reasoning, and writing composition. These assessments necessarily take more time, as they assess more complex skills, but they provide valuable information about the child's current functioning. In addition, it is always important to assess automaticity because the inability to work quickly may require adaptations in classroom instruction. The goal is always to minimize time spent testing and, to the extent possible, assess with tests that have the same normative basis.

The sidebar "Assessment of Major Academic Domains by Norm-Referenced Tests" presents the major achievement domains of SLD and how they are assessed within the three major norm-referenced assessment batteries: the Woodcock-Johnson IV, the Wechsler Individual Achievement Test-III, and the Kaufman Tests of Educational Achievement. We do not address the issues of oral expression and listening comprehension because they really represent domains of language functioning and are better dealt with through the speech and language impairment category of IDEA 2004. In the sections that follow, we provide a description of each domain and its importance to academic success.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Assessment of Major Academic Domains by Norm-Referenced Tests

Construct	Woodcock-Johnson IV	Wechsler Individual Achievement Test-III	Kaufman Tests of Educational Achievement
Core Tests			
Word recognition	Word identification	Word reading	Letter and word recognition
Phonetic decoding	Word attack	Pseudoword decoding	Nonsense word decoding
Reading fluency	Word reading sentence reading	Oral reading	Silent reading
Reading comprehension	Passage comprehension	Reading comprehension	Reading comprehension
Math computations	Calculation	Numerical operations	Computation
Math problem-solving	Applied problems	Problem-solving	Concepts and applications
Written expression	Spelling	Spelling	Spelling
Supplemental Tests			
Math fluency	Math facts	Math fluency	
Writing fluency	Sentence writing	Alphabet writing	Writing fluency
Written expression	Writing samples	Essay composition	Written expression

Reading

In IDEA 2004, students can be identified with SLD in three reading domains: basic reading (dyslexia), reading fluency, and reading comprehension. Basic reading represents a problem with word-recognition accuracy and fluency (and usually spelling). Some children read accurately but slowly (reading fluency). Both of these problems are often referred to as dyslexia, an SLD that involves the ability to read single words accurately and fluently. Still other students do not show problems with reading words or text accurately and fluently but struggle to understand what they read (reading comprehension).

Word-Recognition Accuracy

Most normative assessments include subtests that require the untimed oral reading of isolated real words and pseudowords. These tests measure students' sight word knowledge and capacity for sounding out words. These tests are typically the best single predictor of overall levels of academic achievement and they are vital for the identification of dyslexia, which is defined by problems reading and spelling words accurately and fluently in isolation.

Reading Fluency

There are many quick, affordable measures of reading fluency. The achievement batteries outlined in the sidebar "Assessment of Major Academic Domains by Norm-Referenced Tests" include timed reading fluency measures. Some reading fluency assessments require the student to read single words aloud fluently and accurately; others require the student to read connected text. Other fluency measures are hybrid fluency/comprehension measures and require the student to fluently read text and process for meaning, such as the Woodcock-Johnson Sentence Reading Fluency subtest and the Wechsler Individual Achievement

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Test Oral Reading subtest. Quick alternatives are the Test of Word Reading Efficiency-2, which involves oral reading of real words and pseudowords from a list, and the Test of Reading Fluency, which requires text reading. Grade-appropriate CBMs are also appropriate. The key to assessing reading fluency is for the student to read text aloud quickly and accurately, so that fluency can be measured in terms of words read correctly per minute.

Reading Comprehension

Reading comprehension is difficult to assess with a single measure, and different comprehension tests will give slightly different scores because of differences in how they assess reading comprehension. As a result, it is important to note the nature of the material the person reads as well as the response format. For example, reading comprehension varies according to what the child reads (sentences, passages, genre), how the child demonstrates understanding (cloze, open ended, multiple choice, think-alouds), how much the child must remember (answering questions with and without the text available), and the complexity of the text and the ideas within (vocabulary elaboration vs. knowledge, inferencing, activation of background knowledge). If a test contains text beyond the child's word recognition or fluency skills, it is unlikely to isolate comprehension skill and multiple measures that assess reading comprehension in different ways may be needed.

A good assessment of reading comprehension requires reading complex text. If a student has completed group assessments of reading comprehension, such as state-mandated assessments of reading, the results can be reviewed as part of the evaluation. However, it is important to also evaluate levels of effort.

Planning for Reading Interventions

To the extent possible, evaluating relative skill levels in word recognition and reading comprehension can help to differentiate intervention programs. Students with severe reading difficulties need a comprehensive reading program that includes systematic instruction in foundational reading skills. Students with specific deficits in comprehension may require more text- and language-focused interventions. These determinations can be made by planning an assessment that incorporates tests listed in the sidebar "Assessment of Major Academic Domains by Norm-Referenced Tests."

Mathematics

IDEA 2004 identifies two domains of SLD involving calculations (dyscalculia) and problem solving. Calculations are problems with basic math skills, including fact retrieval and other components of accurate computation. Problem solving usually involves procedural knowledge and is assessed with word problems.

Math Calculations

Measures of math calculations typically include items that range from basic arithmetic to algebra and geometry. Unlike reading, which develops in a more linear fashion, low performance on math computation tests could reflect problems in many areas, including fact retrieval, procedural knowledge, and attention difficulties. This is especially apparent for students with attention-deficit/hyperactivity disorder (ADHD). Math computations typically rely on a paper-and-pencil format with computational problems presented in isolation, making it particularly useful for isolating potential math difficulties in the presence of potential reading and language problems.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Math Problem Solving

A second domain of mathematics is problem solving. Most norm-based assessments include a measure of problem solving, and several are listed in the sidebar "Assessment of Major Academic Domains by Norm-Referenced Tests." These tests typically involve solving real-life math problems or "word problems." These sorts of problems are frequently difficult for children with reading difficulties, especially if they have to read the problem. Additionally, because many children with reading problems have language difficulties, children with reading problems often struggle with math problem solving even when the problems are read to them.

Math Fluency

Unlike reading, there is little evidence of SLD involving math fluency, but fact retrieval is slow for many students who struggle in math. The batteries highlighted in the sidebar include three timed assessments of computational skills that help identify students who lack automaticity in basic arithmetic skills, which can lead to difficulties in mastering more advanced mathematics.

Planning for Math Intervention

Basic math computation and fact retrieval difficulties are best addressed through comprehensive math programs that teach procedural knowledge through word problems. A comprehensive assessment can assist in intervention planning by providing data to guide the amount of time devoted to practicing fact retrieval and basic arithmetic during the problem-solving intervention.

Written Expression

IDEA 2004 specifies a broad category for written expression, which in the research literature involves either transcription (the mechanical act of putting together letters to make words) or composition (essay or story writing). The two are closely linked because automaticity of transcription is important for composition.

Handwriting and Spelling

Difficulties with handwriting and spelling can affect essay composition, highlighting the complex and interrelated nature of the writing task. Most norm-referenced tests include a measure of spelling, which may represent the primary source of difficulty in written expression for many children with word reading difficulties. An analysis of spelling errors may help identify whether the spelling problem is related to underdeveloped phonological awareness or with the student's knowledge of English orthography. Spelling tests, like any writing task, can also be used as an informal assessment of handwriting.

Written Expression

Norm-referenced assessments also typically include an assessment of written expression. These tests vary in their composition requirement, but all are designed to evaluate how well the student is able to express himself in text. This sort of assessment is typically needed only for students who do not demonstrate significant problems with basic writing skills. Students who struggle to transcribe or who have motoric difficulties associated with ADHD and other disorders will struggle with this task, and their score might not be a true reflection of their ability to create a written composition if they were to use a keyboard, for example. For students who demonstrate higher-level composition difficulties, writing a composition is key, which is required by the Essay Composition subtest of the Wechsler Individual Achievement Test. An alternative is the Spontaneous Writing subtests of the Test of Written Language.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Writing Fluency

Automaticity in writing and typing skills is critical for writing success. Therefore, assessing writing fluency with tests like the Sentence Writing Fluency subtest of the Woodcock-Johnson, the Alphabet Writing Fluency subtest of the Wechsler Individual Achievement Test, or the Writing Fluency subtest of the Kaufman Tests of Educational Achievement may be useful. However, such tests are meaningful only for students who do not have problems with the basic motor task.

Planning for Writing Intervention

There are well established methods for teaching transcription (handwriting and spelling). The strongest evidence for programs involving composition is self-regulated strategy development, which teaches strategies for compositing and editing, along with organizational components. A comprehensive assessment of constituent writing skills can help in planning the amount of instruction focused on transcription versus composition skills.

Assessing and Building Automaticity

Across all academic domains, a student must develop automaticity, or fluency. The need for fluency is easily illustrated in reading development, in which some students overcome initial word reading difficulties but continue to struggle with fluency. Automaticity is critical for cognitive efficiency, but also because it allows for greater opportunities to practice academic tasks in reading, writing, and math. Many children with SLD struggle to achieve automaticity because of difficulties with basic skills—difficulties that are compounded because these students have fewer opportunities to access print, complex math, or composition writing. This effect highlights the need for interventions to include multiple quality opportunities for practice and engagement. Automaticity in all academic skills should be assessed and discussed to determine the degree to which practice and engagement need to be incorporated into the intervention and because automaticity is an excellent indicator of progress.

Evaluating Instructional Response

IDEA 2004 stipulates that students cannot be identified with SLD without evidence that they have received adequate instruction in reading and math and data demonstrating inadequate progress. These data are most efficiently collected in a schoolwide RTI framework, but other forms of assessment can meet this requirement, such as grades, the State of Texas Assessments of Academic Readiness, assessments completed in general education, progress monitoring with CBMs that are not part of an RTI framework, or even screening with norm-referenced assessments that are repeated as part of the eligibility process. Although a formal assessment of instructional response is not required, evidence of adequate instruction and insufficient progress must be documented and is an essential part of any SLD eligibility determination.

The most common method of assessing instructional response involves CBMs in reading, math, and spelling. These measures are given as serial probes and are usually time constrained. In reading, a student may be asked to read word lists or stories as quickly as possible every 1 to 4 weeks during an instructional period. Cloze or maze tasks are more closely related to comprehension but are moderately to highly correlated with word reading accuracy and fluency. In math, grade-appropriate calculations are given in a time-limited fashion. In written expression, timed spelling tests, alphabet writing tests, and other procedures are used. As previously noted, normed-referenced assessments can also be used. The critical component for identification is the student's level at the end of an intervention period or another point in the instructional period. For identification, the end point is more important than the slope or amount of change because the information on growth is contained in the end point. For modifying instruction, the slope is important.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Exclusionary Criteria

Academic difficulties may be due to other disabilities, such as a sensory problem, intellectual disability, or another pervasive disturbance of cognition, like autism spectrum disorder. These disorders have specific identification criteria and require interventions that address a much more pervasive impairment of adaptation that contrasts with the narrow impairment in adaptive skills that characterizes SLD. Additionally, contextual factors that may interfere with achievement, such as limited English proficiency, comorbid behavioral problems, and economic disadvantage, should be considered. The goal of this part of the assessment is to determine whether such a condition is a primary cause of low achievement, a comorbid condition, or a result of low achievement.

These considerations can also assist in planning for effective interventions. For example, children with ADHD who are receive interventions to address their attention and academic difficulties achieve better outcomes. Anxiety might also limit the effectiveness of standalone academic interventions. If a child struggling to read exhibits high levels of anxiety, a treatment program that addresses both reading and anxiety is critical.

Limited English proficiency is another issue that must be considered, particularly in Texas, where many children come from homes in which English is not the primary language spoken. Children who grow up in households where the language at home is different from the language of instruction are at greater risk for academic difficulties, primarily due to the difficulties associated with mastering academic content while learning a second language. Yet no clear criteria or assessments differentiate a child with achievement difficulties due to SLD from a child who demonstrates limited English proficiency. One assessment strategy is to include assessments of oral language proficiency and achievement in both languages. However, these results must also be considered in context, as many English learners attend English-only classrooms and have not received academic instruction in their first language. Parsing the interconnected issues of academic difficulties and language proficiency takes careful consideration to ensure that students are not identified with SLD simply because they lack the English proficiency to perform well on achievement tests in English.

To address all potential exclusionary factors and better plan for treatment, the comprehensive assessment should routinely include parent and teacher rating scales of behavior and academic adjustment, along with parent-completed developmental and medical history forms. These scales may identify behavioral comorbidities and historical factors (e.g., history of brain trauma) that are important to screen. If there is evidence for behavioral comorbidity, the guidelines for identifying these disorders in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) should be followed. Simply referring a child for educational interventions without identifying and treating these factors will increase the probability of a poor intervention response.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

References

Beaujean, A. A., Benson, N. F., McGill, R. J., & Dombroski, S. S. (2018). A misuse of IQ scores: Using the dual discrepancy/consistency method for identifying specific learning disabilities. *Journal of Intelligence, 6*, 36. doi:10.3390/jintelligence6030036

Benson, N. F., Beaujean, A., McGill, R. J., & Dombroski, S. C. (2018). Critique of the core-selective evaluation process. *Journal of the Texas Educational Diagnosticians' Association, 47*, 14–18.

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

Burns, M. K., Petersen-Brown, S., Haegele, K., Rodriguez, M., Schmitt, B., Cooper, M., . . . VanDerHeyden, A. M. (2016). Meta-analysis of academic interventions derived from neuropsychological data. *School Psychology Quarterly, 31*(1), 28–42.

Dehn, M. (2013). *Essentials of processing assessment* (2nd ed.). New York, NY: John Wiley.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (Eds.) (2013). *Essentials of cross-battery assessment* (3rd ed.). New York, NY: John Wiley.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2019). *Learning disabilities: From identification to intervention* (2nd ed.). New York, NY: Guilford Press.

Fletcher, J. M., & Miciak, J. (2017). Comprehensive cognitive assessments are not necessary for the identification and treatment of learning disabilities. *Archives of Clinical Neuropsychology, 32*, 2–7.

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

Fletcher, J. M., Stuebing, K. K., Barth, A. E., Denton, C. A., Cirino, P. T., Francis, D. J., & Vaughn, S. (2011). Cognitive correlates of inadequate response to reading intervention. *School Psychology Review, 40*(1), 3–22.

Flowers, L., Meyer, M., Lovato, J., Wood, F., & Felton, R. (2001). Does third grade discrepancy status predict the course of reading development? *Annals of Dyslexia, 51*(1), 49–71.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*, 98–110.

Hale, J. B., & Fiorello, C. A. (2004). *School neuropsychology: A practitioner's handbook*. New York, NY: Guilford Press.

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

Individuals with Disabilities Education Act (IDEA) regulations, 34 C.F.R. *§§ 300.3 et seq.* (2006). IDEA regulations commentary, 71 Fed. Reg. 46651, (2006, August 14).

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Kranzler, J. H., Floyd, R. G., Benson, N., Zaboski, B., & Thibodaux, L. (2016). Classification agreement analysis of cross-battery assessment in the identification of specific learning disorders in children and youth. *International Journal of School & Educational Psychology, 4*(3), 124–136.

Macmann, G. M., & Barnett, D. W. (1985). Discrepancy score analysis: A computer simulation of classification stability. *Journal of Psychoeducational Assessment, 3*(4), 363–375.

Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the verbal-performance dichotomy of the Wechsler scales. *School Psychology Quarterly, 9*(3), 161–197.

Macmann, G. M., Barnett, D. W., Lombard, T. J., Belton-Kocher, E., & Sharpe, M. N. (1989). On the actuarial classification of children: Fundamental studies of classification agreement. *The Journal of Special Education, 23*(2), 127–149.

Messick, S. (1986). The once and future issues of validity: Assessing the meaning and consequences of measurement. *ETS Research Report Series, 1986*(2).

Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly, 29*(1), 21–37.

Miciak, J., Taylor, W. P., Denton, C. A., & Fletcher, J. M. (2015). The effect of achievement test selection on identification of learning disabilities within a patterns of strengths and weaknesses framework. *School Psychology Quarterly, 30*(3), 321–334.

Miciak, J., Taylor, W. P., Stuebing, K. K., & Fletcher, J. M. (2018). Simulation of LD identification accuracy using a pattern of processing strengths and weaknesses method with multiple measures. *Journal of Psychoeducational Assessment, 36*(1), 21–33. doi:10.1177/0734282916683287

Miciak, J., Williams, J. L., Taylor, W. P., Cirino, P. T., Fletcher, J. M., & Vaughn, S. (2016). Do processing patterns of strengths and weaknesses predict differential treatment response? *Journal of Educational Psychology, 108*(6), 898–909.

Naglieri, J. A. (1999). *Essentials of CAS assessment*. New York, NY: John Wiley & Sons.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system*. Chicago, IL: Riverside.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*(3), 181–197.

Schneider, W. J., & Kaufman, A. S. (2017). Let's not do away with comprehensive cognitive assessments just yet. *Archives of Clinical Neuropsychology, 32*, 8–20.

Schultz, E. K., & Stephens, T. L. (2015). Core-selective evaluation process: An efficient and comprehensive approach to identify students with SLD using the WJ-IV. *The Dialog, 44*(2), 5–12.

Schultz, E. K., & Stephens, T. L. (2017). *Core-selective evaluation process (C-SEP): Overview and procedures, Texas edition*. Authors.

Simos, P. G., Fletcher, J. M., Rezaie, R., & Papanicolaou, A. C. (2014). Does IQ affect the functional brain network involved in pseudoword reading in students with reading disability? A magnetoencephalography study. *Frontiers in Human Neuroscience, 7*, 932.

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International

Shepard, L. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. *The Journal of Special Education, 14*(1), 79–91.

Stuebing, K. K., Barth, A. E., Molfese, P. J., Weiss, B., & Fletcher, J. M. (2009). IQ is not strongly related to response to reading instruction: A meta-analytic interpretation. *Exceptional Children, 76*, 31–51.

Stuebing, K. K., Fletcher, J. M., Branum-Martin, L., & Francis, D. J. (2012). Evaluation of the technical adequacy of three methods for identifying specific learning disabilities based on cognitive discrepancies. *School Psychology Review, 41*, 3–21.

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*, 469–518.

Tanaka, H., Black, J., Hulme, C., Leanne, S., Kesler, S., Whitfield, G., . . . Hoeft, F. (2011). The brain basis of the phonological deficit in dyslexia is independent of IQ. *Psychological Science, 22*, 1442–1451.

Taylor, W. P., Miciak, J., Fletcher, J. M., & Francis, D. J. (2017). Cognitive discrepancy models for specific learning disabilities identification: Simulations of psychometric limitations. *Psychological Assessment, 29*(4), 446–457.

Texas Education Agency. (2017). *2016–2017 special education reports*. Electronic resource accessed August 1, 2017, at https://rptsvr1.tea.texas.gov/cgi/sas/broker

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

© 2019 The University of Texas System/Texas Education Agency
Licensed under Creative Commons BY-NC-ND 4.0 International