CHAPTER 1

Introduction

Brian, a second-grade student at Salter Elementary School, was referred to the school psychologist for evaluation. The request for evaluation from the multidisciplinary team noted that he was easily distracted and was having difficulty in most academic subjects. Background information reported on the referral request indicated that he was retained in kindergarten and was on the list this year for possible retention. As a result of his difficulties sitting still during class, his desk has been removed from the area near his peers and placed adjacent to the teacher's desk. Brian currently receives supplemental (i.e., Tier 2) intervention in mathematics.

Susan was in the fifth grade at Carnell Elementary School. She had been in a self-contained classroom for students with learning disabilities since second grade and was currently doing very well. Her teacher referred her to determine her current academic status and potential for increased inclusion.

Jorgé was in the third grade at Moore Elementary School. He was referred by his teacher because he was struggling in reading and had been receiving support through the school's emergent bilingual program for the past year since arriving from Puerto Rico. Jorgé's teacher was concerned that he was not achieving the expected level of progress compared to other students with similar backgrounds.

Dell was a third-grade student at Post Oak Elementary. He was referred for an evaluation for special educational eligibility by his teacher because of persistent reading difficulties and her observation that "he doesn't understand anything he reads."

All of these cases are samples of the many types of referrals for academic problems faced by school personnel. How should the team proceed to conduct the evaluations? The answer to this question lies in how the problems are conceptualized. Years ago, a school psychologist would administer an individual intelligence test (usually, the Wechsler Intelligence Scale for Children—Fourth Edition [WISC-IV; Wechsler, 2003]), an individual achievement test (such as the Peabody Individual Achievement Test—Revised/Normative Update [PIAT-R/NU; Markwardt, 1997], or the Wechsler Individual Achievement

Test—Second Edition [WIAT-II; Wechsler, 2001]), and a test of visual–motor integration (usually, the Bender–Gestalt). In the past, and still today in some situations, the psychologist might add some measure of personality, such as projective drawings. Currently, it is common to see school psychologists administer a large battery of subtests that measure cognitive processes assumed to underlie academic skills, and based on the magnitude of the differences in the student's scores across subtests (i.e., their "pattern of strengths and weakness"), determine eligibility for special education. Other professionals, such as educational consultants or educational diagnosticians, might assess the child's specific academic skills by administering norm-referenced achievement tests, such as the Woodcock–Johnson Psycho-Educational Battery—Revised (Woodcock et al., 2001), the KeyMath—3 Diagnostic Assessment (KeyMath-3; Connolley, 2007), or other instruments. Based on these test results, a determination of eligibility (in the cases of Brian, Jorgé, and Dell) or evaluation of academic performance (in the case of Susan) would be made.

When Brian was evaluated in this traditional way, the results revealed that his scores on tests of academic achievement were not discrepant enough from his scores on tests of intellectual and cognitive functioning, and therefore he was not eligible for special education. Not surprisingly, Brian's teacher requested that the multidisciplinary team make some recommendations for remediating his skills. From this type of assessment, the team found it very difficult to make specific recommendations. The team suggested that because Brian was not eligible for special education, he was probably doing the best he could in his current classroom. They did note that his decoding skills appeared weak and recommended that some consideration be given to switching him out of a phonics-based approach to reading.

When Susan was assessed, the data showed that she was still performing substantially below grade levels in all academic areas. Despite having spent the last 3 years in a self-contained classroom for students with learning disabilities, Susan had made minimal progress when compared to peers of similar age and grade. As a result, the team decided not to increase the amount of time that she be included in general education for academic subjects.

When Jorgé was assessed, the team also administered measures to evaluate his overall language development in addition to a standard psychoeducational test battery. Specifically, the Woodcock–Muñoz Language Survey—Third Edition (WMLS-III; Woodcock et al., 2017) was given to assess his English and Spanish language proficiency in academically related tasks such as listening, speaking, reading, and writing. The data showed that his poor reading skills were a function of less than expected development in English, rather than a general underdeveloped ability in Spanish. Jorgé was therefore not considered eligible for special education services other than programs for second language learners.

Dell, the third grader evaluated for a learning disability in reading, was assessed using a large battery of cognitive skills tests and standardized tests of achievement. Despite significant reading difficulties evident in the assessment, he did not demonstrate a specific pattern of strengths and weaknesses in cognitive skills that, under the school district's model of identification, would make him eligible for special education. Based on the test results, it was determined that he did not have a learning disability and therefore did qualify for special education services. Dell's school did not offer supplemental interventions for students that were not formally identified with a disability; therefore, Dell remained in general education where his teacher did the best they could to support him.

In contrast to viewing the referral problems of Brian, Susan, Jorgé, and Dell only from the question of diagnosis or identification, one could also conceptualize their referrals as questions of "which intervention strategies would be likely to improve their academic skills?" Seen in this way, assessment becomes a problem-solving process and involves a very different set of methodologies. First, to identify intervention strategies, one must have a clear understanding of the child's accuracy and fluency with skills that are foundational to the achievement domain deemed to be a "problem," the student's mastery of skills that have already been taught, the rate at which learning occurs when the child is taught appropriate to their achievement level, and a thorough understanding of the instructional environment in which instruction has occurred. To do this, one must be able to select measures of academic achievement that are relevant to (1) the student's areas of difficulty and (2) skills targeted in instruction. The assessment process must be dynamic and evaluate how the child progresses across time when effective instruction and sufficient practice opportunities are provided. Such an assessment requires an understanding of the developmental nature of academic skills, the precursor skills that are foundational to the development of proficiency within and across academic skill domains, and how to identify specific skill deficits that may explain broader academic difficulties. An understanding is required of the behaviors that facilitate learning, such as task engagement, and the instructional environments that best occasion these skills. The evaluator must understand how to identify measures that are sensitive to the impact of instructional interventions and inform subsequent decision making. A clear understanding of the student's academic difficulties and instructional ecology (i.e., the classroom environment and the variables within it that promote or impede learning) is most clearly achieved through methods of direct assessment, direct observation, teacher and student interviews, and examination of student-generated products. When the assessment is conducted from this perspective, the results are more directly linked to developing effective intervention strategies that are customized to the individual needs of the student.

When Brian was assessed in this way, it was found that he was appropriately placed in the curriculum materials in both reading and math. A lack of fluency in basic addition and subtraction facts was identified, which indicated the need for practice strategies to build his automaticity. Specific problems in spelling and written expression were noted, and specific recommendations for instruction in capitalization and punctuation were made. The assessment also revealed that Brian was not receiving enough opportunities to practice reading new word types with feedback from the teacher as they were taught, and not enough time to practice reading text. Thus, his difficulties with decoding were not a reason to stop a phonics-based reading program, but to increase explicit instruction in reading longer and more complex words and provide more opportunities to practice reading with feedback from a teacher. Moreover, the assessment team members suggested that Brian's seat be moved from next to the teacher's desk to sitting among his peers because the data showed that he really was not as distractible as the teacher had indicated. When these interventions were implemented, Brian showed gains in performance in reading, writing, and mathematics that equaled those of his general education classmates.

The results of Susan's direct assessment were more surprising and in direct contrast to the traditional evaluation. Although it was found that Susan was appropriately placed in the areas of reading, mathematics, and spelling, an examination of her skills in the curriculum showed that she probably could be successful in the general education classroom and that she would benefit from a reading intervention that her teacher provided to three other students in her classroom with similar reading skill needs. In particular, it was found that she had attained fifth-grade mathematics skills within the curriculum, despite scoring below grade level on the standardized test. When her reading group was adjusted, Susan's data over a 13-week period showed that she was making the same level of progress as her fifth-grade classmates without disabilities.

Jorgé's assessment was also a surprise. Although his lack of familiarity with English was evident, Jorgé showed that he was successful in learning, comprehending, and reading when the identical material was presented in Spanish. In fact, Jorge's comprehension of spoken English was comparable to his comprehension of written English, and it was determined that Jorgé was much more successful in learning to read in English once he was able to read the same material in Spanish. In Jorgé's case, monitoring his reading performance in both English and Spanish showed that he was making slow but consistent progress. Although he was still reading English at a level equal to that of first-grade students, goals were set for Jorgé to achieve a level of reading performance similar to middle second grade by the end of the school year. Data collected over that time showed that at the end of the third grade, he was reading at a level similar to students at a beginning second-grade level, having made almost 1 year of academic progress over the past 5 months.

For Dell, the third grader with significant reading difficulties, consider an alternative history in which Dell had attended Rushing Springs Elementary School since kindergarten. This school implemented a schoolwide, multi-tiered system of support (MTSS). In this model, assessment is part of a dynamic, ongoing effort at providing high-quality instruction and intervention intensification. All students in the school are assessed periodically during the school year to identify those who may not be achieving at levels consistent with grade-level expectations. Students who are below grade-level benchmarks are provided with supplemental intervention designed to address their skill needs. The intervention is provided beyond the core instruction offered to all students, and Dell's progress is carefully monitored in response to the interventions. Students who still are struggling despite this added intervention are provided with a more intensive level of intervention. Student eligibility for special education services is determined through a decision-making process called response to intervention (RTI), embedded within the MTSS framework. RTI considers the progress students make in evidence-based interventions as a component of an evaluation for special education eligibility.

For Dell in this alternative history, the multidisciplinary team was alerted to his reading difficulties much earlier than third grade. Data from universal screening measures at the beginning of first grade indicated that he was below expected levels in early reading skills. From September through December of first grade, Dell received an additional 30 minutes of intervention (Tier 2) that specifically targeted skills that were low at the fall assessment. Dell's progress during the intervention was monitored, which revealed that despite this additional level of intervention, he was still not making progress at a rate that would close the achievement gap between him and his same-grade classmates. From January through March, Dell received an even more intensive level of intervention more focused on his specific needs (Tier 3). The ongoing progress monitoring revealed that Dell was still not making enough progress, despite the increased intensity of instruction. As a result, the team decided that Dell should be evaluated to determine if he was eligible for special education services. The school psychologist conducting the comprehensive evaluation used Dell's data from (1) universal screening, (2) progress monitoring during the intervention intensification process, (3) other measures of specific academic skills relevant to understanding Dell's reading difficulties, and (4) evaluation of the classroom environment. All of these data would contribute to decisions regarding Dell's eligibility for special education services.

In Dell's case, unlike the previous cases, the focus of the evaluation was not on a diagnostic process but on a problem-solving process. Dell presented with an achievement profile and rate of learning that, if allowed to continue, would have led to significant reading deficits. These difficulties would have negative implications across other academic domains, including mathematics (e.g., word-problem solving), writing, and achievement in science and social studies. The MTSS framework allowed for intervention to be implemented as soon as academic difficulties were observed, the data collected through the process drove the selection of interventions, and the outcomes of the intervention process informed the next steps in the evaluation process. One possibility is that when Dell was found to not have responded to instructional intervention offered through general education and supplemental support as part of the MTSS model, he might then be considered as potentially eligible for the most intensive level of service: that of a student identified as having a specific learning disability.

THE FOCUS OF THIS TEXT

This text provides a direct assessment and intervention methodology for the evaluation and remediation of academic problems with students like Dell. Specifically, the text includes detailed descriptions of conducting a behavioral assessment of academic skills (as developed by Shapiro, 1989, 1990, 2004, 2011; Shapiro & Lentz, 1985, 1986), assessment frameworks that carefully consider relevant skills and alignment of measures to the curriculum of instruction (Hosp et al., 2014; Howell et al., 1993), evidence-based interventions focused on teaching skills critical for academic success, and interpretation and integration of data collected as part of the intervention process for determining the need for intervention intensification.

BACKGROUND, HISTORY, AND RATIONALE FOR ACADEMIC ASSESSMENT AND INTERVENTION

The percentage of children who consistently experience academic problems has been of continued concern to school personnel. Over the past two decades, the national percentage of students receiving special education services has increased. According to the National Center for Education Statistics (NCES, 2022), in the 2000–2001 school year, 6.3 million children in the United States between 3 and 21 years of age (13% of total school enrollment) were identified as eligible for special education services. By the 2020–2021 school year, the number increased to 7.2 million (15% of school enrollment). Students identified as having learning disabilities comprise the largest classification of those receiving special education, making up approximately 33% of those classified and receiving special education services in 2020–2021, followed by speech and language impairment at 19% (which also plays a role in academic difficulties and learning disability).

The past 30 years have been marked by considerable debate regarding the methods used to identify students as having a learning disability. In particular, the early 2000s saw significant challenges to the validity and practical utility of discrepancy formulae (i.e., making eligibility decisions based on the discrepancies between attained scores on intelligence and achievement tests; Fletcher et al., 2003; Francis et al., 1996, 2005; Peterson & Shinn, 2002; Sternberg & Grigorenko, 2002; Stuebing et al., 2002). Debate persists regarding the role and relevance of tests of cognitive processing in learning disability identification (e.g., Fletcher & Miciak, 2017; Schneider & Kaufman, 2017).

Among alternative approaches to identification, RTI is legally permitted through the 2004 passage of the Individuals with Disabilities Education Act (IDEA). This method requires identifying the degree to which a student responds to academic interventions that are known to be highly effective. Considered in the context of other assessment data and information, students who do not respond positively to such interventions would be considered as potentially eligible for special education services related to learning disability (Fletcher et al., 2002; Fletcher et al., 2019; Fletcher & Vaughn, 2009; Gresham, 2002). Evidence indicates that, when implemented appropriately, RTI-based decision making can improve the rates at which students who are genuinely in need of special education services are identified (Burns & VanDerHeyden, 2006; O'Connor et al., 2013; VanDerHeyden et al., 2007). However, concerns persist regarding schools' ability to implement core instruction, invention, and assessment effectively that make RTI decisions possible (Balu et al., 2015; D. Fuchs & L. Fuchs, 2017; Ruffini et al., 2016), and inconsistency between school districts and state-level policies on procedures and implementation (Hudson & McKenzie, 2016). Others are not convinced that the RTI approach will ultimately improve the identification of students with learning disabilities (O'Connor et al., 2017; Scruggs & Mastropieri, 2002; Kavale & Spaulding, 2008; Reynolds & Shaywitz, 2009). Nevertheless, directly measuring a difficulty to learn in situations where most children are successful is closer to the idea of *learning disability*, in contrast to models that only consider cognitive and achievement discrepancy at a single point in time (Fletcher et al., 2019).

Data have shown that academic skills problems are consistently a major focus of referrals for evaluation. With a nationally representative sample of school psychologists, Benson et al. (2019) found that 25% of all reported referral concerns were for suspected learning disabilities, which was at least twice as frequent as any other referral concern. Bramlett and colleagues (2002) examined the patterns of referrals made for school psychological services. Surveying a national sample of school psychologists, they ranked academic problems as the most frequent reasons for referral, with 57% of total referrals made for reading problems, 43% for written expression, 39% for task completion, and 27% for mathematics. These rates are historically stable; Ownby et al. (1985) examined the patterns of referrals made for school psychological services within a small school system (school population = 2,800). Across all grade levels except preschool and kindergarten (where few total referrals were made), referrals for academic problems exceeded referrals for behavior problems by almost 5 to 1.

Clearly, there are significant needs for effective assessment and intervention strategies to address academic problems among school-age children. Indeed, the number of commercially available standardized achievement tests (e.g., Salvia et al., 2016) suggests that evaluation of academic progress has been a longstanding need among educators. Over the years, survey studies have observed an increase in the use of measures of academic achievement (Benson et al., 2019; Goh et al., 1981; Hutton et al., 1992; Stinnett et al., 1994; Wilson & Reschly, 1996). Most recently, the survey data published by Benson and colleagues (2019) indicated that five of the ten most administered assessment instruments by school psychologists on a monthly basis were measures of academic skills (three of which were curriculum-based measures). These recent data also noted an increase in the use of behavior rating scales and a marked decrease in the use of projective personality tests in contrast to previous survey studies.

Despite the historically strong concern about assessing and remediating academic problems, controversy remains regarding the most effective methods for conducting useful assessments and choosing the most effective intervention strategies. Historically, educational professionals have expressed dissatisfaction with commercially available, norm-referenced tests (e.g., Donovan & Cross, 2002; Heller et al., 1982; Hively & Reynolds, 1975; Wiggins, 1989). Likewise, strategies that attempt to remediate deficient learning processes identified by these measures have historically not been useful in effecting change in academic performance (e.g., Arter & Jenkins, 1979; Good et al., 1993). However, the landscape of academic assessment has changed a great deal in the time since the first edition of this text was published in 1989. A greater understanding of how academic skills like reading, mathematics, and written expression develop (and how they break down) has led to the development of commercially available, norm-referenced tests that more directly measure relevant skills (e.g., the Comprehensive Test of Phonological Processing; Wagner et al., 2013), as well as new subtests added to longstanding broadband measures of academic skills. Curriculum-based measures are now readily available and are routinely used by many school psychologists (Benson et al., 2019). The question is no longer about whether school psychologists should or should not use a certain class of assessments, but rather how to select an ideal set of measures that will (1) provide the best assessment of a student's academic strengths and difficulties, (2) help identify the reason(s) for their skill difficulties, and (3) point to how to best intervene.

ASSESSMENT AND DECISION MAKING FOR ACADEMIC PROBLEMS

Salvia et al. (2007) defined assessment as "the process of collecting data for the purpose of (1) specifying and verifying problems, and (2) making decisions about students" (p. 5). They identified five types of decisions that can be made from assessment data: referral, screening, classification, instructional planning, and monitoring of pupils' progress. They also add that decisions about the effectiveness of programs (program evaluation) can be made from assessment data.

Not all assessment methodologies for evaluating academic behavior equally address each type of decision needed. For example, some norm-referenced instruments may be useful for understanding a student's level of achievement compared to peers of the same grade or age, but may not be valuable for decisions regarding instructional programming. Likewise, criterion-referenced tests that offer intrasubject comparisons may be useful in identifying a student's relative strengths and weaknesses but may not be sensitive to monitoring student progress within a curriculum. Methods that use frequent, repeated assessments may be valuable tools for monitoring progress but may not offer sufficient data on identifying the cause of a student's academic problem. Clearly, the use of a particular assessment strategy should be linked to the type of decision one wishes to make, and no specific measure type is likely to be sufficient. A framework that can be used across measures and types of decisions would be extremely valuable.

The various types of decisions described by Salvia et al. (2007) should require the collection of different kinds of data. Interesting trends have been observed over the years in the types of assessment data that are commonly collected. In an early study, Goh et al. (1981) reported data suggesting that regardless of the reason for referral, most school psychologists administer an individual intelligence test, a general test of achievement, a test of perceptual–motor performance, and a projective personality measure. A replication of the Goh et al. study 10 years later (Hutton et al., 1992) found that little had changed. Psychologists still spent more than 50% of their time engaged in assessment.

Hutton et al. (1992) reported that the emphasis on intelligence tests noted by Goh et al. (1981) had lessened, whereas the use of achievement tests had increased. Hutton et al. (1992) also found that the use of behavior rating scales and adaptive behavior measures had increased somewhat. Stinnett et al. (1994), as well as Wilson and Reschly (1996), again replicated the basic findings of Goh et al. (1981). In a survey of assessment practice, Shapiro and Heick (2004) observed some shifting of assessment practices in relation to students referred for behavior disorders toward the use of measures such as behavior rating scales and systematic direct observation. Shapiro et al. (2004) also found some self-reported movement of school psychologists toward the use of curriculum-based assessment (CBA) measures when the referral was for academic skills problems. Even so, almost 47% of those surveyed reported not using CBA in their practice. Similarly, Lewis et al. (2008), as part of a national telephone survey of practicing school psychologists, examined the self-reported frequency of functional behavioral assessment (FBA) and CBA in their practice over the past year. Outcomes of their survey found that between 58 and 74% reported conducting fewer than 10 FBAs in the year of the survey, and between 28 and 47% of respondents reported using CBA.

Trends toward greater use of measures that more directly assess a skill of interest have continued. Benson et al. (2019), in their survey of a nationally representative sample of practicing school psychologists, observed a significant increase compared to previous surveys in practitioners' routine use of curriculum-based measures, behavior observations, behavior rating scales, problem-solving interviews, and functional assessment interviews. Concurrently, increased use of these methods was associated with decreased use of projective tests and tests of psychomotor functioning. However, Benson et al. found that tests of intellectual and cognitive functioning remain common. Thus, across nearly 40 years, assessment practices changed dramatically since the findings of Goh et al. (1981) and moved away from high-inference projective tests toward methods that more closely evaluate students' behavior and achievement as a function of their environment and instruction. Nevertheless, comprehensive tests of cognitive and intellectual functioning remain common in evaluations of learning difficulties, despite evidence that their use is not necessary (and can lead to problems) in identifying learning disabilities (Fletcher & Miciak, 2017).

This chapter provides an overview of the conceptual issues of academic assessment and remediation. The framework on which behavioral assessment and intervention for academic problems are based is described. First, however, the current state of academic assessment and intervention is examined.

TYPES OF INDIVIDUAL ASSESSMENT METHODS

Norm-Referenced Tests

One of the most common methods for evaluating individual academic skills involves administering commercially available, norm-referenced, standardized tests of achievement. For simplicity, we refer to this class of tests as "standardized" tests of achievement.[1]

[1]Throughout the text, we use the term *standardized tests* to refer to published, commercially available, norm-referenced, standardized tests. We acknowledge that this term is imprecise; standardization simply refers to establishing a set of administration and scoring procedures; therefore, any test can be standardized. Additionally, measures not commercially available can be norm-referenced. Nevertheless, *standardized* tends to be associated most often with the class of commercially available measures described in this section, and thus we use the term for simplicity.

Examples of commonly used standardized tests of achievement include batteries of multiple skills such as the Woodcock–Johnson IV Tests of Achievement (WJ IV; Schrank et al., 2014), or the Wechsler Individual Achievement Test—Third Edition (WIAT-III; Wechsler, 2009), which include subtests that assess reading, mathematics, spelling, writing, and provide overall scores for each skill domain. Other standardized norm-referenced tests are designed to assess more specific skills, such as the Test of Word Reading Efficiency—Second Edition (TOWRE-2; Torgesen et al., 2012), KeyMath—3 Diagnostic Assessment (KeyMath-3; Connolley, 2007), and the Test of Written Language—Fourth Edition (TOWL-4; Hammill & Larsen, 2009).

One of the primary purposes of standardized tests is to determine an individual's "relative standing," in other words, where a student's level of performance stands in relation to other individuals of the same age or grade. Scores on the test are derived by comparing the scores of the child being tested to scores obtained across a large sample of children representing a full range of achievement. Standard scores and percentile ranks are used to describe the relative standing of the target child in relation to other students of the same age or grade in the normative sample (e.g., above average, average, below average, well below average).

On a test appropriate for a given age or grade level, the performance of individuals in the normative sample is expected to be "normally distributed." This means that scores of most individuals will cluster in the middle of the range around the mean (i.e., average), with increasingly smaller numbers of individuals scoring at the upper end (higher achievers) and lower end (lower achievers) of the distribution. This distribution is assumed to approximate the overall population of individuals from which the normative sample was drawn. Thus, normative data give the assessor a reference point for identifying the degree to which the responses of a specific student differ significantly from those of the average same-age/same-grade peer. In other words, a comparison of an individual's achievement to a normative sample indicates the size of the difference between a student's performance on the skill and how most other students perform.

Despite the popular and widespread use of standardized tests for assessing individual academic skills, some issues limit their usefulness. Suppose a test is purports to evaluate a student's acquisition of knowledge. In that case, the test should assess the knowledge student should be expected to demonstrate given the content and curriculum used for instruction. If there is little overlap between what the student is expected to know on the test and what the student was taught, a child's low score on the measure may not necessarily reflect a failure to learn. Instead, the child's score may be due to the test's poor relation to the curriculum the child was provided, or content that is not representative of skills important for the academic domain the test was designed to measure.

In the 1980s and early 1990s, Shapiro and several other scholars in school psychology and special education identified problems with the lack of alignment between what students were taught and what they were tested on in commercial, norm-referenced tests of achievement. In a replication and extension of the work of Jenkins and Pany (1978), Shapiro and Derr (1987) examined the degree of overlap between five commonly used basal reading curricula and four commercial, standardized achievement tests. At each grade level (first through fifth), the number of words appearing on each subtest and in the reading curricula was counted. The resulting score was converted to a standard score (M = 100, SD = 15), percentile, and grade equivalent, using the standardization data provided for each subtest. The results of this analysis are reported in Table 1.1. Across subtests and reading series, there appeared to be little overlap between the words appearing in the series and on the tests.

TABLE 1.1. Overlap between Basal Reader Curricula and Tests

	PIAT				WRAT-R				K-TEA				WRM			
	RS	GE	%tile	SS	RS	GE	%tile	SS	RS	GE	%tile	SS	RS	GE	%tile	SS
Ginn-720																
Grade 1	23	1.8	58	103	40	IM	47	99	14	1.6	37	95	38	1.8	38	96
Grade 2	28	2.8	50	100	52	2M	39	96	23	2.6	42	97	69	2.5	33	94
Grade 3	37	4.0	52	101	58	2E	27	91	27	3.2	32	93	83	3.0	24	90
Grade 4	40	4.4	40	96	58	2E	16	85	27	3.2	16	85	83	3.0	10	81
Grade 5	40	4.4	25	90	61	3B	12	82	28	3.4	9	80	83	3.0	4	74
Scott, Foresman																
Grade 1	20	1.4	27	91	39	IM	41	97	12	1.5	27	91	33	1.8	30	92
Grade 2	23	1.8	13	83	44	IE	16	85	17	1.9	18	86	63	2.3	27	91
Grade 3	23	1.8	7	78	46	2B	4	73	17	1.9	5	70	63	2.3	9	80
Grade 4	23	1.8	3	72	46	2B	1	67	17	1.9	2	70	63	2.3	2	70
Grade 5	23	1.8	1	65	46	2B	.7	59	17	1.9	1	66	63	2.3	.4	56
Macmillan-R																
Grade 1	23	1.8	58	103	35	IB	30	92	13	1.6	32	93	42	1.9	44	98
Grade 2	24	2.0	20	87	41	IM	10	81	18	2.0	19	87	58	2.2	22	89
Grade 3	24	2.0	9	80	48	2B	5	76	21	2.3	12	82	66	2.4	10	81

Grade 4	24	2.0	4	74	48	2B	2	70	21	2.3	5	75	67	2.5	3	72
Grade 5	24	2.0	2	69	50	2M	1	65	21	2.3	2	70	67	2.5	2	69
Keys to Reading																
Grade 1	24	2.0	68	107	41	1M	50	100	15	1.7	42	97	42	1.9	44	98
Grade 2	28	2.8	50	100	51	2M	37	95	20	2.2	27	91	68	2.5	33	94
Grade 3	35	3.8	47	99	59	3B	30	92	24	2.7	19	87	84	3.1	26	91
Grade 4	35	3.8	26	91	59	3B	18	86	24	2.7	9	80	84	3.1	11	82
Grade 5	35	3.8	14	84	59	3B	8	79	25	2.8	5	76	84	3.1	4	74
Scott, Foresman—Focus																
Grade 1	23	1.8	58	103	35	IB	30	92	13	1.6	32	93	37	1.8	35	94
Grade 2	25	2.2	28	91	46	2B	21	88	17	1.9	18	86	56	2.1	20	89
Grade 3	27	2.6	21	88	49	2B	6	77	20	2.2	10	81	68	2.5	11	82
Grade 4	28	2.8	11	82	54	2M	8	79	22	2.4	6	77	76	2.8	7	78
Grade 5	28	2.8	6	77	55	2E	4	73	24	2.7	4	64	81	2.9	3	72

Note. The grade-equivalent scores "B, M, E" for the WRAT-R refer to the assignment of the score to the beginning, middle, or end of the grade level. RS, raw score; GE, grade equivalent; SS, standard score (M = 100; SD = 15); PIAT, Peabody Individual Achievement Test; WRAT-R, Wide Range Achievment Test—Revised; K-TEA, Kaufman Test of Educational Achievement; WRM, Woodcock Reading Mastery Test. From Shapiro and Derr (1987, pp. 60–61). Copyright © 1987 PRO-ED, Inc. Reprinted by permission.

Although these results suggest that the overlap between what is taught and what is tested on reading subtests is questionable, the data examined by Shapiro and Derr (1987) and by Jenkins and Pany (1978) were hypothetical. Good and Salvia (1988) and Bell and colleagues (1992) provided evidence that with actual students evaluated on standardized norm-referenced achievement measures, there is inconsistent overlap between the basal reading series employed in their studies and the different measures of reading achievement.

In the Good and Salvia (1988) study, 65 third- and fourth-grade students who were all being instructed in the same basal reading series (Allyn & Bacon's Pathfinder Program, 1978) were administered four standardized norm-referenced reading subtests popular at the time. Their analysis showed significant differences in test performance for the same students on different reading tests, predicted by the test's content.

Using a similar methodology, Bell et al. (1992) examined the content validity of three standardized tests of word reading: the Reading Decoding subtest of the Kaufman Test of Educational Achievement (KTEA), the Reading subtest of the Wide Range Achievement Test—Revised (WRAT-R; Jastak & Wilkinson, 1984), and the Word Identification subtest of the Woodcock Reading Mastery Tests—Revised (WRMT-R; Woodcock, 1987). All students (n = 181) in the first and second grades of two school districts were administered these tests. Both districts used the Macmillan-R reading series (Smith & Arnold, 1986). Results showed dramatic differences across tests when a word-by-word content analysis (as conducted by Jenkins & Pany, 1978) was conducted. Perhaps more importantly, significant differences were evident across tests for students within each grade level. For example, students in one district obtained an average standard score of 117.19 (M = 100, SD = 15) on the WRMT-R and a score of 102.44 on the WRAT-R, a difference of a full standard deviation.

Problems of overlap between test and text content are not limited to reading. For example, Shriner and Salvia (1988) examined the curriculum overlap between two elementary mathematics curricula and two commonly used individual norm-referenced standardized tests (KeyMath and Iowa Tests of Basic Skills) across grades 1–3. Hultquist and Metzke (1993) examined the overlap across grades 1–6 between standardized measures of spelling performance (subtests from the KTEA, Woodcock–Johnson—Revised [WJ-R], PIAT-R, Diagnostic Achievement Battery—2, Test of Written Spelling—2) and three basal spelling series, as well as the presence of high-frequency words. An assessment of the correspondence for content and the type of learning required revealed a lack of content correspondence at all levels in both studies.

Caveats and Cautions to Interpreting the Studies on Standardized Test and Curriculum Overlap

The studies described above have appeared in this text since its early editions, published during a time when criticism of standardized norm-referenced tests of achievement was strongest. This critique was most often based on a lack of direct correspondence between the items on a test and whether those specific items appeared in the curriculum used for instruction. There is certainly a rationale for the criticism; one should always try to test what one teaches. The previous work in this area was valuable in helping educators and test developers focus on the most relevant test content. However, theory and evidence that emerged in the 30 years since the first edition of this text have indicated that, in some skill areas such as reading, a lack of extensive and specific overlap between a test and the curriculum is not as problematic as once suggested.

Studies that focused on the overlap between tests of word reading skills and reading curricula (e.g., Bell et al., 1992; Good & Salvia, 1988; Jenkins & Pany, 1978; Shapiro & Derr, 1985) were based on a problematic notion that a test was more valid for assessing reading skills if the specific words in the test also appeared in the reading curriculum the student received. The problem with this perspective is the faulty assumption that reading development is solely the product of instruction, and that one's ability to read a word depends on having been taught that specific word. This is clearly not the case. No reading curricula could ever hope to target every single word that students will encounter in print, nor is that even necessary. Share's self-teaching hypothesis and research on it (Share, 1995, 2008), as well as connectionist models of reading acquisition (Harm & Seidenberg, 1999; Seidenberg, 2005, 2017), illuminated how children develop the ability to read thousands of words without having been exposed to instruction in every single word. Word reading is about learning a code. As discussed in Chapters 2 and 6, effective reading instruction involves directly teaching how sounds connect to letters and letter combinations, using that information to decode words, and providing opportunities for students to read lots of *types* of words and texts while providing affirmative and corrective feedback. Equipped with those skills and provided extensive opportunities to read, typically developing readers become better readers the more they read. Their strong foundational skills allow them to read words they have never seen in print based on their knowledge of spelling–sound correspondence and pronunciation of letter combinations (and help on some occasions from vocabulary and reading comprehension). They have "cracked the code"; their development of reading proficiency was due more to *how* they were taught to read words, not the specific words that appeared in a curriculum. Struggling readers and students with learning disabilities commonly experience difficulties with the basic skills that facilitate word reading acquisition, and they often have problems generalizing basic decoding skills to words or similar spelling patterns that were not targeted in instruction. These are hallmarks of reading disability. Therefore, a typically developing reader will perform better on a standardized test of word reading than a student with a reading problem, regardless of the proportion of words that overlap between the test and the curriculum materials they received. To be useful, a test should measure the types of words a student should reasonably be expected to read; a lack of overlap between the specific words on the test and those that appeared in the curriculum of instruction is not a reason to say a test lacks validity or has no utility for understanding a student's reading difficulties.

The issue applies to other academic areas as well. Spelling is similar to word reading; performance on a spelling test is influenced more by a student's knowledge of letter–sound correspondence, knowledge of letter combinations, word-specific spellings, and reading and writing experience, and less about having been directly taught how to spell the specific words in the test. Reading comprehension depends on a host of factors related to word reading, language comprehension, and background knowledge, not only on specific comprehension strategies taught in a curriculum or on the types of texts provided. Therefore, performing well on a test of reading comprehension can be influenced by previous instruction in the subject matter of the passages, but it is certainly not dependent on high test-instruction overlap.

Mathematics, vocabulary, and some aspects of early literacy are areas in which overlap between what was taught and what is tested is more relevant. Unlike word reading, there is much less "self-teaching" possible in mathematics—correctly completing specific mathematics operations often depends on having been taught that operation. In vocabulary, unless context is provided that allows a test-taker to infer word meaning or meaning can be inferred by understanding the morphemes within the word (i.e., affixes, roots), demonstrating knowledge of vocabulary terms depends on having been exposed to them in language, reading, or having been directly taught them. In early literacy, associating a letter or letter combination with a sound is the basis of the alphabetic code and must be taught. Therefore, the relevance of test-instruction overlap varies across different academic skills.

Judicious Consideration of Standardized Norm-Referenced Tests

Previous editions of this text conveyed little optimism regarding the utility of standardized norm-referenced tests of achievement in individual student evaluations. But much has changed since the late 1980s in terms of our understanding of academic skills development and the quality and availability of academic achievement tests. Readers familiar with that history will note that the perspective taken in this edition is more open to the possibility that *some* standardized tests and subtests *of relevant academic skills* can contribute meaningfully to an academic assessment and intervention plan. The preceding and subsequent discussions are not meant to fully absolve standardized tests of achievement, but no test (standardized or not) should be above critique. Rather, any test should be scrutinized regarding its relevance and usefulness for understanding a student's academic difficulty and what to do about it.

As noted earlier, individual standardized tests are helpful in indicating the relative standing of an individual within a peer group. Provided the test assesses skills relevant to a student's grade level or stage of development, and the academic domain of interest, understanding the extent to which a student's current level of performance differs from that of their peers can help establish the need for supplemental intervention support. Across academic domains, current performance is one of the strongest predictors of subsequent achievement unless instruction intercedes. Moreover, one of the defining features of learning disabilities is performance significantly below that of peers (Fletcher et al., 2019). Understanding the extent of a student's skill deficit informs both the need for and the intensity of an intervention. When I (Clemens) work with a student with reading difficulties, one of the first things I want to see is their performance relative to grade-level peers on the Test of Word Reading Efficiency (Second Edition), given how important accuracy and efficiency in reading isolated words are for reading proficiency. To be clear, many curriculum-based measures have normative information available; therefore, this need is not satisfied exclusively by standardized tests. However, in some cases, standardized tests provide high-quality information on specific skills that is not available with other forms of assessment.

Although normative comparisons and relative standing information can be helpful for determining the need for supplemental or intensified intervention supports, standardized achievement tests may have limited use in other types of assessment decisions. An important consideration in assessing academic skills is determining how much progress students have made across time. This determination requires that periodic assessments be conducted. Not only would the costs of frequently administering standardized norm-referenced tests be prohibitive, but these measures were never designed to be repeated at frequent intervals. Doing so would create practice effects that would compromise the integrity of the test.

In addition to the problem of bias from the frequent repetition of the tests, the content and stimuli assessed on standardized tests may result in a very poor sensitivity to small changes in student behavior. Typically, these tests contain items that sample a large array of skills. As instruction or intervention occurs, gains evident on a daily or weekly basis may not appear on the standardized norm-referenced test since these skills may not be reflected on test items. Vocabulary knowledge provides a good example because learning vocabulary terms depends on being exposed to them in language, text, or instruction. Unless an intervention targeted skills in inferring word meaning from surrounding text *and* the test was set up to allow students to use context (e.g., vocabulary terms situated within a sentence), a test of vocabulary knowledge will not be sensitive to effects of an intervention if none of the vocabulary terms on the test were targeted in instruction. Indeed, studies of interventions designed to improve students' vocabulary knowledge have consistently demonstrated meaningful effect sizes on researcher-developed vocabulary tests aligned with the words targeted in the intervention, but minimal effects on standardized tests of vocabulary (Marulis & Neuman, 2010). The effects observed on the researcher-developed tests indicate that interventions were effective in improving vocabulary knowledge, but a standardized vocabulary test is not a fair index of intervention effects if students were never exposed to the tested words in the first place.

In summary, standardized tests of specific and relevant academic skills can provide information for understanding the extent and severity of a student's academic difficulties at a given time. This can help establish the need for supplementary intervention support and how intensive those supports should be. Additionally, many contemporary standardized tests have often undergone an extensive development process using advanced methodologies such as item response theory, have a broad normative base, and therefore demonstrate strong psychometric properties. On the other hand, standardized norm-referenced tests are not sensitive to small changes in student functioning, are not designed for frequent administration, and may not test content covered in instruction. These advantages and limitations must be considered in academic evaluations. The judicious use of standardized norm-referenced tests in academic assessment is discussed in more detail in Chapter 4.

Criterion-Referenced Tests

Another method for assessing individual academic skills is to examine a student's mastery of specific skills. "Mastery" in this context refers to skills learned to the point of very high accuracy and automaticity (i.e., fluency). This procedure requires a comparison of student performance against an absolute standard that reflects proficiency of a skill (i.e., a "criterion" of performance), rather than the normative comparison made to peers of the same age or grade. Many of the statewide, high-stakes tests administered by schools use criterion-referenced scoring procedures, which identify students as scoring in categories such as "below basic," "basic," "proficient," or "advanced" based on their scores relative to predetermined criteria.

Scores on criterion-referenced measures are interpreted by determining whether a student's score meets a criterion that reflects mastery or proficiency of that skill. By looking across the different skills assessed, one can determine the particular components of the content area (e.g., reading, mathematics, writing, learning-related behaviors) that represent strengths or weaknesses in a student's academic profile.

Examples of individual criterion-referenced instruments are a series of inventories developed by Brigance. Each of these measures is designed for a different age group, with the Brigance Inventory of Early Development— (Brigance, 2013a) containing subtests geared for children from birth through age 7, the Comprehensive Inventory of Basic Skills—III (CIBS-III; Brigance, 2013b) providing inventories for skills development between PreK and grade 9, and the Transition Skills Inventory (TSI; Brigance, 2013c) that contains inventories for informing plans for students in middle and high school to transition to postsecondary school life. Each measure includes skills in academic areas such as readiness, speech, listening, reading, spelling, writing, mathematics, and study skills. The inventories cover a wide range of subskills, and each inventory is linked to specific behavioral objectives.

Another example of an individual criterion-referenced test is the Phonological Awareness Literacy Screening (PALS; Invernizzi et al., 2015) developed at the University of Virginia. The PreK, kindergarten, and grades 1–3 versions were designed to measure young children's early literacy development, such as phonological awareness, alphabetic knowledge, letter sounds, spelling, word concepts, word reading in isolation, and passage reading. As with most criterion-referenced measures, the PALS is given to identify those students who have not developed these skills to levels that are predictive of future success in learning to read. The PALS measures were originally designed as broad-based screening measures, although they can also be used for diagnostic and progress monitoring decisions (Invernizzi et al., 2015).

There are situations in which criterion-referenced tests have utility. For example, criterion-referenced measures may be useful for screening decisions. If the interest is in identifying children who may be at risk for academic failure, the use of a criterion-referenced measure should provide a direct comparison of the skills a student possesses against the range of skills expected of same-age or same-grade peers, skills important for a particular period in development, or skills that are key for an upcoming transition.

The type of decision in which criterion-referenced tests may contribute most is the identification of target areas for the development of educational interventions. Given that these measures contain assessments of subskills within a domain, they may be useful in identifying specific strengths and weaknesses in a student's academic profile. In this way, students who have substantially fewer or weaker skills can be targeted for more in-depth evaluation and possible intervention. The measures do not, however, offer direct assistance in identifying intervention strategies. Instead, by indicating a student's strengths, they may aid in the development of interventions capitalizing on these subskills to remediate weaker areas of academic functioning. It is up to the evaluator and educators to (1) determine which skills identified by the test as "weaknesses" are the most important targets for additional assessment and intervention, and (2) align relevant, evidence-based interventions to address the most important areas of weakness.

Criterion-referenced tests can also be helpful for program evaluation. These types of decisions involve the examination of the progress of a large number of students. As such, any problem of limited curriculum–test overlap or sensitivity to short-term growth of students would be unlikely to affect the outcome. For example, one could use the measure to determine the percentage of students in each grade meeting the preset criteria for different subskills, and the extent to which that percentage changes across a school year. This information may be of use in evaluating the effectiveness of instruction or a curriculum. When statewide assessment data are reported, they are often used in this way to identify districts or schools that are meeting or exceeding expected standards.

Criterion-referenced tests are less helpful in other situations, such as determining eligibility for special education. If criterion-referenced measures are involved in such decisions, it is critical that skills expected to be present in students without disabilities be identified. Because these measures do not typically have a normative base, it becomes difficult to make statements about a student's relative standing to peers. For example, to use a criterion-referenced test in kindergarten screening, it is necessary to know the type and level of subskills children should possess as they enter kindergarten. If this information were known, the obtained score of a specific student could be compared to the expected score, and a decision regarding their probability of success could be derived. Of course, the empirical verification of this score would be necessary since identifying subskills needed for kindergarten entrance would most likely be obtained through a teacher interview. Although criterion-referenced tests could be involved in classification decisions, they should not be employed in isolation in this way.

A challenge to interpreting some criterion-referenced tests is that it is not clear how the criterion representing mastery was derived. Although one method for establishing this criterion may be a normative comparison (e.g., criterion = number of items passed by 80% of same-age/same-grade peers), most criterion-referenced tests establish the acceptable criterion score based on logical, rather than empirical, analysis.

Another area in which individual criterion-referenced tests require careful consideration is in monitoring student progress. It would seem that since these measures only make intrasubject comparisons, they would be valuable for monitoring student progress across time. Indeed, some measures like the PALS can be used for monitoring students' progress toward important milestones and benchmarks in early literacy. However, some criterion-referenced tests are not set up for repeated administration. Most criterion-referenced measures are developed by examining standards or published curricula and pulling a subset of items together to assess a subskill. As such, student gains in a specific curriculum may or may not be related directly to performance on the criterion-referenced test. Our earlier discussion regarding test-curriculum overlap is also relevant here.

Another consideration related to monitoring student progress is the limited range of subskills included in a criterion-referenced test. Typically, most criterion-referenced measures contain a limited sample of subskills and a limited number of items assessing any particular subskill. This may require frequently changing the progress monitoring measure as skills are mastered. Nevertheless, as we discuss further in Chapter 7, there are times when monitoring a student's progress toward mastery of a specific skill (which may be defined as meeting a particular accuracy or fluency criterion) that is directly targeted by instruction can be a valuable component of a progress monitoring plan and may benefit the instructional decision-making processes (VanDerHeyden & Burns, 2018).

In summary, like standardized norm-referenced tests, criterion-referenced tests have utility in some situations. Criterion-referenced tests have strong relationships to intrasubject comparison methods and ties to behavioral assessment strategies (Cancelli & Kratochwill, 1981; Elliott & Fuchs, 1997). Furthermore, because the measures offer assessments of subskills within broader areas, they may provide useful mechanisms for the identification of remediation targets in the development of intervention strategies. Criterion-referenced tests may also be useful in a screening process and can help monitor progress in some situations.

Conversely, criterion-referenced tests have limited utility for educational classification, the source of their criteria may not be known, and the development of intervention strategies may not be addressed adequately with these measures alone. In individual assessment situations, like any type of measure, criterion-referenced tests should be used judiciously and viewed as contributors to an overall assessment, are not necessary in all situations, and should not be used in isolation.

Both norm- and criterion-referenced measures provide an indirect evaluation of skills by assessing students on a sample of items taken from expected grade-level performance. This may be helpful when one wishes to gauge a student's overall achievement or evaluate the extent to which a student can generalize skills learned in instruction to content not directly taught. In other cases, however, the items selected on a norm- or criterion-referenced test may not have strong relationships to what students were expected to learn. Additionally, because the measures provide samples of behavior, they may not be sensitive to small gains in student performance across time. As such, they cannot directly tell when an intervention method is successful.

The use of norm- or criterion-referenced tests on their own do not reveal the influence of the instructional environment on student academic performance. Both norm- and criterion-referenced instruments may tell us certain things about a student's individual skills, but they do not assess variables that affect academic performance, such as instructional methods for presenting material, feedback mechanisms, classroom structure, competing contingencies, and so forth (Lentz & Shapiro, 1986).

Clearly, what is needed in the evaluation of academic skills is a method that more directly assesses student performance within the academic curriculum. This methodology should directly assess both the student's skills and the academic environment. It should be able to incorporate a variety of assessment methods and address a wide range of educational decisions. This type of methodology is the basis for this text.

DIRECT ASSESSMENT OF ACADEMIC SKILLS: CURRICULUM-BASED ASSESSMENT

Many assessment models have been derived for the direct evaluation of academic skills (e.g., Blankenship, 1985; Deno, 1985; Gickling & Havertape, 1981; Howell & Nolet, 1999; Salvia & Hughes, 1990; Shapiro, 2004; Shapiro & Lentz, 1986). These models are driven by an underlying assumption that *one should test what one teaches*. As such, the content or skills evaluated by the assessments employed for each model are based on the instructional curriculum and skills critical for academic success. In contrast to the potential problem of limited correspondence between the curriculum and the skills measured by the test in other forms of academic assessment, evaluation methods based on the curriculum offer direct evaluations of student performance on skills that students are expected to acquire. Thus, *direct assessment* refers to methods that directly assess skills relevant to the student's academic achievement. The close connection of these methods to the curriculum of instruction is why these models are generally referred to as methods of *curriculum-based assessment* (CBA). Various direct assessment and CBA models have been conceptualized over the years, and each model provided a somewhat different emphasis to the evaluation process. In addition, various terms have been used by these investigators to describe their respective models.

Gickling's CBA Model

Gickling and colleagues (Gickling & Havertape, 1981; Gickling & Rosenfield, 1995; Gickling & Thompson, 1985; Gravois & Gickling, 2002; Rosenfield & Kuralt, 1990) described a subskill mastery model of CBA that seeks to determine the instructional needs of struggling students based on their performance within the curriculum and content used for instruction. The primary goal is to identify and eliminate a mismatch between the skill level of a target student and the demands of the curriculum in which the student is taught (Gickling & Thompson, 1985). Gickling's CBA model concentrates on the selection of instructional objectives and content based on assessment. This model tries to control the level of instructional delivery carefully so that student success is maximized. To accomplish this task, a student's academic skills are evaluated to identify items that are "known" (i.e., immediate and correct response), "hesitant" (i.e., response was correct, but the student struggled or was unsure), and "unknown" (i.e., incorrect response). The proportion of known to hesitant and unknown items is then used to inform adjustments to the content used for instruction so it is at an "instructional" level suitable for use during teaching, as compared to "independent" (very high accuracy or number of "knowns") or "frustrational" levels (very low accuracy or high number of "unknowns"; Betts, 1946; Gickling & Thompson, 1985).

Shapiro (1992) illustrated how this CBA model can be applied to a student with problems in reading fluency. Burns and colleagues (Burns, 2001, 2007; Burns et al., 2000; MacQuarrie et al., 2002; Szadokierski & Burns, 2008) have published studies offering empirical support for various components of Gickling's model of CBA.

Howell's Curriculum-Based Evaluation Model

Howell and colleagues (Hosp et al., 2014; Howell et al., 2002; Howell & Nolet, 1999) established a subskill-mastery model and problem-solving approach called *curriculum-based evaluation* (CBE) that is wider in scope and application than the Gickling model. Howell's CBE model is a process of inquiry that uses a series of if–then precepts explicitly aimed at understanding a student's problem and what to do about it. It uses strategies such as task analysis, skill probes, direct observation, and other evaluation tools to gather information. Extensive suggestions for intervention programs based on CBE are offered for various subskills such as reading comprehension, decoding, mathematics, written communication, and social skills. In addition, details are provided for changing intervention strategies.

CBE is a problem-solving process that involves five steps: (1) fact-finding/problem identification, (2) hypothesizing assumed causes of the problem, (3) confirming or revising hypotheses through problem validation, (4) summative decision-making and intervention identification, and (5) formative decision-making during intervention (Howell et al., 2002; Kelley et al., 2008). The initial step of the process involves identifying the extent of a difference between the student's current and expected performance, thereby defining the problem the student is experiencing. For example, in mathematics, one would conduct what are called "survey-level assessments." These assessments examine student performance on various mathematics skill domains that are considered related to the area of math in which the student is having difficulty, and evaluating the size of the discrepancy between the student's skills and what is expected at that point in time. In addition to the administration of specific mathematics problems, data would be obtained through interviews with teachers and students, observations of the student completing the problems, and an error analysis of the student's performance.

The data from the initial step of the process leads the examiner to develop hypotheses related to the root causes of the student's problems in math. This step involves pinpointing skill or knowledge gaps and specific areas of weakness that are causing the student's difficulties in the broader academic domain. For example, a CBE might lead the examiner to hypothesize that, based on assessments collected in Step 1, a student's reading difficulties are caused in large part by insufficient knowledge of letter–sound correspondence, which is critical for the acquisition of decoding and fluent word recognition skills. This "best guess" serves as the guide in the next part of the evaluation.

In the next step of the process, validation of the problem is conducted by developing specific-level assessments that focus on the hypothesized skill problems that were identified. If insufficient letter–sound knowledge is hypothesized as the root cause of the student's reading difficulties, the assessment would specifically determine the number of letters and letter combinations the student can correctly associate with their common sounds. Very low accuracy with the task would be viewed as support for the hypothesis. If the hypothesis is not validated, for example, the student demonstrates 95% accuracy in letter–sound correspondence, it suggests that the hypothesized cause was not the root of the problem. The process would cycle back to Step 2 to revise the hypothesis and gather additional assessment data.

Once the hypothesis is validated, the data collected are summarized and a plan for remediation is developed. Validation of the root cause of the problem facilitates the identification of relevant evidence-based interventions for targeting the need. To provide baseline data, a statement indicating the student's present level of achievement and functional performance, and a list of goals and objectives for instruction, are developed. Once the instructional plan is developed and implemented, formative assessment (i.e., progress monitoring) data are collected to determine whether the intervention is having its desired effect. Lack of progress signals the need to back up to Step 4 to reconsider the assessment data and devise a new intervention that will better meet the student's needs.

A key to CBE is a recognition that the remediation of academic skills problems requires a systematic problem-solving process. CBE involves a thorough examination of the subskills required to reach mastery and pinpoints interventions that are designed specifically to remediate those skills. The if–then nature of the process provides an empirical, inquiry-based framework that utilizes assessments relevant to and aligned with the referral concern. This process can result in an intervention aimed specifically at the root of the student's academic difficulty.

Deno's Curriculum-Based Measurement Framework

Among models of CBA, the approach with the most substantial research base is that developed by Deno and colleagues at the University of Minnesota (e.g., Deno, 1985). Derived from earlier work on "data-based program modification" (Deno & Mirkin, 1977), Deno's model, called *curriculum-based measurement* (CBM), was primarily designed as a progress monitoring framework to guide instructional decisions (i.e., signaling decisions to continue effective programs or make adjustments to ineffective ones), rather than as a system designed to develop intervention strategies. However, as will be discussed later in this text, information useful for informing a student's skill difficulties and what kinds of adjustments may be warranted may still be gleaned from the ongoing collection of CBM data.

The model involves repeated and frequent administration of a standard set of skills probes that are known to be important indices of student proficiency within the skill domain in which the child is being taught (i.e., reading, mathematics, writing). Research has shown that the model is effective in monitoring student progress even when the measures used for purposes of monitoring are not derived directly from the material used for instruction (L. S. Fuchs & Deno, 1994). The skill measured by the probes (e.g., oral reading fluency) may not necessarily be the skill being directly taught or practiced, but is viewed as an academic "vital sign" that reflects improvement and overall skill acquisition in the academic domain. Although CBM was originally designed for monitoring students' progress within interventions, subsequent work by Deno and colleagues supported the use of CBM tools and decision-making frameworks for this and other related purposes, such as screening decisions, eligibility decisions, progress monitoring, and program evaluation (e.g., Deno, L. S. Fuchs, et al., 2001; Deno, Marston, & Mirkin, 1982; Deno, Marston, & Tindal, 1985–1986; Deno, Mirkin, & Chiang, 1982; Filderman et al., 2018; Foegen et al., 2007; L. S. Fuchs, Deno, et al., 1984; L. S. Fuchs & D. Fuchs, 1986a; L. S. Fuchs, D. Fuchs, et al., 1994; Jung et al., 2018; Shinn et al., 1993; Stecker & L. S. Fuchs, 2000; Stecker et al., 2005; Wayman et al., 2007).

General Outcomes versus Specific Subskill Mastery Models

L. S. Fuchs and Deno (1991) differentiated between *general outcomes measurement* and *specific subskill-mastery measurement* models used for monitoring progress. General outcomes measures are central to Deno's CBM framework and focus on skills reflective of overall achievement (e.g., the measurement of oral reading rate as an indicator of overall reading achievement). As such, they provide measurement of students' growth toward year-end "general" outcomes in the overall academic domain. General outcomes measures are presented in a standardized format. Material for assessment is controlled for difficulty by grade levels and may or may not come directly from the curriculum of instruction (L. S. Fuchs & Deno, 1994). Typically, measures are presented as brief, timed samples of performance, using rate as the primary metric. Although outcomes derived from this model may suggest when and if instructional modifications are needed, and the information gained by monitoring a student with a general outcomes approach can inform the need for additional assessment and identifying the kinds of instructional decisions that should be made, the model was designed to signal the need to either continue an effective program or adjust an ineffective one, and not to suggest what those instructional modifications should be.

In contrast, a specific subskill-mastery measurement approach (L. S. Fuchs & Deno, 1991) uses measures that include items contained in the instructional program, such as vocabulary terms included in a curriculum. Measures are often administered to evaluate a student's progress toward meeting a specific mastery criterion, for example, 100% accuracy completing a set of mathematics computation problems targeted in instruction. Specific subskill-mastery measures are generally not considered part of the CBM framework (L. S. Fuchs & Deno, 1991), but still can provide highly useful data for making instructional decisions (VanDerHeyden et al., 2018). The primary objective of the subskill-mastery model is to determine whether students are meeting the short-term instructional objectives of the curriculum, in other words, whether the student is acquiring the skill or content that is being targeted in instruction. Specific subskill-mastery measures may be teacher-made or embedded within the curriculum of instruction, and the metric used to determine student performance can include accuracy, rate, or analysis of error patterns. This type of approach usually requires a shift in measurement with the teaching of each new objective. However, the measures are often highly sensitive to student progress with the skills being targeted, and can help identify areas in which instruction should be adjusted.

Shapiro's Model of CBA

Although the previously described models each differ in various ways, they all focus on the direct assessment of students' academic performance to evaluate academic problems. Certainly, the importance of assessing individual academic skills cannot be denied. However, it seems equally important to examine the instructional environment in which the student is being taught. Lentz and Shapiro (Shapiro, 1987a, 1990, 1996a, 1996b, 2004;

Shapiro & Lentz, 1985, 1986) provided a model for academic assessment that incorporated the evaluation of the academic environment and student performance. Calling the model "behavioral assessment of academic skills," they (Shapiro, 1989, 1996a; Shapiro & Lentz, 1985, 1986) drew on the principles of behavioral assessment employed for assessing social–emotional problems (Mash & Terdal, 1997; Ollendick & Hersen, 1984; Shapiro & Kratochwill, 2000) but applied them to the evaluation of academic problems. Teacher interviews, student interviews, systematic direct observation, and examinations of student-produced academic products played a significant part in the evaluation process. Specific variables examined for the assessment process were selected from the research on effective teaching (e.g., Denham & Lieberman, 1980) and applied behavior analysis (e.g., Sulzer-Azaroff & Mayer, 1986). In addition, the CBM methodology developed by Deno and colleagues was used to evaluate individual student performance on an ongoing basis but was combined with the assessment of the instructional environment in making recommendations for intervention. Indeed, this assessment of the instructional ecology differentiates the model from other models of CBA.

In a refinement of this model, Shapiro (1990) described a four-step process for assessing academic skills that integrates several of the existing models of CBA into a systematic methodology for conducting direct academic assessment. As illustrated in Figure 1.1, the process begins with an evaluation of the instructional environment through the use of systematic observation, teacher interviewing, student interviewing, and a review

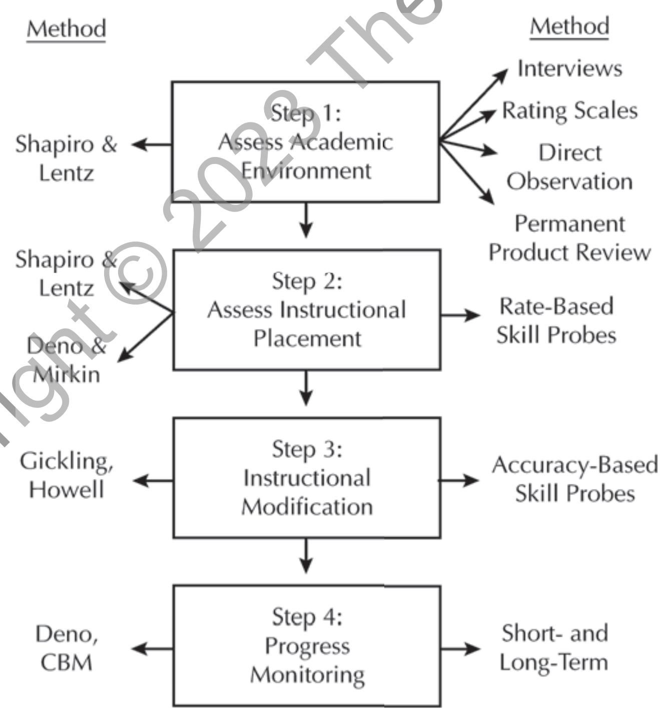

FIGURE 1.1. Integrated model of CBA. Adapted from Shapiro (1990, p. 334). Copyright © National Association of School Psychologists, Inc. Reprinted by permission of Taylor & Francis Ltd, *http://www.tandfonline.com* on behalf of National Association of School Psychologists.

of student-produced academic products. The assessment continues by determining the student's current instructional level in curriculum materials and directly assessing relevant academic skills. Next, instructional modifications, designed to maximize student success, are implemented with ongoing assessment of the acquisition of instructional objectives (subskill and short-term goals). The final step of the model involves monitoring student progress toward general outcomes (year-end) curriculum goals. Shapiro's model remains essentially unchanged in this edition of the text. As described in more detail in Chapter 3, readers will find the core aims, stages of assessment, and methodologies the same. Within the model, however, readers will discover some expansions to the assessment and intervention methods conducted within the steps based on research and theory that have emerged in the 30 years since the model was first established.

One of the important considerations in adopting a methodology for conducting academic assessment is the degree to which the proposed change is acceptable to consumers who will use the method. Substantial attention in the literature has been given to the importance of the acceptability of intervention strategies recommended for school- and home-based behavior management (e.g., Clark & Elliott, 1988; Miltenberger, 1990; Reimers, Wacker, Cooper, et al., 1992; Reimers, Wacker, Derby, et al., 1995; Witt & Elliott, 1985). Eckert and Shapiro (Eckert, Hintze, et al., 1999; Eckert & Shapiro, 1999; Eckert, Shapiro, et al., 1995; Shapiro & Eckert, 1994) extended the concept of treatment acceptability to assessment acceptability. Using a measure derived from the Intervention Rating Profile (Witt & Martens, 1983), the studies demonstrated that both teachers and school psychologists found CBA, compared to the use of standardized norm-referenced tests, to be relatively more acceptable in conducting academic skills assessments. In a nationally derived sample, Shapiro and Eckert (1993) also showed that 46% of school psychologists surveyed indicated that they had used some form of CBA in their work. In a replication and extension of the original survey 10 years later, 53% of those psychologists surveyed indicated that they had used CBA in conducting academic assessments (Shapiro, Angello, et al., 2004). However, surveys in both 1990 and 2000 also revealed that school psychologists still had limited knowledge of the actual methods used in conducting a CBA. Although there was a statistically significant ($p < .05$) increase in the percentage reporting use of CBA in 2000 compared to 1990, a large proportion of psychologists still reported not using CBA. At the same time, over 90% of those surveyed who had graduated in the past 5 years had been trained in the use of CBA through their graduate programs.

Chafouleas et al. (2003) found a similarly stronger acceptability of CBA methods among 188 members of the National Association of School Psychologists for developing intervention strategies compared to norm-referenced tests or brief experimental analysis (a methodology discussed further in Chapter 4). O'Bryon and Rogers (2010) determined that 53% of school psychologists surveyed reported using CBA as part of their assessment practices with English learners. Filter et al. (2013) found that practicing school psychologists reported CBA was an activity and assessment model they wished they could use more often, but were unable due to other responsibilities such as intelligence testing and report writing. In their comprehensive survey of 1,317 school psychologists, Benson et al. (2019) observed that CBM measures represented 3 of the top 10 forms of assessment that respondents reported using on a monthly basis.

Overall, these studies suggest that although CBA and related methods are highly acceptable to practitioners (teachers and school psychologists), school psychologists usually prefer to use CBA, and it is increasingly being taught as part of the curriculum in training school psychologists. As of the most recent survey by Benson et al. (2019), it appears that elements of CBA are commonly used (mainly CBM tools). However, systematic models of CBA that take a comprehensive approach to assessment have yet to assume a prominent role in the assessment methods of practicing psychologists.

Summary of Assessment Approaches

Strategies for the assessment of academic skills range from the more indirect norm- and criterion-referenced methods through direct assessment, which is based on evaluating a student's performance on skills critical for achievement and consistent with the curriculum of instruction. The type of decision to be made must be tied to the assessment strategy employed. Although some norm-referenced standardized assessment methods may be helpful in understanding the severity of a student's academic difficulties and may contribute to making eligibility decisions, these strategies are lacking in their ability to assist evaluators in recommending appropriate intervention strategies or in their sensitivity to improvements in academic skills over short periods of time. Many alternative assessment strategies, designed to provide closer links between the assessment data and intervention methods, are available. Although these strategies may also be limited to certain types of decision making, their shared characteristic of using the curriculum as the basis for assessment allows these methods to be employed more effectively for several different types of assessment and evaluation decisions. Shapiro's model of CBA is the basis of this text, and it is described in detail across the remaining chapters.

INTERVENTION METHODS FOR ACADEMIC SKILLS

Intervention strategies developed for academic problems can be conceptualized on a continuum from indirect to direct procedures. Intervention techniques that attempt to improve academic performance by targeting underlying cognitive, neuropsychological, motor, or sensory processes can be characterized as *indirect* interventions because they are considered to affect academic performance by way of improvement in an underlying process. In particular, programs that target working memory, executive functions, perceptual–motor skills, or interventions based on *aptitude by treatment interactions* (ATIs), would be considered indirect methods of intervention.

In contrast, *direct* interventions attempt to improve the area of academic difficulty by directly teaching the academic skills missing from the student's repertoire. These skills are targeted because assessment data indicate they are reasons for the student's difficulties in the larger academic domain. These types of interventions usually are based on an examination of variables that are centrally involved in academic performance.

Indirect Interventions for Academic Skills

Working Memory Training

Computerized working memory training programs are commercially available, and their use in schools has increased significantly. Programs typically consist of games and activities designed to improve working memory and attention. Cogmed (Neural Assembly, 2021) is an example of a computer-based program actively marketed to schools and clinicians as an intervention for students with learning and attention difficulties. Developers of working memory training programs like Cogmed make two assumptions: (1) They assume that working memory and attention can be improved by playing games that involve working memory (i.e., "near-transfer" effects), and (2) they assume that improvements in working memory and attention will lead to improvements in attention and on-task behavior in the classroom, cognitive abilities, and academic skills such as reading and mathematics (i.e., "far-transfer" effects). The products are promoted through extensive marketing materials, which often include quotes from unnamed educators regarding the dramatic improvements they observed in their students' attention and academic achievement following use of the program.

Unfortunately, the research evidence on computerized working memory programs is not nearly as encouraging as the marketing materials suggest. High-quality randomized controlled trials and several meta-analyses (i.e., quantitative summaries of multiple research studies) have reached similar conclusions: Working memory training tends to result in short-term near-transfer improvement on working memory tasks aligned with the activities in the training programs, but weak to negligible far-transfer improvements in more general areas of functioning including verbal or nonverbal ability, reading skills, mathematics performance, attention, or behavior regulation (Anderson et al., 2018; Chacko et al., 2014; Gray et al., 2012; Hitchcock & Westwell, 2017; Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016; Redick et al., 2013; Sala & Gobet, 2017; Shipstead, Hicks, et al., 2012; Shipstead, Redick, et al., 2012; Soveri et al., 2017). To illustrate, a comprehensive meta-analysis of 145 experimental comparisons by Melby-Lervåg et al. (2016) found that any improvements in working memory following training were short-lived, and when they did occur, transfer to improvements in generalized attention, cognitive functioning, or academic skills was not observed. Thus, experimental evidence has not supported a core assumption on which working memory training is based.

An additional problem with assessment and intervention for working memory, which extends to other cognitive processes, is that working memory is embedded and highly integrated within human language, cognition, and behaviors. In fact, you are using working memory right now as you read, and you will use it later when you are talking with someone, writing an email, or listening to the news. It is indispensable for complex, intellectual tasks and activities. This deep integration makes working memory difficult to isolate and measure as a unique, discrete construct. As a result, working memory measures tend to be highly specific, rudimentary indices that reveal very little about a student's academic skill difficulties, and certainly do not reveal more about a student's problems than a test of the actual academic skill of interest. Moreover, a working memory test administered to a student with academic difficulties will probably indicate low performance, but this knowledge is of little practical use because improving working memory has not yet been shown to reliably improve academic performance. Instead, assessment is better focused on the skills relevant to reading, mathematics, and writing development and difficulties and will translate to more relevant, practical, and effective intervention recommendations.

Interventions Targeting "Executive Functioning"

Although its precise definitions vary, executive functioning (EF) refers to a class of processes that allow individuals to accomplish goal-directed activities (Cirino et al., 2019; Miciak et al., 2019). EF is generally agreed to include inhibitory control, cognitive flexibility, and working memory (Diamond, 2013).

EF has long been hypothesized to play a causal role in academic achievement (Bierman et al., 2008), and numerous correlational studies have observed that EF, even from very young ages, is correlated with academic skills measured at the same time or at some point in the future (e.g., Blankenship et al., 2019; Blair & Razza, 2007; Cirino et al., 2019; Cutting et al., 2009; Jacob & Parkinson, 2015; Morgan et al., 2019). However, as with other processes that indirectly affect academic skills, a correlation between EF and achievement does not necessarily mean that targeting EF will lead to academic improvement. EF and its related subskills have not been found to predict student responsiveness to academic intervention (Miciak et al., 2019), and comprehensive meta-analyses by Jacob and Parkinson (2015), Rapport et al. (2013), and Kassai et al. (2019) found no convincing evidence that interventions targeting EF led to improved academic achievement or attention.

A problem with EF training, in most studies, is that it targets basic processes through games, activities, and training routines that are highly distal and abstracted from tasks and situations that students navigate in the classroom. Interventions have targeted self-regulation, inhibitory control, or attention in computer games or in clinical training sessions outside of the classroom, and thus have observed limited effects on generalized classroom behavior or academic skills (Jacob & Parkinson, 2015; Rapport et al., 2013). These effects are consistent with that of most other indirect interventions—the further one separates a task or process from how the skill should be demonstrated in real-life academic situations, the less likely that any improvement in the abstracted process will improve academic skills.

There is one aspect of EF that shares a good deal of overlap with behaviors that *have* been demonstrated to be causally related to students' overall functioning and achievement in school. Most definitions of EF include *inhibitory control*, which represents one's ability to suppress a dominant response when needed. For example, a student suppressing their urge to blurt out an answer to a teacher's question, and raising their hand instead, is demonstrating inhibitory control. Inhibitory control is part of *self-regulation*, also considered a component of EF. Self-regulation, and behavioral self-regulation more specifically, refers to a set of acquired skills involved in intentionally activating, directing, sustaining, and controlling one's attention, emotions, and behaviors (Morrison et al., 2010). Self-regulation skills are critical for accomplishing complex academic tasks, which often require sustained attention, remembering rules and expectations, determining appropriate solutions, reflection and monitoring understanding, task persistence, and detecting errors. As such, strategies to teach and support behavioral self-regulation required of complex tasks have been successfully integrated into interventions in reading (Joseph & Eveleigh, 2011), written expression (Harris & Graham, 2016), and mathematics (L. S. Fuchs et al., 2003). Important aspects of self-regulation have been studied that pertain directly to learning-related skills and students' engagement within academic tasks. Behavior self-regulation can be viewed as a point at which behavior and learning intersect. As will be discussed in Chapters 5 and 6, interventions aimed at improving students' task engagement have been shown to improve academic skills—but the crucial aspect is, unlike many applications of general EF training, the student's self-regulation behavior is targeted within the specific academic situations in which it is most relevant and needed.

In summary, interventions that target executive functions with tasks that are isolated and discrete from the situations in which they are needed are unlikely to improve academic skills. However, directly targeting EF-related behaviors necessary for success in academic tasks (i.e., behavior self-regulation), and doing so in the actual situation where the behaviors are relevant, is more likely to be associated with improved academic performance.

Perceptual–Motor Training and Interventions Targeting Visual and Auditory Modalities

A number of intervention approaches have targeted gross and fine motor skills, and foundational perceptual abilities such as visual and auditory processing, in efforts to improve academic achievement and related skills (e.g., task engagement, cognitive functioning). The rationale for targeting motor skills is based on hypotheses that complex, higher-order cognitive skills, such as reading or mathematics proficiency, have their roots in brain areas and systems such as the cerebellum and the vestibular system that are responsible for coordinating movement, balance, and processing incoming sensory information. Deficits or dysfunction in these systems is thought to cause learning difficulties and disabilities; it is also thought that targeting underlying perceptual and motor skills should improve academic skills (e.g., Kibby et al., 2008; Levinson, 1988; Nicolson et al., 2001; Westendorp et al., 2011). Hypotheses such as these were most notably reflected in programs that emerged in the 1960s, such as the Doman–Delacato patterning technique (Delacato, 1963), and intervention strategies that involved gross and fine motor coordination, such as crawling, walking, hand–eye coordination, and balancing. Like many other interventions that are far removed from the involvement of academic skills, evidence for the efficacy of these approaches was not observed. In 1982, the American Academy of Pediatrics condemned the Doman–Delacato technique as ineffective and potentially harmful, and the seminal meta-analysis by Kavale and Mattson (1983) observed no benefits of perceptual–motor training on students' functioning.

Nevertheless, programs continue to emerge that market kinesthetic activity and perceptual–motor training as ways to "repattern" the brain to improve learning. Brain Gym® is one such example. But, like before, reviews of research show no evidence of benefit on learning or academic achievement (Hyatt, 2007; Hyatt et al., 2009; Spaulding et al., 2010). Furthermore, well-controlled studies and meta-analyses have failed to find evidence that cerebellar functions play a causal role in learning difficulties, nor is there evidence that students with learning difficulties demonstrate deficits, structural abnormalities, or dysfunction in this area (Barth et al., 2010; Irannejad & Savage, 2012; Paulesu et al., 2014; Richlan et al., 2009).

Other intervention approaches have focused on visual or auditory processing. These methods assume that because vision and language are so heavily involved in performing academic tasks such as reading, writing, and mathematics (1) learning difficulties must be caused by visual or auditory processing deficits, and (2) training these specific areas will result in academic skills improvement. The magnocellular theory (Stein, 2001) is one such example; it postulates that reading disabilities are caused by abnormal functioning of large neurons in the visual pathway, which causes text to appear crowded and jittery. There is evidence that individuals with reading disabilities perform more poorly on visual tasks compared to individuals without reading disabilities (Tafti et al., 2014); however, research indicates that poorer performance on visual tasks is more likely a *consequence* of impoverished reading experience, not a cause of it (Creavin et al., 2015; Olulade et al., 2013).

Other approaches in reading have targeted the visual modality through the use of colored eyeglass lenses or colored overlays placed over text (Irlen & Lass, 1989), vision exercises (e.g., Badami et al., 2016), or special fonts designed for individuals with dyslexia (e.g., Dyslexie, OpenDyslexic) that are believed to improve visual processing. None of these approaches have demonstrated reliable improvements in reading. Colored overlays demonstrate no differential benefit for improving reading skills (Hyatt et al., 2009); special fonts like Dyslexie do not improve reading skills for students with reading disabilities, nor do students prefer them over typical fonts (e.g., Kuster et al., 2018; Wery & Diliberto, 2017); and vision-specific exercises or interventions do not improve reading and are not endorsed by major pediatric and ophthalmologist organizations for treating reading disabilities (American Academy of Pediatrics, 2011). Vision is obviously involved in sighted reading, but as will be discussed in Chapter 4, vision is how information is taken in, where other areas and brain systems responsible for rapidly connecting symbols with phonological information, language, reasoning, and self-regulation are primarily involved in reading achievement.

Similar arguments have been made regarding the role of auditory processing as an explanation for learning difficulties (i.e., "auditory processing disorder"), and that targeting this domain will result in academic skills improvements for students with learning difficulties (e.g., Tallal, 1980). Fast ForWord (Tallal, 2000) is an example of an auditory processing program designed to improve reading and language skills, which includes audio-visual training exercises using acoustically modified language to target memory, timing, sequencing, and discrimination of sounds in oral language. However, a meta-analysis of experimental studies revealed no benefits of the Fast ForWord intervention on students' language or reading difficulties (Strong et al., 2011).

Aptitude by Treatment Interactions

Historically, many indirect interventions for remediating academic skills have been based on the concept of ATI. The basis for ATI is the theory that individuals differ in the ways they best learn (i.e., "aptitudes"), which were commonly thought to include visual, auditory, or kinesthetic modalities. The ATI concept emerged in the 1960s and enjoyed significant popularity during the 1970s, but its influence can still be observed today. In education, one often hears reference to children's "learning styles," reflecting a belief that one student may be more of a "visual" learner, while another may be more of an "auditory" learner. Another commonly held belief is that high-ability learners benefit more from instructional environments that are less structured, whereas lower-achieving students require greater structure. These notions originated with the ATI theory.

By extension, the ATI hypothesis posited that different aptitudes require different treatment. If one properly matches the correct treatment to the correct learning style (i.e., aptitude), then gains will be observed in the child's learning and behavior. For example, an evaluation by an advocate of learning styles may suggest the student is a "visual learner" and possesses a preference for learning in the visual over auditory modality. In light of this information, and ignoring the fact that the vast majority of learners could be classified as visual learners given how much humans rely on sight for learning, the learning styles advocate would predict that the student will succeed if their instruction is based more on a visual than an auditory approach to teaching reading. In this example, the aptitude (a strength in the visual modality) is matched to a treatment procedure (change to a more visually oriented reading curriculum). Thus, a better ATI is sought in hopes of improving the student's skills.

Experimental studies that examined the validity of ATIs have not yielded encouraging results. In a meta-analysis of 39 studies of ATI-based assessment and instruction, Kavale and Forness (1987) found that (1) groups of individuals displaying a particular modality preference could not be consistently or reliably differentiated, meaning there was considerable overlap in modality preferences across groups, and therefore no evidence that one student could be reliably classified as an "auditory learner" and another classified a "visual learner"; and (2) there were no benefits associated with instruction linked to the assessed modality of learning.

Pashler et al. (2009) conducted an updated, comprehensive review of research on learning styles and ATIs. They determined that although children and adults may differ in how they process different forms of information (i.e., auditory, visual, kinesthetic), and will report a preference in the way they like information to be presented, there was no evidence that aligning instruction to an individual's learning style improved learning or achievement. Pashler et al. concluded that "at present, there is no adequate evidence base to justify incorporating learning styles assessments into general educational practice" (p. 105).

THE PERSISTENCE OF THE ATI CONCEPT: MODELS OF COGNITIVE PROCESSING STRENGTHS AND WEAKNESSES

Despite a lack of evidence that a student's learning aptitude contributes to instruction planning, significant efforts are still being made to use the ATI paradigm to explain academic failure. One type of effort to maintain the ATI concept is a group of methods that seek to identify a student's pattern of cognitive *processing strengths and weaknesses* (PSW) as a basis for identifying learning disabilities and designing intervention (Flanagan et al., 2013; Hale & Fiorello, 2004; Hale et al., 2008). PSW approaches include three types: cross-battery assessment (XBA; Flanagan et al., 2013), the concordance–discordance method (C-DM; Hale et al., 2008), and the discrepancy/consistency method (D/CM; Naglieri & Das, 1997). Although the three models differ in specifics, they share a common conceptual framework (Fiorello et al., 2014; Flanagan et al., 2010, 2013). PSW models involve administering a large set of cognitive processing tests, in addition to tests of academic achievement. PSW models assume that a student with a learning disability has strengths (i.e., average scores or better) in many cognitive domains but a weakness in a specific cognitive process that is related to the academic skill that was the reason for the referral. A statistically significant difference between the cognitive process strengths and the specific cognitive weakness is viewed as evidence for the *specific* nature of the learning disability, as opposed to a general difficulty in learning. At the same time, the evaluation must reveal a weakness in an academic skill that is significantly different from the cognitive strength, but not significantly different from the specific cognitive process weakness. Thus, the academic difficulty is *unexpected* and *specific*, given that the academic difficulty occurs in the presence of multiple cognitive strengths (hence, "specific learning disability"). Determining a student's cognitive profile is thought to facilitate the identification of interventions (Flanagan et al., 2010; Hale et al., 2008).

Research has revealed significant problems with PSW methods of evaluation. Studies have revealed that PSW methods demonstrate poor accuracy in correctly identifying students who have a learning disability (Kranzler et al., 2016a; Miciak, Fletcher, et al., 2014; Miciak, Taylor, et al., 2018; Stuebing et al., 2012). There are psychometric concerns with PSW; evidence indicates problems with the validity and reliability of profile analysis (Fletcher et al., 2019; McGill & Busse, 2017; McGill et al., 2018), and that standardized achievement tests from different batteries are not as interchangeable as PSW models assume (Miciak et al., 2015). Maki and Adams (2020) found that when presented with case studies, school psychologists demonstrated significantly less consistency in interpreting the results of the same PSW evaluations compared to results based on a student's responsiveness to intervention. Even more problematic is the lack of evidence that a PSW evaluation meaningfully informs intervention (Fletcher et al., 2019; Miciak et al., 2016); some have argued that attempts to connect cognitive profile assessment to intervention by proponents of PSW are perhaps more likely to be driven by confirmation bias or clinical illusion (Kranzler et al., 2016b; McGill & Busse, 2017). Although advocates of PSW have acknowledged that other sources of information are used in evaluation decisions and that sole reliance on a cognitive profile will result in inaccuracies (Fiorello et al., 2014), to date there is no convincing evidence that determining a student's cognitive processing profile adds any value to justify its expense in time and resources (Fletcher & Miciak, 2017). Overall, very little empirical evidence supports the utility of PSW models, which led Kranzler et al. (2016b) to conclude that PSW approaches have more characteristics of pseudoscience than evidence-based practice.

A fundamental problem with models of PSW is the assumption that matching a student to a treatment that targets the underdeveloped or poorly functioning cognitive process will lead to improvements in academic skills. Syntheses of research evidence do not support this assumption. A meta-analysis by Burns et al. (2016) of 37 studies that investigated the use of neuropsychological assessment data to design instruction observed an overall effect size of $d = 0.17$ on student achievement (i.e., a small effect). On the other hand, the meta-analysis revealed that instruction based on direct assessment of relevant skills, such as reading fluency or phonological awareness, was associated with educationally meaningful effect sizes of 0.43 and 0.48, respectively. Kearns and Fuchs (2013), after systematically reviewing 50 studies in which instruction that targeted a cognitive function, or combined a focus on cognitive function with academic components, failed to find support in improving students' academic achievement.

Make no mistake: Cognitive processing is absolutely involved in academic achievement, as it is with any complex human activity. Advances in neuroscience have occurred through advances in imaging techniques that have provided insights on the brain regions, structures, systems, and networks involved in reading skills, both in terms of patterns of activation associated with typical functioning as well as reading disabilities (Richlan, 2019, 2020), and neurological differences that precede or follow reading interventions (Barquero et al., 2014; Nugiel et al., 2019; Partanen et al., 2019). This work is accumulating in mathematics (Arsalidou et al., 2018; De Smedt & Ghesquière, 2019) and written expression (e.g., Brownsett & Wise, 2010; Richards et al., 2011). However, brain imaging is not yet sophisticated enough to reliably identify learning disabilities (Seidenberg, 2017), let alone design interventions from the results.

The difficulty with efforts focused on cognitive and neuropsychological assessment in identifying learning difficulties is the poor link between the assessment findings and the intervention strategies that flow from the assessment. Very little well-designed research on the treatment validity of cognitive and neuropsychological assessment or other such approaches can be found. Although advocates will point to specific case studies illustrating their findings, or to examples with small samples of how cognitive processing assessment is applied to intervention development, systematic and well-controlled research on the link between cognitive processing and neuropsychological methods of assessment is yet to be evident in the literature. This may change in the future as we continue to increase our understanding of how the brain works. However, at present, the appeal of basing intervention recommendations on neuropsychological processing assessments remains popular but not scientifically substantiated.

Summary of Indirect Interventions

Overall, the best way to summarize the effects of indirect interventions on academic skills, and academic evaluation models that focus on cognitive processes, is as follows: If the intervention does not involve teaching or practicing the reading, mathematics, or writing skills it is intended to improve, do not expect improvement in reading, mathematics, or writing. The best chances for improving academic skills are to target those skills directly.

Direct Interventions for Academic Problems

Interventions for academic problems are considered *direct* if the skills and behaviors targeted for change are directly involved in the performance of the skill in the natural environment. For example, in reading, interventions that specifically target the skills and knowledge sources necessary for reading words (e.g., letter–sound correspondence, phonemic blending and segmenting, spelling, orthographic knowledge, morphemes) and comprehending text (e.g., fluent reading, vocabulary knowledge, background knowledge) would be considered direct interventions. This contrasts with indirect interventions discussed earlier, which may target sensory processes, motor skills, or cognitive processes. Additionally, instead of intervention based on *aptitude* by treatment interactions, more productive interventions consider skill × treatment interactions that match an ideal strategy to a student's specific academic needs (e.g., Burns et al., 2010; Connor et al., 2009, 2011).

The remainder of this book is focused on Shapiro's four-step model for assessing academic skills problems and developing interventions that directly target them. Most interventions will include strategies that fall within two domains:

1. **Explicit instruction in relevant academic skills.** Explicit instruction refers to clear, unambiguous instruction in which a teacher clearly teaches and demonstrates the skill ("I do"), leads students in performing the skill ("We do"), and monitors and provides immediate feedback on student performance of the skill ("You do"). The skills targeted are those that are directly involved and important for success in the academic skill area. Decades of research have demonstrated that explicit instruction is associated with stronger student outcomes compared to other less explicit or student-centered instruction, especially with struggling learners. The benefits of direct and explicit instructional strategies have been demonstrated for improving reading (Ehri 2020b; Stockard et al., 2018), mathematics computation (Codding et al., 2011; Gersten, Chard, et al., 2009), mathematics word-problem solving (Zheng et al., 2013), and written expression (Gillespie & Graham, 2014). Interventions that involve explicit instruction are discussed across the text.

2. **Active engagement, opportunities to respond, and practice.** The time during which students are actively rather than passively engaged in academic responding, or "engaged time," has a long and consistent history of finding significant relationships to academic performance. Humans learn by *doing* and obtaining feedback on the result. Although observing is important for learning, complex skills also require opportunities to practice and learn through trial and error. Practice is essential because it provides us with experience and allows us to receive feedback, either through observing the outcome or from someone else such as a peer or teacher, on whether our response was accurate.

The association of active engagement in academic tasks to academic skill acquisition has an extensive history (e.g., Berliner, 1979, 1988; Fisher & Berliner, 1985; Frederick, Walberg, & Rasher, 1979; Goodman, 1990; Greenwood, 1991; Greenwood, Horton, et al., 2002; Hall et al., 1982; MacSuga-Gage & Simonsen, 2015; Myers, 1990; Pickens & McNaughton, 1988; Stanley & Greenwood, 1983; Thurlow, Ysseldyke, et al., 1983, 1984; Ysseldyke, Thurlow, et al., 1984, 1988). Accordingly, a number of academic interventions designed specifically to increase opportunities to respond have been developed and are discussed across this text.

In the chapters that follow, a full description is provided of Shapiro's four-step model for assessing children referred for academic skills problems. The reader will learn how assessment can be considered in the context of the students' environment, how it is best focused on skills and behaviors most relevant for academic tasks, and how to identify the most relevant, evidence-based interventions from the assessment findings. Interventions are of central importance to the four-step model; not only are interventions the reason for conducting the assessment in first place, but a student's responsiveness to intervention provides a better understanding of their academic achievement and needs for support moving forward. In this model, and across this text, assessment and intervention are considered together in a dynamic manner.