

DISCUSSION 2

Prior to beginning work on this discussion, read Chapters 5 and 6 in the textbook and the required articles for this week, and view the [IQ: A history of deceit](#) (Links to an external site.) video. For your initial post, you will present at least two viewpoints debating professional approaches to assessment used in psychology for your assigned age group. Please see the list below for your assigned age group. In addition to the required reading, research a minimum of one peer-reviewed article from the Ashford University Library on ability testing research that pertains to your assigned age group.

In your initial post, you must

- Briefly compare and discuss at least two theories of intelligence and the contemporary assessment measures related to those theories.
- Analyze challenges related to assessing individuals in your assigned age group and describe any special ethical and sociocultural issues which must be considered.
- Analyze and provide evidence from validation studies supporting and opposing the use of specific instruments with your assigned population.
- Present the pros and cons of individual versus group assessment of ability.
- Summarize the implications of labelling and mislabeling individuals in your assigned age group as a result of testing and assessment.

Preschool-aged children through age 7 (Assigned Group)

Evidence-Based Assessment of Learning Disabilities in Children and Adolescents

Jack M. Fletcher

*Department of Pediatrics and the Center for Academic and Reading Skills,
University of Texas Health Science Center at Houston*

David J. Francis

*Department of Psychology and the Texas Institute for Measurement, Evaluation and Statistics,
University of Houston*

Robin D. Morris

Department of Psychology, Georgia State University

G. Reid Lyon

Child Development and Behavior Branch, National Institute of Child Health and Human Development

The reliability and validity of 4 approaches to the assessment of children and adolescents with learning disabilities (LD) are reviewed, including models based on (a) aptitude–achievement discrepancies, (b) low achievement, (c) intra-individual differences, and (d) response to intervention (RTI). We identify serious psychometric problems that affect the reliability of models based on aptitude–achievement discrepancies and low achievement. There are also significant validity problems for models based on aptitude–achievement discrepancies and intra-individual differences. Models that incorporate RTI have considerable potential for addressing both the reliability and validity issues but cannot represent the sole criterion for LD identification. We suggest that models incorporating both low achievement and RTI concepts have the strongest evidence base and the most direct relation to treatment. The assessment of children for LD must reflect a stronger underlying classification that takes into account relations with other childhood disorders as well as the reliability and validity of the underlying classification and resultant assessment and identification system. The implications of this type of model for clinical assessments of children for whom LD is a concern are discussed.

Assessment methods for identifying children and adolescents with learning disabilities (LD) are multiple, varied, and the subject of heated debates among practitioners. Those debates involve issues that extend beyond the value of specific tests, often reflecting different views of how LD is best identified. These views reflect variations in the definition of LD and, therefore, variations in what measures are selected to operationalize the definition (Fletcher, Foorman, et al., 2002). Any focus on the “best tests” leads to a hopeless

morass of confusion in an area such as LD that has not successfully addressed the classification and definition issues that lead to identification of who does and who does not possess characteristics of LD. Definitions always reflect an implicit classification indicating how different constructs are measured and used to identify members of the class in terms of similarities and differences relative to other entities that are not considered members of the class (Morris & Fletcher, 1988). For LD, children who are members of this class are historically differentiated from children who have other achievement-related difficulties, such as mental retardation, sensory disorders, emotional or behavioral disturbances, and environmental causes of underachievement, including economic disadvantage, minority language status, and inadequate instruction (Fletcher, Francis, Rourke, Shaywitz, & Shaywitz, 1993; Lyon, Fletcher, & Barnes, 2003). If the classification is valid, children with LD may share characteristics that are similar with other groups of underachievers, but they

Grants from the National Institute of Child Health and Human Development, P50 21888, Center for Learning and Attention Disorders, and National Science Foundation 9979968, Early Reading Development: A Cognitive Neuroscience Approach supported this article.

We gratefully acknowledge contributions of Rita Taylor to preparation of this article.

Requests for reprints should be sent to Jack M. Fletcher, Department of Pediatrics, University of Texas Health Science Center at Houston, 7000 Fannin Street, UCT 2478, Houston, TX 77030. E-mail: Jack.Fletcher@uth.tmc.edu

should also differ in ways that can be measured and that can serve to define and operationalize the class of children and adolescents with LD.

In this article, we consider evidence-based approaches to the assessment of LD in the context of different approaches to the classification and identification of LD. We argue that the measurement systems that are used to identify children and adolescents with LD are inseparable from the classifications from which the identification criteria evolve. Moreover, all measurement systems are imperfect attempts to measure a construct (LD) that operates as a latent variable that is unknowable independently of how it is measured and therefore of how LD is classified. The construct of LD is imperfectly measured simply because the measurement tools themselves are not error free (Francis et al., 2005). Different approaches to classification and definition capitalize on this error of measurement in ways that reduce or increase the reliability of the classification itself. Similarly, evaluating similarities and differences among groups of students who are identified as LD and not LD is a test of the validity of the underlying classification, so long as the variables used to assess this form of validity are not the same as those used for identification (Morris & Fletcher, 1988). As with any form of validity, adequate reliability is essential. Classifications can be reliable and still lack validity. The converse is not true; they cannot be valid and lack reliability. A valid classification of LD predicts important characteristics of the group. Consistent with the spirit of this special section, the most important characteristic is whether the classification is meaningfully related to intervention. For LD, a classification should also predict a variety of differences on cognitive skills, behavioral attributes, and achievement variables not used to form the classification, developmental course, response to intervention (RTI), neurobiological variables, or prognosis (Fletcher, Lyon, et al., 2002).

To address these issues, we consider the reliability and validity of four approaches to the classification and assessment of LD: (a) IQ discrepancy and other forms of aptitude-achievement discrepancy, (b) low achievement, (c) intra-individual differences, and (d) models incorporating RTI and some form of curriculum-based measurement. We consider how each classification reflects the historically prominent concept of "unexpected underachievement" as the key construct in LD assessment (Lyon et al., 2001), that is, what many early observers characterized as a group of children unable to master academic skills despite the absence of known causes of poor achievement (sensory disorder, mental retardation, emotional disturbances, economic disadvantages, inadequate instruction). From this perspective, a valid classification and measurement system for LD must identify a unique group of underachievers that is clearly differentiated from groups with other forms of underachievement.

Defining LD

Historically, definition and classification issues have haunted the field of LD. As reviewed in Lyon et al. (2001), most early conceptualizations viewed LD simply as a form of "unexpected" underachievement. The primary approach to assessment involved the identification of intra-individual variability as a marker for the unexpectedness of LD, along with the exclusion of other causes of underachievement that would be expected to produce underachievement. This type of definition was explicitly coded into U.S. federal statutes when LD was identified as an eligibility category for special education in Public Law 94-142 in 1975; essentially the same definition is part of current U.S. federal statutes in the Individuals with Disabilities Education Act (1997).

The U.S. statutory definition of LD is essentially a set of concepts that in itself is difficult to operationalize. In 1977, recommendations for operationalizing the federal definition of LD were provided to states after passage of Public Law 94-142 to help identify children in this category of special education (U. S. Office of Education, 1977). In these regulations, LD was defined as a heterogeneous group of seven disorders (oral language, listening comprehension, basic reading, reading comprehension, math calculations, math reasoning, written language) with a common marker of intra-individual variability represented by a discrepancy between IQ and achievement (i.e., unexpected underachievement). Unexpectedness was also indicated by maintaining the exclusionary criteria present in the statutory definition that presumably lead to expected underachievement. Other parts of the regulations emphasize the need to ensure that the child's educational program provided adequate opportunity to learn. No recommendations were made concerning the assessment of psychological processes, most likely because it was not clear that reliable methods existed for assessing processing skills and because the field was not clear on what processes should be assessed (Reschly, Hosp, & Smied, 2003).

This approach to definition is now widely implemented with substantial variability across schools, districts, and states in which students are served in special education as LD (MacMillan & Siperstein, 2002; Mercer, Jordan, Allsop, & Mercer, 1996; Reschly et al., 2003). It is also the basis for assessments of LD outside of schools. Consider, for example, the definition of reading disorders in the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994), which indicates that the student must perform below levels expected for age and IQ, and specifies only sensory disorders as exclusionary:

- A. Reading achievement, as measured by individually administered standardized tests of read-

ing accuracy or comprehension, is substantially below that expected given the person's chronological age, measured intelligence, and age-appropriate education.

- B. The disturbance in Criterion A significantly interferes with academic achievement or activities of daily living that require reading skills.
- C. If a sensory deficit is present, the reading difficulties are in excess of those usually associated with it.

The International Classification of Diseases-10 has a similar definition. It differs largely in being more specific in requiring use of a regression-adjusted discrepancy, specifying cut points (achievement two standard errors below IQ) for identifying a child with LD, and expanding the range of exclusions.

Although these definitions are used in what are often disparate realms of practice, they lead to similar approaches to the identification of children and adolescents as LD. Across these realms, children commonly receive IQ and achievement tests. The IQ test is commonly interpreted as an aptitude measure or index against which achievement is compared. Different achievement tests are used because LD may affect achievement in reading, math, or written language. The heterogeneity is recognized explicitly in the U.S. statutory and regulatory definitions of LD (Individuals With Disabilities Education Act, 1997) and in the psychiatric classifications by the provision of separate definitions for each academic domain. However, it is still essentially the same definition applied in different domains. In many settings, this basic assessment is supplemented with tests of processing skills derived from multiple perspectives (neuropsychology, information processing, and theories of LD). The approach boils down to administration of a battery of tests to identify LD, presumably with treatment implications.

Underlying Classification Hypotheses

Implicit in all these definitions are slight variations on a classification model of individuals with LD as those who show a measurable discrepancy in some but not all domains of skill development and who are not identified into another subgroup of poor achievers. In some instances, the discrepancy is quantified with two tests in an aptitude-achievement model epitomized by the IQ-discrepancy approach in the U.S. federal regulatory definition and the psychiatric classifications of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) and the International Classification of Diseases-10. Here the classification model implicitly stipulates that those who meet an IQ-discrepancy inclusionary criterion are different in meaningful ways from those who are underachievers and do not meet the

discrepancy criteria or criteria for one of the exclusionary conditions. Some have argued that this model lacks validity and propose that LD is synonymous with underachievement, so that it should be identified solely by achievement tests (Siegel, 1992), often with some exclusionary criteria to help ensure that the achievement problem is unexpected. Thus, the contrast is really between a two-test aptitude-achievement discrepancy and a one-test chronological age-achievement discrepancy with achievement low relative to age-based (or grade-based) expectations. If processing measures are added, the model becomes a multitest discrepancy model. Identification of a child as LD in all three of these models is typically based on assessment at a single point in time, so we refer to them as "status" models. Finally, RTI models emphasize the "adequate opportunity to learn" exclusionary criterion by assessing the child's response to different instructional efforts over time with frequent brief assessments, that is, a "change" model. The child who is LD becomes one who demonstrates intractability in learning characteristics by not responding adequately to instruction that is effective with most other students.

Dimensional Nature of LD

Each of these four models can be evaluated for reliability and validity. Unexpected underachievement, a concept critically important to the validity of the underlying construct of LD, can also be examined. The reliability issues are similar across the first three models and stem from the dimensional nature of LD. Most population-based studies have shown that reading and math skills are normally distributed (Jorm, Share, Matthews, & Matthews, 1986; Lewis, Hitch, & Walker, 1994; Rodgers, 1983; Shalev, Auerbach, Manor, & Gross-Tsur, 2000; Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992; Silva, McGee, & Williams, 1985). These findings are buttressed by behavioral genetic studies, which are not consistent with the presence of qualitatively different characteristics associated with the heritability of reading and math disorders (Fisher & DeFries, 2002; Gilger, 2002). As dimensional traits that exist on a continuum, there would be no expectation of natural cut points that differentiate individuals with LD from those who are underachievers but not identified as LD (Shaywitz et al., 1992).

The unobservable nature of LD makes two-test and one-test discrepancy models unreliable in ways that are psychometrically predictable but not in ways that simply equate LD with poor achievement (Francis et al., 2005; Stuebing et al., 2002). The problem is that the measurement approach is based on a static assessment model that possesses insufficient information about the underlying construct to allow for reliable classifications of individuals along what is essentially an unobservable dimension. If LD was a

manifest concept that was directly observable in the behavior of affected individuals, or if there were natural discontinuities that represented a qualitative breakpoint in the distribution of achievement skills or the cognitive skills on which achievement depends, this problem would be less of an obstacle. However, like achievement or intelligence, LD is a latent construct that must be inferred from the pattern of performance on directly observable operationalizations of other latent constructs (namely, test scores that index constructs like reading achievement, phonological awareness, aptitude, and so on). The more information available to support the inference of LD, the more reliable (and valid) that inference becomes, thus supporting the fine-grained distinctions necessitated by two-test and one-test discrepancy models. To the extent that the latent construct, LD, is categorical, by which we mean that the construct indexes different classes of learners (i.e., children who learn differently) as opposed to simply different levels of achievement, then systems of identification that rely on one measurable variable lack sufficient information to identify the latent classes and assign individuals to those classes without placing additional, untestable, and unsupported constraints on the system. It is simply not possible to use a single mean and standard deviation and to estimate separate means and standard deviations for two (or more) unobservable latent classes of individuals and determine the percentage of individuals falling into each class, let alone to classify specific individuals into those classes. Without constraints, such as specifying the magnitude of differences in the means of the latent classes, the ratio of standard deviations, and the odds of membership in the two (or more) classes, the system is under-identified, which simply means that there are many different solutions that cannot be distinguished from one another.

When the system is under-identified, the only solution is to expand the measurement system to increase the number of observed relations, which in one sense is what intra-individual difference models attempt by adding assessments of processing skills. Other criteria are necessary because it is impossible to uniquely identify a distinct subgroup of underachieving individuals consistent with the construct of LD when identification is based on a single assessment at a single time point. Adding external criteria, such as an aptitude measure or multiple assessments of processing skills, increases the dimensionality of the measurement system and makes latent classification more feasible, even when the other criteria are themselves imperfect. But the main issues for one-test, two-test, and multitest identification models involve the reliability of the underlying classifications and whether they identify a unique subgroup of underachievers. In the next section, we examine variations in reliability and validity for each of these models, fo-

cus on the importance of reliability, as the validity of the classifications can be no stronger than their reliability.

Models Based on Two-Test Discrepancies

Although the IQ-discrepancy model is the most widely utilized approach to identifying LD, there are many different ways to operationalize the model. For example, some implementations are based on a composite IQ score, whereas others utilize either a verbal or nonverbal IQ score. Other approaches drop IQ as the aptitude measure and use a measure such as listening comprehension. In the validity section, we discuss each of these approaches. The reliability issues are similar for each example of an aptitude-achievement discrepancy.

Reliability

Specific reliability problems for two-test discrepancy models pertain to any comparison of two correlated assessments that involve the determination of a child's performance relative to a cut point on a continuous distribution. Discrepancy involves the calculation of a difference score (D) to estimate the true difference (Δ) between two latent constructs. Thus, discussions about discrepancy must distinguish between problems with the manifest (i.e., observed) difference (D) as an index of the true difference (Δ) but also must consider whether the true difference (Δ) reflects the construct of interest. Problems with the reliability of D based on differences between two tests are well known, albeit not in the LD context (Bereiter, 1967). However, there is nothing that fundamentally limits the applicability of this research to LD if we are willing to accept a notion of Δ as a marker for LD. There are major problems with this assumption that are reviewed in Francis et al. (2005). The most significant is regression to the mean. On average, regression to the mean indicates that scores that are above the mean will be lower when the test is repeated or when a second correlated test is used to compute D . In this example, individuals who have IQ scores above the mean will obtain achievement test scores that, on average, will be lower than the IQ test score because the achievement score will move toward the mean. The opposite is true for individuals with IQ scores below the mean. This leads to the paradox of children with achievement scores that exceed IQ, or the identification of low-achieving, higher IQ children with achievement above the average range as LD.

Although adjusting for the correlation of IQ and achievement helps correct for regression effects (Reynolds, 1984-1985), unreliability also stems from the attempt to assess a person's standing relative to a cut point on a continuous distribution. As discussed in the

following section on low achievement models, this problem makes identification with a single test—even one with small amounts of measurement error—potentially unreliable, a problem for any status model.

None of this discussion addresses the validity question concerning Δ . Specifically, does Δ embody LD as we would want to conceptualize it (e.g., as unexpected underachievement), or is Δ merely a convenient conceptualization of LD because it is a conceptualization that leads directly to easily implemented, operational definitions, however flawed they might be?

Validity

The validity of the IQ-discrepancy model has been extensively studied. Two independent meta-analyses have shown that effect sizes on measures of achievement and cognitive functions are in the negligible to small range (at best) for the comparison of groups formed on the basis of discrepancies between IQ and reading achievement versus poor readers without an IQ discrepancy (Hoskyn & Swanson, 2000; Stuebing et al., 2002), findings similar to studies not included in these meta-analyses (Stanovich & Siegel, 1994). Other validity studies have not found that discrepant and nondiscrepant poor readers differ in long-term prognosis (Francis, Shaywitz, Stuebing, Shaywitz, & Fletcher, 1996; Silva et al., 1985), response to instruction (Fletcher, Lyon, et al., 2002; Jiménez et al., 2003; Stage, Abbott, Jenkins, & Berninger, 2003; Vellutino, Scanlon, & Jaccard, 2003), or neuroimaging correlates (Lyon et al., 2003; but also see Shaywitz et al., 2003, which shows differences in groups varying in IQ but not IQ discrepancy). Studies of genetic variability show negligible to small differences related to IQ-discrepancy models that may reflect regression to the mean (Pennington, Gilger, Olson, & DeFries, 1992; Wadsworth, Olson, Pennington, & DeFries, 2000). Similar empirical evidence has been reported for LD in math and language (Fletcher, Lyon, et al., 2002; Mazzocco & Myers, 2003). This is not surprising given that the problems are inherent in the underlying psychometric model and have little to do with the specific measures involved in the model except to the extent that specific test reliabilities and intertest correlations enter into the equations.

Despite the evidence of weak validity for the practice of differentiating discrepant and nondiscrepant students, alternatives based on discrepancy models continue to be proposed, and psychologists outside of schools commonly implement this flawed model. However, given the reliability problems inherent in IQ discrepancy models, it is not surprising that these other attempts to operationalize aptitude-achievement discrepancy have not met with success. In the Stuebing et al. (2002) meta-analysis, 32 of the 46 major studies had a clearly defined aptitude measure. Of these studies,

19 used Full Scale IQ, 8 used Verbal IQ, 4 used Performance IQ, and 1 study used a discrepancy of listening comprehension and reading comprehension. Not surprisingly, these different discrepancy models did not yield results that were different from those when a composite IQ measure was utilized. Neither Fletcher et al. (1994) nor Aaron, Kuchta, and Grapenthin (1988) were able to demonstrate major differences between discrepant and low achievement groups formed on the basis of listening comprehension and reading comprehension.

The differences in these models involve slight changes in who is identified as discrepant or low achieving depending on the cut point and the correlation of the aptitude and achievement measures. The changes simply reflect fluctuations around the cut point where children are most similar. It is not surprising that effect sizes comparing poor achievers with and without IQ discrepancies are uniformly low across these different models. Current practices based on this approach to identification of LD epitomized by the federal regulatory definition and psychiatric classifications are fundamentally flawed.

One-Test (Low Achievement) Models

Reliability

The measurement problems that emerge when a specific cut point is used for identification purposes affect *any* psychometric approach to LD identification. These problems are more significant when the test score is not criterion referenced, or when the score distributions have been smoothed to create a normal univariate distribution. To reiterate, the presence of a natural breakpoint in the score distribution, typically observed in multimodal distributions, would make it simple to validate cut points. But natural breaks are not usually apparent in achievement distributions because reading and math achievement distributions are normal. Thus, LD is essentially a dimensional trait, or a variation on normal development.

Regardless of normality, measurement error attends any psychometric procedure and affects cut points in a normal distribution (Shepard, 1980). Because of measurement error, any cut point set on the observed distribution will lead to instability in the identification of class members because observed test scores will fluctuate around the cut point with repeated testing or use of an alternative measure of the same construct (e.g., two reading tests). This fluctuation is not just a problem of correlated tests or simply a matter of setting better cut scores or developing better tests. Rather, no single observed test score can capture perfectly a student's ability on an imperfectly measured latent variable. The fluctuation in identifications will vary across different tests, depending in part on the measurement

error. In both real and simulated data sets, fluctuations in up to 35% of cases are found when a single test is used to identify a cut point. Similar problems are apparent if a two-test discrepancy model is used (Francis et al., 2005; Shaywitz et al., 1992).

This problem is less of an issue for research, which rarely hinges on the identification of individual children. Thus, it does not have great impact on the validity of a low achievement classification because, on average, children around the cut point who may be fluctuating in and out of the class of interest with repeated testing are not very different. However, the problems for an individual child who is being considered for special education placement or a psychiatric diagnosis are obvious. A positive identification in either example often carries a poor prognosis.

Validity

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

Despite this evidence for validity, concerns emerge about definitions based solely on achievement cut points. Simply utilizing a low achievement definition, even when different exclusionary criteria are applied, does not operationalize the true meaning of unexpected underachievement. Although such an approach to identification is deceptively simple, it is arguable whether the subgroups that remain represent a unique group of underachievers. For example, how well are underachievers whose low performance is attributed to LD differentiated from underachievers whose low performance is attributed to emotional disturbance, economic disadvantage, or inadequate instruction (Lyon et al., 2001)? To use the example of word recognition,

there is little evidence that these subgroups vary in terms of phonological awareness or other language tasks, RTI, or even neuroimaging correlates. In this respect, the validity is weak because the underlying construct of LD is not adequately assessed. Additional criteria are needed, but simply adding a single aptitude measure *decreases* reliability and does not add to the validity of a low achievement definition.

Models Based on Intra-Individual Differences

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

This statement proposes intra-individual differences as a marker for unexpected underachievement. As opposed to a single marker such as IQ discrepancy or low achievement, unexpectedness is operationalized as unevenness in scores across multiple tests. The person identified as LD (by definition) has strengths in many areas of cognitive or neuropsychological function but weaknesses in core attributes that lead to underachievement. The LD is unexpected because the weaknesses lead to selected and narrow difficulties with achievement and adaptive functions. Proponents of this view believe that such approaches identify children as LD based on profiles across tests that differentiate types of LD and also differentiate LD from other childhood disorders, such as mental retardation and behavioral disorders such as attention deficit hyperactivity disorder (ADHD). This approach leads to definitions based on inclusionary criteria in which children are identified as LD based on characteristics that relate to intra-individual differences (Lyon et al., 2001).

Reliability

In essence, the intra-individual difference model employs a multitest discrepancy approach and carries with it the problems involved with estimation of discrepancies and cut points. These problems are inherent in any attempt to identify a person as LD (Fletcher et

al., 2003). However, examining patterns of test scores has long been favored by clinical neuropsychologists, largely because it seems to correspond more closely with clinical practice and because it adds information to the decision-making process (see the elegant discussion of differential test scores, discrepancies, and profiles in Rourke, 1975). The unique reliability issue involves the idea that LD is represented by unevenness in test profiles. This may be true, but does this observation mean that children with flatter profiles are not LD? Severity is correlated with the shape of a profile due to the lack of independence of different tests that might be used to construct the profile (Morris, Fletcher, & Francis, 1993). Children with increasingly severe reading problems, for example, will show increasingly flat profiles across processing measures (e.g., phonological awareness, rapid naming, and vocabulary) in direct correspondence to severity because all these measures are moderately correlated. Thus, if the inclusionary criterion for the presence of LD is evidence of a discrepancy in neuropsychological or processing skills, such an approach may exclude the most severely impaired children, irrespective of global measures such as IQ, because more severely impaired children are less likely to show skill discrepancies due to the inter-correlation of the tests (Morris et al., 1993, 1998).

Validity

A major assumption of a multitest intra-individual differences model is that identification based on performance patterns will lead to enhanced treatment of children with LD. It is commonly assumed that such tests point out areas that need intervention. However, there is little evidence that strengths and weaknesses in processing skills are related to intervention outcomes. It is well established that training in underlying processes does not usually generalize into the related academic area (Lyon & Moats, 1988; Reschly, Tilly, & Grimes, 1999; Vellutino, 1979). For example, training on phonological awareness skills without explicit transfer to a letter component produces gains in phonological awareness but not in reading (National Reading Panel, 2000). Training in auditory or visual perceptual skills does not lead to better outcomes for children identified as "auditory" or "visual" learners (Lyon, Fletcher, Fuchs, & Chhabra, in press; Vellutino, Fletcher, Scanlon, & Snowling, 2004).

There is support for the idea that intra-individual differences identify some children as LD, epitomized by the link of dyslexia with word recognition and phonological processing (Vellutino et al., 2004). Even here the intra-individual differences model focuses on skills that are only correlated with the achievement domain. Simply identifying children with LD based solely on processing skills is questionable and would likely yield

many false positive identifications of children as LD without achievement difficulties (Torgesen, 2002). The reliability of many processing measures is lower than those associated with the achievement (or IQ) domain, so such false positives should be expected. Other than the word recognition-phonological processing link, relations of processing and other forms of LD are not well established (Torgesen, 2002). Finally, what do we learn about variability in processing skills that is not apparent in profiles across achievement domains (Fletcher et al., 2003)? In fact, the model has the most validity at the level of achievement markers but simply collapses into a low achievement model in the absence of processing measures. Thus, if we accept the notion that specific discrepancies in cognitive domains are a unique marker for LD, given that the processing measures are usually linked to an achievement domain, what is unique about variations in processing skills that is not apparent in variations in achievement domains? Would we eliminate as LD students who have difficulties in reading, math, and writing? This is not viable, as impairments in all domains often occur in non-mentally retarded children with language-based difficulties.

Models Incorporating RTI

An alternative approach to status models that would increase the reliability of these would increase the number of time points whereby a child was assessed. Shepard (1980), for example, proposed that IQ discrepancies could be assessed more reliably if a child was tested four times. The impracticality of such an approach, which would require about 10 to 12 hr per child, is obvious, not to mention that even more resources would be devoted to determination of eligibility, taking away funds and time needed for intervention.

Another approach to increasing the number of time points would involve much shorter assessments of key achievement skills over time. These approaches, or RTI models, typically involve identification practices based in part on multiple short assessment probes of knowledge and performance in a specific academic domain, such as reading or math (Fuchs & Fuchs, 1998). By linking multiple assessments to specific attempts to intervene with the child, the construct of unexpected underachievement can be operationalized, in part, on the basis of nonresponsiveness to instruction to which most other students respond (Gresham, 2002). In fact, this is still a variation of a discrepancy model, but the advantage is that the model is better identified because of multiple short assessments of a key attribute (e.g., reading, math) over time.

Such models have been proposed in several recent consensus reports that address LD identification (Bradley, Danielson, & Hallahan, 2002; President's

Commission on Excellence in Special Education, 2002), most notably in a recent report of the National Research Council (Donavon & Cross, 2002). These reports suggest that one criterion for LD identification is when a student does not respond to high-quality instruction and intervention. The implementation of this approach requires frequent monitoring of progress as the student receives the intervention (Fuchs & Fuchs, 1998). Speece and Case (2001) found that serial assessments of growth and level of performance in reading fluency predicted reading problems in at-risk children better than a single assessment of fluency. This approach is anchored in a system known as curriculum-based measurement, where the assessments themselves have adequate reliability and are constantly improving (Fuchs & Fuchs, 1999; Shinn, 1998).

Reliability

Are RTI approaches that involve multiple assessments over time psychometrically more reliable than traditional approaches to LD identification? An approach based on multiple measures over time has the potential to reduce the difficulties encountered with reliance on a single assessment at a single time point. Certainly the reliability of the multiple assessment approach is greater than if the single assessment is used to form a discrepancy, because typically the discrepancy will be a poorer (i.e., less reliable) measure of the true difference than the observed measures of their respective underlying constructs. Focusing on successive measurements over time has the effect of moving the identification process from "ability-ability" comparisons (two different abilities compared at one point in time) to "ability change" models (same ability over time). Such approaches have the potential to ameliorate the difficulties associated with ability-ability discrepancies, whether univariate or bivariate, because they involve the use of more than two assessment time points. Generally, the more information that is brought to bear on any eligibility or diagnostic decision, the more reliable the decision, although it is certainly possible to create counterexamples by combining information from irrelevant or confounding sources. Such irrelevancies are not likely to be introduced by assessing the same skill over time as in a model that incorporates RTI, when that skill was previously deemed relevant to assess at a single time point.

Conceptually, the study of change is made more feasible by the collection of multiple assessments because the precision by which change can be measured improves as the number of time points increases (Rogosa, 1995). When more than two assessment time points are collected, the reliability of estimated change can also be estimated directly from the data, and the imprecision inherent in individual estimates can be used to provide im-

proved estimates of growth parameters for individual students as well as for groups of students. If change is not linear, the use of four or more time points can map the form of growth. And for those who favor status models over change or learning models, it remains possible to use the intercept term in the individual growth model as an estimate of status. This intercept provides a more precise estimate of true status at any single point in time than would any single assessment.

These approaches are not without difficulty. The introduction of serial assessments has not eliminated the necessity of indirect estimation of the parameters of interest. In the discrepancy model, D is used to estimate Δ . A model incorporating RTI uses a complex function of the observed data for individual i as well as the data from many other individuals to estimate each of the π_{ij} , the j true learning parameters for individual i . Different approaches to this estimation problem have varying strengths and weaknesses but all will make assumptions about the arithmetic form of the model, the distribution of the learning parameters, and the distributions of the errors. The ramifications of these assumptions for inferences about individual learning parameters must be studied in the LD context.

Models based on RTI also involve imperfect measures that include measurement error (Fletcher et al., 2003). However, this problem is reduced because of the use of multiple assessments and the borrowing of precision from the entire collection of data to provide a more precise estimate of the growth parameters of each individual. Thus, it becomes possible to estimate a child's "true" status more precisely as well as to estimate the rate of skill acquisition and to use these estimates as indicators of LD. In addition, this approach to estimation makes assumptions about the distribution of errors of measurement. In some cases, errors might be assumed to be uncorrelated. Again, this assumption must be examined in terms of its importance to inferences about individual status and rates of learning. In many cases, the inclusion of multiple assessment time points will allow this assumption to be relaxed, and the correlation among errors of measurement can be estimated and taken into account in forming inferences about individual status and rates of learning.

There still could be a need to identify individual children as LD based on cut points unless the entire process devolves to clinical judgment. Models that include RTI do not solve the issue of the dimensional versus categorical nature of LD. Determining cut points and benchmarks, for example, will continue to be an arbitrary process until cut points are linked to functional outcomes (Cisek, 2001), an issue never really addressed in LD identification for any identification model. However, models that include RTI have the promise of incorporating functional outcomes because they are tied to intervention response.

Validity

The introduction of serial assessments has an advantage beyond any statistical advantage it may confer for the estimation of individual's true status. Specifically, the introduction of serial assessments brings learning and the measurement of change to the forefront in conceptualizations of LD. The collection of serial assessments under specified conditions of effective instruction simultaneously focuses the definition of LD on a failure to learn, where learning can be measured more directly. Moreover, the specific instructional elements and the conditions under which they are implemented can be described, thereby providing a clearer basis for the expectation of learning and the unexpectedness of any failure to learn. Finally, focusing on multiple assessments in a RTI model has the advantage of clearly tying the identification process to the most important component of the construct of LD, which is unexpected underachievement. Models that incorporate RTI may identify a unique group of children that can be clearly differentiated from other low achievers in terms of cognitive correlates, prognosis, and even neurobiological factors.

Studies of children defined using different methods as responders and nonresponders clearly show significant differences in cognitive skills. For example, Stage et al. (2003), Vellutino et al. (2003), and Vaughn, Linan-Thompson, and Hickman-Davis (2003) found that nonresponders to early intervention differed from responders in both preintervention achievement scores and preintervention cognitive tasks. Nonresponders typically had more severe deficits in both reading-related factors (e.g., phonemic awareness, fluency) and reading skills. In recent imaging studies involving both early intervention and remediation of older students (see Fletcher, Simos, Papanicolaou, & Denton, 2004), we likewise found that individuals who were nonresponders showed more severe reading difficulties prior to intervention. More dramatic were the differences in neuroimaging correlates between those who responded to intervention and those who did not. We have found that nonresponders persist with a brain activation pattern that generally demonstrates a failure to activate left hemisphere areas known to be involved in the development of reading skills. In fact, those who were nonresponders showed predominant right-hemisphere activity much like that observed in children and adults with identified reading disabilities (Fletcher et al., 2004).

Implications for Clinical Assessments

This review of classification models may seem removed from the question of how to conduct clinical assessments of children suspected of LD. In fact, when a

psychologist conducts any assessment for LD, the selection of tests reflects the underlying classification model and the constructs it specifies. If the psychologist or educator adopts an aptitude-achievement discrepancy model, the primary tools will be the tests used to operationalize aptitude (e.g., IQ) and achievement. If the clinician adopts a low achievement model, aptitude will not be measured—just achievement. An intra-individual differences model will require neuropsychological or cognitive processing measures. If a model is used that incorporates RTI, assessments of the integrity of the implementation of the intervention and progress monitoring assessments are necessary.

In evaluating models, we found little evidence that supports the aptitude-achievement and intra-individual difference models. Both involve the assessment of cognitive processes that do not contribute to the identification of a unique group of underachievers with LD and have serious reliability problems. The low achievement model has more reliability and validity but does not identify a unique group of underachievers. RTI criteria may permit identification of a unique group of underachievers but by themselves are not sufficient for identification of LD. Combining the strengths of the low achievement and RTI models leads to a hybrid model that invokes concepts of low achievement and RTI. This model can be expanded to incorporate assessment of contextual factors and other disorders that should be evaluated because of the need for differential treatment (Fletcher, Foorman, et al., 2002).

Learning disorders attributable to mental retardation, sensory problems (blindness, deafness), language status (e.g., English as a second language), or transient factors (adjustment difficulties, disruption of the home or school environment) should not be identified as LD. We have not included economic disadvantage, comorbid emotional and behavior disorders, or established neurological disorders as exclusionary criteria and would stipulate that the only way to exclude LD in children with these associated conditions is to provide an intervention that is appropriate and evaluate RTI. A classification of LD may exclude children with emotional or neurological disorders, or those who are economically disadvantaged from the LD category because of policy or resource issues—all are eligible for special education—but children with these associated conditions have forms of underachievement that are difficult to distinguish from those in children with LD. In the end, LD should be identified only after adequate opportunity to learn has been systematically evaluated. Those who do not respond to intervention need more specialized, individualized, and intensive treatments, as well as the probable conferment of disability status and the civil rights protections that come with identification. It is the intractability as indicated by an inadequate response to quality instruction that must be present to identify a child as LD. If a child responds, LD is

not indicated. Any child who has achievement difficulties should receive intervention, whether it is tutoring with support by a college student or intensive intervention by an experienced, well-trained academic therapist.

This model is quite different from the one that child clinical and other psychologists have utilized for the past few decades, and some may respond by suggesting that this model can only be implemented in schools. In fact, we argue that in the absence of an evaluation of RTI, LD should not be identified in any setting—school, clinic, hospital, and so on. We conceptualize traditional clinical evaluations as an opportunity to identify children as “at risk” for LD and to intervene with any child who is struggling to achieve. In schools, screening for reading problems can be done on a large-scale basis in kindergarten and Grade 1 as advocated in Donovan and Cross (2002) and implemented in states such as Texas (Fletcher, Foorman, et al., 2002). Those who are identified as at-risk have their progress monitored and receive increasingly intense, multitiered interventions that may eventuate in identification for special education in LD (Vaughn & Fuchs, 2003). In a multitiered intervention approach, children are screened for risk characteristics, such as weaknesses in letter-sound knowledge and phonological awareness in kindergarten and word reading in Grade 1, with immediate monitoring of progress (Torgesen, 2002). Depending on the rate of progress, interventions are intensified and modified in an effort to accelerate the rate of development of an academic skill. Children are not identified with disabilities until the final tier of the process.

Evaluations outside of schools should utilize a similar approach based on measurement of the three components of the hybrid model proposed by the consensus group in Bradley et al. (2002): (a) low achievement, (b) RTI, and (c) consideration of contextual factors and exclusions. Any psychological evaluation of a child or adolescent should consider the relevant achievement constructs (see following section) that represent the different types of LD. If there is evidence of low achievement, the focus should not be on extensive assessments of processing skills but on referral to an appropriate source for intervention. The psychologist should expect to have a working relationship with the intervention source so that RTI will be measured. This means that clinical child psychologists must be knowledgeable about educational interventions and prepared to develop a treatment plan that incorporates this form of intervention, just as they may be prepared to work with a physician around medication for problems with attention or anxiety. It is perfectly reasonable to ask the child to return every 4 to 6 months to repeat achievement tests and independently evaluate progress in conjunction with more frequent assessments of progress obtained by the intervention source.

The psychologist should also evaluate for other problems that may be associated with low achievement to adequately plan treatment. If mental retardation is suspected, IQ, adaptive behavior, and related assessments consistent with this classification can be administered. But note that if the child or adolescent has achievement scores in reading comprehension or math that are within two standard deviations of the mean (consistent with traditional legal definitions of mental retardation), or development of adaptive behavior obviously inconsistent with mental retardation, assessment of IQ is not necessary as such levels of performance preclude mental retardation. Some children may have oral language disorders that require speech and language intervention that will require referral and additional evaluation. Screening with vocabulary measures and through interacting with the child will help identify these children; the vocabulary screen will also help identify children who may benefit from additional intellectual screening. Many children with achievement difficulties or LD also have comorbid difficulties with attention and both internalizing and externalizing psychopathology. These disorders need to be assessed and treated, as simply referring a child for educational intervention without addressing comorbidities will surely increase the probability of a poor RTI. We believe that no clinical evaluation of a child should be conducted without a documentation of achievement levels through direct assessment or school report of such an assessment. If achievement deficits are apparent, intervention of some sort should be provided. It is not likely that treating a child for a comorbid disorder, such as ADHD, will result in improved levels of achievement in the absence of educational intervention.

Altogether, we are suggesting that from the perspective of LD, psychologists should perform assessments for emotional and behavioral disorders consistent with other articles in this special section. For LD, they need to administer achievement tests and evaluate RTI. This is regardless of subdiscipline (e.g., school psychologist, child clinical psychologist, neuropsychologist) or setting. To evaluate achievement, individualized norm-referenced assessments should be conducted. RTI requires assessments of intervention integrity and monitoring of progress.

Evaluating Achievement

Identifying specific achievement tests is not difficult, although tests for some domains are better developed than others. Lyon et al. (2003) suggested that LD represented six major achievement types, including (a) word recognition; (b) reading fluency; (c) reading comprehension; (d) mathematics computations; (e) reading and math, which is not really a comorbid association but a more severe reading problem with distinct math difficulties; and (f) written expression, which

could involve spelling, handwriting, or text generation. These patterns were drawn from the research literature (e.g., Rourke & Finlayson, 1978; Siegel & Ryan, 1988; Stothard & Hulme, 1996), but an extensive discussion of the evidence for these types is beyond the scope of this article (see Lyon et al., 2003). The assessment implications are straightforward. Many children and adolescents will have difficulties in more than one domain, so a thorough assessment of academic achievement is very important.

A set of achievement tests should be used. It is helpful to use tests from the same battery because the normative group is the same, which facilitates comparisons across tests. However, the battery from which these tests are chosen is less important than the constructs that are measured. Any single battery has strengths and weaknesses that can be supplemented with measures from other assessments. Given the suggestion that six types of LD may exist, the important constructs are word recognition, reading fluency, reading comprehension, math computations, and written expression. We usually assess spelling as a screen for written expression and handwriting difficulties and math and writing fluency as supplemental assessments.

Table 1 outlines these constructs and how they can be assessed with the commonly used Woodcock-Johnson Achievement Battery-III (WJ; Woodcock, McGrew, & Mather, 2001) or the Wechsler Individual Achievement Test-II (WIAT; Wechsler, 2001). We use the WJ and WIAT because they meet established criteria for reliability (internal consistency and test-retest) and validity (construct and concurrent) and were developed to account for variations in ethnicity and socioeconomic status. In particular, the normative sampling took into account this type of variation, and analyses (differential item functioning) were conducted to identify items that were not comparable across these sources of normative variation. There are also other norm-referenced assessments that can be used to supplement the WJ or WIAT, which we discuss later. Few of these supplemental measures have been developed with the

care of the WJ or the WIAT, particularly with regard to the adequacy of the normative base and attempts to address different forms of normative variation.

Table 1 should not be taken to indicate that there are 11 different types of LD, one for each test. To reiterate, many children have problems in multiple domains. The pattern of academic strengths and weaknesses is an important consideration (Fletcher, Foorman, et al., 2002; Rourke, 1975). Table 1 identifies constructs and core tests that would be administered to every child and supplemental tests that would be used if there were concerns about a particular academic domain. If the referral indicated concerns about a particular area, additional tests from other measures would be used. Most children with significant academic problems where LD may eventually be a concern have difficulty with word recognition and consequently tend to have problems across domains of reading. Going beyond the core tests is usually not necessary if the child has problems with word recognition. Isolated problems with reading comprehension and written expression occur infrequently. If the problem is specifically math, using assessments in addition to the WJ or WIAT is helpful in ensuring that the deficiency is not just a matter of attention difficulties.

An advantage of the WJ and WIAT is the assessment of word recognition for both real words and pseudowords, the latter permitting an assessment of the child's ability to apply phonics rules to sound out words. Most achievement batteries assess recognition of real words, which is the essential component. These measures tend to be highly intercorrelated across different assessment batteries, including the Wide Range Achievement Test-III (Wilkinson, 1993), and the Accuracy measure from the Gray Oral Reading Test-Fourth Edition (Wiederholt & Bryant, 2001).

The WJ also has a silent reading speed subtest that, in our assessments, is highly correlated with other fluency measures despite the fact that it is not simply oral reading speed, requiring the child to answer some questions while reading a series of passages for 3 min.

Table 1. Achievement Constructs in Relation to Subtests From the WJ and the WIAT

Construct	WJ Subtest	WIAT Subtest
Core Tests		
Word Recognition	Word Identification	Word Reading
Reading Fluency	Word Attack	Pseudoword Decoding
Reading Comprehension	Reading Fluency	—
Math Computations	Passage Comprehension	Reading Comprehension ^a
Written Expression	Calculation	Numerical Operations
Supplemental Tests		
Math Fluency	Spelling	Spelling
Writing Fluency	Math Fluency	—
Math Concepts	Writing Fluency	—
Written Expression	Quantitative Concepts	—
	Writing Samples	Written Expression ^a

^aAlso assesses fluency

The WIAT permits assessment of reading speed during silent-reading comprehension. Both assessments are easily supplemented with the Test of Word Reading Efficiency (Torgesen, Wagner, & Raschotte 1999), which involves oral reading of real words and pseudowords on a list. The Test of Reading Fluency (Deno & Marston, 2001) is an option that requires text reading. Both measures are quick and efficient, and the former was designed with item analyses addressing differential item responses across ethnic groups. Whenever text is read out loud, fluency can be assessed as words read correctly per minute. The Gray Oral Reading Test—Fourth Edition includes a score for fluency of oral text reading.

Reading comprehension can only be screened with the WJ Passage Comprehension subtests which is a cloze-based assessment in which the child reads a sentence or passage and fills in a blank with a missing word. The Reading Vocabulary subtest is used to create a reading comprehension composite, but it places such a premium on decoding that we usually do not administer it. The WIAT also does not demand much reading of text. Some children who struggle to comprehend text in the classroom do not have difficulties on these subtests because the level of complexity rarely parallels what children are expected to read on an everyday basis. Supplemental assessments using the Group Reading Assessment and Diagnostic Education (Williams, Cassidy, & Samuels, 2001), the Gray Oral Reading Test—Fourth Edition, or even one of the well-constructed reading comprehension assessments from the group-based Stanford Achievement Test—10th Edition (Harcourt Assessment, 2002), Iowa Test of Basic Skills (Hoover, Hieronymous, Frisbie, & Dunbar, 2001), or similar instrument is essential. Often children have had these assessments in school, and it is useful to review results as part of the overall evaluation.

Reading comprehension is a difficult construct to assess (Francis, Fletcher, Catts, & Tomblin, in press). In evaluating comprehension skills, the assessor should attend to the nature of the material the child is asked to read and the response format. Reading comprehension tests vary in what the child reads (sentences, paragraphs, pages), the response format (cloze, open-ended questions, multiple-choice, think aloud), memory demands (answering questions with and without the text available), and how deeper aspects of meaning are evaluated (understanding of the essential meaning vs. literal understanding, vocabulary knowledge and elaboration, ability to infer or predict). It may be difficult to determine the source of the child's difficulties based on a single measure. Thus, if the issue is comprehension and the source is not in the child's word recognition or fluency skills, multiple measures that assess reading comprehension in different ways are needed.

For math, the Calculations subtest of the WJ and Numerical Operations subtest of the WIAT are paper-and-pencil tests of math computations (Table 1). Low scores on this type of task predict variation in cognitive skills depending on other academic strengths and weaknesses (Rourke, 1993). However, low scores could reflect problems with fact retrieval and verbal working memory if word recognition is comparably lower, as opposed to problems with procedural knowledge if word recognition is significantly higher and not deficient. Deficient scores can also reflect problems paying attention, especially in children with ADHD. The math computations subtests from the Wide Range Achievement Test—III is also frequently used and is useful because it is timed and the problems are less organized. The key is the paper-and-pencil assessment of math computations, which is how difficulties in math are typically manifested in children who do not have reading problems. As in reading, assessments of fluency are helpful, although there is no evidence suggestive of a math fluency disorder. In Table 1, the WJ Math Fluency subtest is identified as a supplemental measure, representing a timed assessment of single-digit arithmetic facts that may be helpful for identifying children who lack speed in basic arithmetic skills. Such difficulties make it difficult to master more advanced aspects of mathematics. If an assessment of math concepts is needed, which we would do only if math was an overriding concern, the Quantitative Concepts subtest of the WJ is more useful than the WJ Applied Problems or WIAT Math Reasoning subtests, which introduce word problems that are difficult for children with reading difficulties.

Written expression is most difficult to assess, partly because it is not clear what constitutes a disorder of written expression—spelling, handwriting, or text generation (Lyon et al., 2003). Obviously problems with the first two components will constrain text generation. Spelling should be assessed as it may represent the primary source of difficulty with written expression for children, especially if they also have word-recognition difficulties. The analysis of spelling errors (Rourke, Fisk, & Strang, 1986) can be informative in understanding whether the problem is with the phonological component of language or with the visual form of letters (i.e., orthography). Spelling also permits an informal assessment of handwriting.

Table 1 identifies WJ and WIAT measures of written expression. The utility of these measures is not well established, and the significant generation of text in terms of construction and writing of passages and stories is not really required. As with reading comprehension, it may be important to supplement or even replace this assessment with a test such as the Thematic Maturity subtest of the Test of Written Language (Hammill & Larsen, 1998). Measuring fluency with a measure such as the WJ Writing Fluency subtest may also be

useful. As in reading and math, fluency of writing predicts the quality of composition.

From this type of assessment, characteristic patterns emerge that will demarcate the classification and indicate a need for specific kinds of intervention. For each of the six types of LD, there are interventions with evidence of efficacy that should be utilized in or out of a school setting (Lyon et al., in press). The goal is not to diagnose LD, which is not feasible in a one-shot evaluation for the psychometric and conceptual reasons outlined previously, but to identify achievement difficulties that can be addressed through intervention. If the assessor is knowledgeable about these patterns, very specific intervention recommendations, as well as the need for other assessments, can be made.

Table 2 summarizes achievement patterns that are well established in research (Fletcher, Foorman, et al., 2002; Lyon et al., 2003). Intervention should be considered for any child who performs below the 25th percentile on a well-established assessment, with an understanding that these are not firm cut points and should be evaluated across all the measures. We are not indicating that 25% of all children have a LD, only that scores below the 25th percentile are commonly associated with low performance in school, assuming the cut point is reliably attained. In examining Table 2, the decision rules should not be rigidly applied and are simply guidelines to assist clinicians. Identifying LD is always based on factors beyond just the test scores. The decision process should focus on what is needed for intervention, which requires an assessment of contextual

variables and the presence of comorbid disorders that influence decisions about what sort of plan will be most effective for an individual child. Low achievement is related to many contextual variables, which is why the flexibility in special education guidelines allows interdisciplinary teams to base decisions on factors that go beyond test scores. The purpose of assessment is ultimately to develop an intervention plan.

Evaluating Response to Instruction

Once a child is screened or tested for achievement deficits, progress should be monitored if a problem is identified. It is astonishing that U.S. special education guidelines do not require at least yearly readministration of the achievement tests that were used to justify the placement as one method of assessing the efficacy of the intervention plan. If a child is responding to intervention, his or her rate of development should be accelerated relative to the normative population (i.e., the achievement gap is closed). As part of this assessment of RTI, progress should be monitored on a frequent basis if the problem is with word recognition or fluency, math computations, or spelling. Reading comprehension and higher forms of written expression will show less rapid change and progress, as monitoring tools for these types of problems have not been adequately developed.

Most of the tests mentioned here have alternative forms. But some have been developed to permit assessments with even more frequency and are referred to "curriculum-based assessments" (Fuchs & Fuchs,

Table 2. *Eight Achievement Patterns Associated With Intervention*

1. Decoding and Spelling < 90; Arithmetic one half standard deviation higher than word recognition and spelling and at least 90. This is a pattern characterized by problems with single word decoding skills and better arithmetic ability. Reading comprehension will vary depending on how it is assessed but is usually impaired. Children with this pattern have significant phonological language problems and strengths in spatial and motor skills (Rourke & Finlayson, 1978).
2. Arithmetic < 90, Decoding and Spelling > 90 and at least 7 points higher. Children with difficulties that only involve math show this pattern, which is associated with problems with motor and spatial skills, problem-solving deficiencies, and disorganization (Rourke & Finlayson, 1978). It usually represents problems with math procedures as opposed to math facts (Lyon et al., 2003).
3. Decoding, Comprehension, Spelling, and Arithmetic < 90. This pattern represents a problem with word recognition characterized by language and working memory problems more severe than in children with poor decoding and better development of math skills (Rourke & Finlayson, 1978). The math problem involves learning and retrieving math facts (Lyon et al., 2003).
4. Spelling and Arithmetic < 90, Decoding > 90 and 7 points higher. Essentially the same pattern as Number 3 except the motor (and writing) component is more severe.
5. Reading Comprehension < 90 and 7 points below decoding. This pattern often reflects long-term oral language disorder. Problems with receptive language, short-term memory, and attention are apparent, with strengths in phonological language skills (Stothard & Hulme, 1996).
6. Decoding skills 7 points lower than Comprehension skills and < 90. This pattern reflects a phonological language problem with usually better than average semantic language and spatial skills (Stothard & Hulme, 1996). The pattern is not apparent for reading comprehension measures that are timed or require significant amounts of text reading.
7. Reading Fluency < 90 and < Decoding by one half standard deviation will reflect a problem where accuracy of word reading is less of a problem than automaticity of word reading (Lyon et al., 2003).
8. Spelling < 90. This pattern reflects (a) motor deficits in a young child or (b) residuals of earlier phonological language problems that have been remediated or compensated in older children and adults. The pattern is common in adults with a history of word recognition difficulties. Fluency is often impaired.

Note: The patterns are based on relations of reading decoding, reading fluency, reading comprehension, spelling, and arithmetic. It is assumed that any score below the 25th percentile (standard score = 90) is impaired and that a difference of one half standard deviations is important (± 7 standard score points). The patterns should be considered prototypes and the rules loosely applied (adapted from Fletcher, Foorman, Boudousquie, Barnes, Schatschneider, & Francis, 2002). These patterns are not related to IQ scores.

1999). Such measures are often used by the intervener (e.g., teacher) to document how well a child is responding to instruction. Typically a child would read a short reading passage appropriate for grade level (or do a set of math computations) for 2 to 3 min. The number of words (or math problems) correctly read (or computed) would be graphed over time and compared against grade-level benchmarks, representing a criterion-referenced form of assessment. A child may be screened with these measures, and those performing below the benchmark may be candidates for intervention, especially in schools.

Such assessments should also be accompanied by observations of the integrity of the implementation of the intervention, including the amount of time spent on supplemental instruction, especially if the child does not appear to be making progress. School psychologists are often well prepared in this area of assessment. Although a psychologist operating outside of a school may not be in a position to do curriculum-based assessments or to personally evaluate the intervention, such assessments should be expected, especially if the referral is to a private academic therapist.

A variety of methods have been developed, and the assessments with the most widespread utilization are the Monitoring Basic Skills Progress (Fuchs, Hamlett, & Fuchs, 1990), which assesses reading, math, and spelling fluency, and the Dynamic Indicators of Basic Early Reading Skills (Good, Simmons, & Kame'enui, 2001), a battery of different reading fluency measures. Some of these tools are focused primarily on the lower grades, but the norm-referenced assessments of fluency identified previously—especially if they have alternative forms—can be used with older students. These measures meet accepted psychometric criteria for reliability and validity. The curriculum-based assessment measures have not been assessed as formally for differential item functioning but have been widely employed with school populations that are quite diverse (Fuchs & Fuchs, 1999; Shinn, 1998).

Conclusions

Based on our evaluation of models, we propose a hybrid model that incorporates features of low achievement and RTI models for the identification of children as LD. We specifically do not find evidence to support extensive assessments of cognitive, neuropsychological, or intellectual skills to identify children as LD. Although some may view this model as only for schools, we reject the idea that the routine evaluations done in the past by psychologists and educators outside of school settings are useful for LD. We find little value in the idea of evaluating a child in a single assessment and concluding that the child has LD based on an IQ-achievement discrepancy, low achievement, or profiles on neuropsychy-

chological tests, largely because such assessments are not directly related to treatment and the diagnosis itself is not reliable. As soon as it is apparent that the child has an achievement problem, a referral for intervention should be made and the resources that might be spent on diagnosis should be spent on intervention. Children should not be diagnosed as LD until a proper attempt at instruction has been made. Assessment of achievement skills should be a routine part of any psychological evaluation of a child and cannot be seen as the province of just the schools. Serial monitoring of RTI and the integrity of instruction should be completed before children are identified as LD. There are issues involved in the intervention component, estimation of slope and intercept effects, as well as decisions that have to be made about cut points that will differentiate responders and non-responders (Gresham, 2002). For these reasons alone, RTI cannot be the sole criterion for identification, and flexibility in decision making is required. At the same time, there appears to be considerable validity to this approach, implying that it is indeed possible to reliably identify nonresponders as a group with unexpected underachievement.

In addition to the evidence for validity (and the greater reliability of the underlying psychometric model), the model does not require the use of exclusionary criteria (especially emotional disturbance and economic disadvantage) to operationalize unexpected underachievement, thus capturing the construct of LD. This is an important consideration given the lack of evidence validating classifications that utilize these particular exclusions (Kavale, 1988; Lyon et al., 2001). The model does operationalize the concept of opportunity to learn, which is rarely directly assessed as part of LD identification. It is also a model that can only be implemented in an instructional setting, such as a school, or in clinical settings outside of public schools where remediation is utilized, such as an academic therapy setting. But it is not consistent with the traditional approach to LD identification based on a single administration of a test battery and consideration of a diagnosis, which we believe is an outmoded model that detracts from intervention. In the absence of an attempt to systematically instruct the child, LD cannot be "diagnosed," obviating the traditional "test and treat" model, as identifying LD must be the end product of an attempt to instruct the child (i.e., "treat and test"). This is not a post hoc approach but rather an argument that in the absence of the opportunity to learn exclusion, the concept of LD has no basis in evidence, and low achievement per se is not adequate evidence for LD. Such an approach ties the concept of LD to treatment, which is important. It may be that a single assessment may indicate "risk" or even an achievement disorder. But such an assessment cannot indicate a "disability" in the absence of functional criteria that would include opportunity to learn.

A final comment involves what some will see as equating LD with measurable deficits on achievement tests. Some will argue that the mere presence of a deficit on a measure of processing skills means that person should be identified by LD, in part because of the belief that such deficits indicate a brain anomaly. The most common example is the linking of "executive function" deficits with LD. We argue that the concept of LD is empty without a focus on achievement, largely because it becomes more difficult to identify a unique subgroup representing LD that would be different from other classes of childhood disorders. Executive functions, for example, are often linked to ADHD, but classifications of ADHD based on executive function deficits as assessed by cognitive tests do not have much validity (Barkley, 1997). Moreover, executive function deficits characterize many childhood populations.

More fundamentally, consider an overarching classification of childhood learning and behavioral difficulties. For LD, achievement deficits represent markers for the underlying classification. What distinguishes the LD prototype from, for example, a behavioral disorder such as ADHD is the presence of an achievement deficit. If a child with ADHD has an achievement deficit, it is usually reflective of a comorbid association (Fletcher, Shaywitz, & Shaywitz, 1999). If we expand our classification to mental retardation, the key for differentiating mental retardation from LD (or ADHD) is not just the intelligence test score. Rather, the major difference is in adaptive behavior, where mental retardation should reflect a pervasive deficit in adaptive behavior and LD as a relatively narrow deficit (Bradley et al., 2002). So a classification of these three major disorders requires markers for achievement, attention-related behaviors, and adaptive behavior. In the absence of these types of markers, and a focus on classification, all children with problems are simply disordered and there is no need for assessment because they would all require the same interventions. When LD is tied to levels and patterns of achievement, an evidence base for differential interventions focused on learning in specific academic domains emerges. This is the strongest evidence for the validity of the concept of LD, its classification, and the source of evidence-based approaches to assessment and identification.

References

- Aaron, P. G., Kuchta, S., & Grapenthin, C. T. (1988). Is there a thing called dyslexia? *Annals of Dyslexia*, 38, 33-49.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Barkley, R. A. (1997). *ADHD and the nature of self-control*. New York: Guilford.
- Bereiter, C. (1967). Some persisting dilemmas in the measurement of change. In C. W. Harris (Ed.), *Problems in the measurement of change*. Madison: University of Wisconsin Press.
- Bradley, R., Danielson, L., & Hallahan, D. P. (Eds.). (2002). *Identification of learning disabilities: Research to practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Christensen, C. A. (1992). Discrepancy definitions of reading disability: Has the quest led us astray? *Reading Research Quarterly*, 27, 276-278.
- Cisek, G. J. (2001). Conjectures on the rise and fall of standard setting: An introduction to context and practice. In G. J. Cisek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 3-18). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Deno, S. L., & Marston, D. (2001). *Test of Oral Reading Fluency*. Minneapolis: Educators Testing Service.
- Donavon, M. S., & Cross, C. T. (2002). *Minority students in special and gifted education*. Washington, DC: National Academy Press.
- Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: Genetic dissection of a complex cognitive trait. *Nature: Review of Neuroscience*, 10, 767-780.
- Fletcher, J. M., Fooman, B. R., Boudousquie, A. B., Barnes, M. A., Schatschneider, C., & Francis, D. J. (2002). Assessment of reading and learning disabilities: A research-based, intervention-oriented approach. *Journal of School Psychology*, 40, 27-63.
- Fletcher, J. M., Francis, D. J., Rourke, B. P., Shaywitz, B. A., & Shaywitz, S. E. (1993). Classification of learning disabilities: Relationships with other childhood disorders. In G. R. Lyon, D. Gray, J. Kavanagh, & N. Krasnegor (Eds.), *Better understanding learning disabilities* (pp. 27-55). New York: Brookes.
- Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R., et al. (2002). Classification of learning disabilities: An evidence-based evaluation. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 185-250). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Fletcher, J. M., Morris, R. D., & Lyon, G. R. (2003). Classification and definition of learning disabilities: An integrative perspective. In H. L. Swanson, K. R. Harris, & S. Graham. *Handbook of learning disabilities* (pp. 30-56). New York: Guilford.
- Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., et al. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology*, 85, 1-18.
- Fletcher, J. M., Shaywitz, S. E., & Shaywitz, B. A. (1999). Comorbidity of learning and attention disorders: Separate but equal. *Pediatric Clinics of North America*, 46, 885-897.
- Fletcher, J. M., Simos, P. G., Papanicolaou, A. C., & Denton, C. (2004). Neuroimaging in reading research. In N. K. Duke & M. H. Mallette (Eds.), *Literacy research methods* (pp. 252-286). New York: Guilford.
- Francis, D. J., Fletcher, J. M., Catts, H., & Tomblin, B. (in press). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Current issues in reading comprehension and assessment*. Mahwah, NJ: Lawrence Erlbaum Associate, Inc.
- Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of learning disabilities: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*, 38, 98-108.
- Francis, D. J., Shaywitz, S. E., Stuebing, K. K., Shaywitz, B. A., & Fletcher, J. M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curves analysis. *Journal of Educational Psychology*, 88, 3-17.

- Fuchs, L., & Fuchs, D. (1998). Treatment validity: A simplifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice, 4*, 204-219.
- Fuchs, L., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *The School Psychology Review, 28*, 659-671.
- Fuchs, L. S., Hamlett, C. L., & Fuchs, D. (1990). *Monitoring basic skills progress*. Austin, TX: PRO-ED.
- Gilger, J. W. (2002). Current issues in the neurology and genetics of learning-related traits and disorders: Introduction to the special issue. *Journal of Learning Disabilities, 34*, 490-491.
- Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.
- Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 467-564). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Hammill, D. D., & Larsen, S. C. (1996). *Test of Written Language*. Austin, TX: PRO-ED.
- Harcourt Assessment. (2002). *Stanford Achievement Test* (10th ed.). New York: Author.
- Hoover, A. N., Hieronymous, A. N., Frisbie, D. A., & Dunbar, S. B. (2001). *Iowa Test of Basic Skill*. Itasca, IL: Riverside.
- Hoskyn, M., & Swanson, H. L. (2000). Cognitive processing of low achievers and children with reading disabilities: A selective meta-analytic review of the published literature. *The School Psychology Review, 29*, 102-119.
- Individuals With Disabilities Education Act, 20 U. S. C. 1400 *et seq.* Stat. 34 C. F. R. 300 (1997).
- Jiménez, J. E., del Rosario Ortiz, M., Rodrigo, M., Hernández-Valle, I., Ramirez, G., Estévez, A., et al. (2003). Do the effects of computer-assisted practice differ for children with reading disabilities with and without IQ-achievement discrepancy? *Journal of Learning Disabilities, 36*, 34-47.
- Jorm, A. F., Share, D. L., Matthews, M., & Matthews, R. (1986). Cognitive factors at school entry predictive of specific reading retardation and general reading backwardness: A research note. *Journal of Child Psychology, 27*, 45-54.
- Kavale, K. A. (1988). Learning disability and cultural disadvantage: The case for a relationship. *Learning Disability Quarterly, 11*, 195-210.
- Lewis, C., Hitch, G. J., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9- to 10-year-old boys and girls. *Journal of Child Psychology Psychiatry, 35*, 283-292.
- Lyon, G. R., Fletcher, J. M., & Barnes, M. C. (2003). Learning disabilities. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (2nd ed., pp. 520-586). New York: Guilford.
- Lyon, G. R., Fletcher, J. M., Fuchs, L., & Chhabra, V. (in press). Treatment of learning disabilities. In E. Mash & R. Barkley (Eds.), *Treatment of childhood disorders*. New York: Guilford.
- Lyon, G. R., Fletcher, J. M., Shaywitz, S. E., Shaywitz, B. A., Torgesen, J. K., Wood, F. B., et al. (2001). Rethinking learning disabilities. In C. E. Finn, Jr., R. A. J. Rotherham, & C. R. Hokanson, Jr. (Eds.), *Rethinking special education for a new century* (pp. 259-287). Washington, DC: Thomas B. Fordham Foundation and Progressive Policy Institute.
- Lyon, G. R., & Moats, L. C. (1988). Critical issues in the instruction of the learning disabled. *Journal of Consulting and Clinical Psychology, 56*, 830-835.
- MacMillan, D. L., & Siperstein, G. N. (2002). Learning disabilities as operationally defined by schools. In R. Bradley, L. Danielson, & D. P. Hallahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 287-368). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school age years. *Annals of Dyslexia, 53*, 218-253.
- Mercer, C. D., Jordan, L., Allsop, D. H., & Mercer, A. R. (1996). Learning disabilities definitions and criteria used by state education departments. *Learning Disability Quarterly, 19*, 217-232.
- Morris, R., & Fletcher, J. M. (1988). Classification in neuropsychology: A theoretical framework and research paradigm. *Journal of Clinical and Experimental Neuropsychology, 10*, 640-658.
- Morris, R. D., Fletcher, J. M., & Francis, D. J. (1993). Conceptual and psychometric issues in the neuropsychological assessment of children: Measurement of ability discrepancy and change. In I. Rapin & S. Segalovitz (Eds.), *Handbook of neuropsychology* (Vol. 7, pp. 341-352). Amsterdam: Elsevier.
- Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*, 347-373.
- National Center for Learning Disabilities. (2002). *Achieving better outcomes—maintaining rights: An approach to identifying and serving students with specific learning disabilities*. New York: Author.
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington, DC: National Institute of Child Health and Human Development.
- Pennington, B. F., Gilger, J. W., Olson, R. K., & DeFries, J. C. (1992). The external validity of age- versus IQ-discrepancy definitions of reading disability: Lessons from a twin study. *Journal of Learning Disabilities, 25*, 562-573.
- President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families*. Washington, DC: Department of Education.
- Reschly, D. J., Hosp, J. L., & Smied, C. M. (2003). *And miles to go State SLD requirements and authoritative recommendations*. National Research Center on Learning Disabilities. Retrieved July 22, 2003, from the World Wide Web: www.nrcld.org
- Reschly, D. J., Tilly, W. D., & Grimes, J. P. (1999). *Special education in transition: Functional assessment and noncategorical programming*. Longmont, CO: Sopris West.
- Reynolds, C. (1984-1985). Critical measurement issues in learning disabilities. *Journal of Special Education, 18*, 451-476.
- Rodgers, B. (1983). The identification and prevalence of specific reading retardation. *British Journal of Educational Psychology, 53*, 369-373.
- Rogosa, D. (1995). Myths about longitudinal research (plus supplemental questions). In J. M. Gottman (Ed.), *The analysis of change* (pp. 3-66). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Rourke, B. P. (1975). Brain-behavior relationships in children with learning disabilities: A research program. *American Psychologist, 30*, 911-920.
- Rourke, B. P. (1993). Arithmetic disabilities specific or otherwise: A neuropsychological perspective. *Journal of Learning Disabilities, 26*, 214-226.
- Rourke, B. P., & Finlayson, M. A. J. (1978). Neuropsychological significance of variations in patterns of academic performance: Verbal and visual-spatial abilities. *Journal of Pediatric Psychology, 3*, 62-66.
- Rourke, B. P., Fisk, J., & Strang, J. (1986). *Neuropsychological assessment of children*. New York: Guilford.
- Shalev, R. S., Auerbach, J., Manor, O., & Gross-Tsur, V. (2000). Developmental dyscalculia: Prevalence and prognosis. *European Adolescent Psychiatry, 9*, 58-64.

- Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Distribution and temporal stability of dyslexia in an epidemiological sample of 414 children followed longitudinally. *New England Journal of Medicine*, *326*, 145–150.
- Shaywitz, S. E., Shaywitz, B. A., Fulbright, R. K., Skudlarski, P., Mencl, W. E., Constable, R. T., et al. (2003). Neural systems for compensation and persistence: Young adult outcomes of childhood reading disability. *Biological Psychiatry*, *54*, 25–33.
- Shepard, L. (1980). An evaluation of the regression discrepancy method for identifying children with learning disabilities. *Journal of Special Education*, *14*, 79–91.
- Shinn, M. R. (1998). *Advanced applications of curriculum-based measurement*. New York: Guilford.
- Siegel, L. S. (1992). An evaluation of the discrepancy definition of dyslexia. *Journal of Learning Disabilities*, *25*, 618–629.
- Siegel, L. S., & Ryan, E. (1988). Working memory in subtypes of learning disabled children. *Journal of Clinical and Experimental Neuropsychology*, *10*, 55.
- Silva, P. A., McGee, R., & Williams, S. (1985). Some characteristics of 9-year-old boys with general reading backwardness or specific reading retardation. *Journal of Child Psychology and Psychiatry*, *26*, 407–421.
- Speece, D. L., & Case, L. P. (2001). Classification in context: An alternative approach to identify early reading disability. *Journal of Educational Psychology*, *93*, 735–749.
- Stage, S. A., Abbott, R. D., Jenkins, J. R., & Berninger, V. W. (2003). Predicting response to early reading intervention from verbal IQ, reading-related language abilities, attention ratings, and verbal IQ-word reading discrepancy: Failure to validate discrepancy method. *Journal of Learning Disabilities*, *36*, 24–33.
- Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, *86*, 24–53.
- Stothard, S. E., & Hulme, C. (1996). A comparison of reading comprehension and decoding difficulties in children. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 93–112). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Stuebing, K. K., Fletcher, J. M., LeDoux, J. M., Lyon, G. R., Shaywitz, S. E., & Shaywitz, B. A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal*, *39*, 469–518.
- Torgesen, J. K. (2002). Empirical and theoretical support for direct diagnosis of learning disabilities by assessment of intrinsic processing weaknesses. In R. Bradley, L. Danielson, & D. Halahan (Eds.), *Identification of learning disabilities: Research to practice* (pp. 565–650). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Torgesen, J., Wagner, R., & Raschotte, C. (1999). *Test of Word Reading Efficiency*. Austin, TX: PRO-ED.
- U. S. Office of Education. (1977). Assistance to states for education for handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register*, *42*, G1082–G1085.
- Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction. *Learning Disabilities Research and Practice*, *18*, 137–146.
- Vaughn, S. R., Linan-Thompson, S., & Hickman-Davis, P. (2003). Response to treatment as a means of identifying students with reading/learning disabilities. *Exceptional Children*, *69*, 391–409.
- Vellutino, F. R. (1979). *Dyslexia: Theory and research*. Cambridge, MA: MIT Press.
- Vellutino, F. R., Fletcher, J. M., Scanlon, D. M., & Snowling, M. J. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*, 2–40.
- Vellutino, F. R., Scanlon, D. M., & Jaccard, J. (2003). Toward distinguishing between cognitive and experiential deficits as primary sources of difficulty in learning to read: A two year follow-up of difficult remediate and readily remediate poor readers. In B. R. Foorman (Ed.), *Preventing and remediating reading difficulties: Bringing science to scale* (pp. 73–120). Baltimore: York.
- Wadsworth, S. J., Olson, R. K., Pennington, B. F., & DeFries, J. C. (2000). Differential genetic etiology of reading disability as a function of IQ. *Journal of Learning Disabilities*, *33*, 192–199.
- Wechsler, D. (2001). *Wechsler Individual Achievement Test* (2nd ed.). San Antonio, TX: Psychological Corporation.
- Wiederholt, J. L., & Bryant, B. R. (2001). *Gray Oral Reading Test* (4th ed.). Austin, TX: PRO-ED.
- Wilkinson, G. (1993). *Wide Range Achievement Test-3*. Wilmington, DE: Wide Range.
- Williams, K. T., Cassidy, J., & Samuels, S. J. (2001). *Group reading assessment and diagnostic education*. Circle Pines, MN: American Guidance Services.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Received March 4, 2004

Accepted November 10, 2004

Copyright of Journal of Clinical Child & Adolescent Psychology is the property of Lawrence Erlbaum Associates and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

THE RELATIONSHIPS AMONG STERNBERG'S TRIARCHIC ABILITIES, GARDNER'S MULTIPLE INTELLIGENCES, AND ACADEMIC ACHIEVEMENT

BIRSEN EKINCI
Marmara University

In this study I investigated the relationships among Sternberg's Triarchic Abilities (STA), Gardner's multiple intelligences, and the academic achievement of children attending primary schools in Istanbul, Turkey. Participants were 172 children (93 boys and 81 girls) aged between 11 and 12 years. STA Test (STAT) total scores were significantly and positively related to linguistic, logical-mathematical, and intrapersonal test scores. Analytical ability scores were significantly positively related to only logical-mathematical test scores, practical ability scores were only related to intrapersonal test scores, and the STAT subsections were significantly related to each other. After removing the effect of multiple intelligences, the partial correlations between mathematics, social science, and foreign language course grades and creative, practical, analytical, and total STAT scores, were found to be significant for creative scores and total STAT scores, but nonsignificant for practical scores and analytical STAT scores.

Keywords: Sternberg's Triarchic Abilities Test, multiple intelligences, academic achievement, children, intelligence.

Since 1980 there has been increasing interest in the role of intelligence in learning and its impact on student achievement. Similarly to education theorists, many researchers on intelligence have been conducting studies to apply theories about intelligence, to education in general and, in particular, to the instructional context of the classroom (Castejón, Gilar, & Perez, 2008). The main difference between contemporary and older approaches to the role of intelligence is that,

Birsen Ekinci, Atatürk Education Faculty, Marmara University.

This study was supported by the Marmara University, Scientific Research Projects Center, research number EGT-D-110913-0387.

Correspondence concerning this article should be addressed to: Birsen Ekinci, Atatürk Education Faculty, Department of Primary Education, Marmara University, Göztepe Campus, 34722 Kadıköy, Istanbul, Turkey. Email: birsenp@marmara.edu.tr

in earlier conceptualizations, intelligence was described as involving one factor of general mental ability that encompasses the common variance among all the contributing factors. The existence of this general intelligence factor was originally hypothesized by Spearman in 1927 and labeled as "g" (see Jensen, 1998). It was hypothesized that this g factor exists over and above the various abilities that make up intelligence, including verbal, spatial visualization, numerical reasoning, mechanical reasoning, and memory (Carroll, 1993). However, according to contemporary theories, intelligence must be regarded as existing in various forms and the levels of intelligence can be improved through education. The most widely accepted comparative theories of intelligences in recent literature are Gardner's (1993) multiple intelligences theory and Sternberg's (1985) triarchic theory of intelligence. Researchers have reported significant differences between student outcomes for classroom instruction conducted following the principles of multiple intelligences, and student outcomes under traditionally designed courses of instruction in science (Özdermir, Güneysu, & Tekkaya, 2006), reading (Al-Balhan, 2006), and mathematics (Douglas, Burton, & Reese-Durham, 2008).

Gardner (1993) developed a theory of multiple intelligences that comprises seven distinct areas of skills that each person possesses to different degrees. Linguistic intelligence (LI) is the capacity to use words effectively, either orally or in writing. Logical-mathematical intelligence (LMI) is the capacity to use numbers effectively and to reason well. Spatial intelligence (SI) is the ability to perceive the visual-spatial world accurately and to interpret these perceptions. Bodily-kinesthetic intelligence (KI) involves expertise in using one's body to express ideas and feelings. Musical intelligence (MI) is the capacity to perceive, discriminate, and express musical forms. Interpersonal intelligence (INPI) is the ability to perceive, and make distinctions in, the moods, intentions, motivations, and feelings of other people. Intrapersonal intelligence (INTI) is self-knowledge and the ability to act adaptively on the basis of that knowledge. Naturalist intelligence (NI) is expertise in the recognition and classification of the numerous species – the flora and fauna – of a person's environment (Armstrong, 2009).

Researchers have addressed the relationship between multiple intelligences and metrics of different abilities, and of various psychological constructs. Reid, Romanoff, Algozzine, and Udall (2000) showed that SI, LI, and LMI were related to scores in a test to measure the nonverbal abilities of pattern completion, reasoning by analogy, serial reasoning, and spatial visualization, among a group of handicapped and nonhandicapped children aged between 5 and 17 years. Furthermore, the effects of multiple intelligences-based teaching strategies on students' academic achievement have been studied extensively (Al-Balhan, 2006; Douglas et al., 2008; Greenhawk, 1997; Mettetal, Jordan, & Harper, 1997; Özdermir et al., 2006). In addition, some researchers have investigated the relationship between multiple intelligences and academic achievement (McMahon, Rose, & Parks, 2004; Snyder, 1999). McMahon and colleagues

found that, compared with other students, fourth-grade students with higher scores on LMI were more likely to demonstrate reading comprehension scores at, or above, grade level. In a similar study, Snyder reported a positive correlation between high school students' grade point averages and KI. In the same study results showed that there was a positive correlation between the total score for the Metropolitan Achievement Test-Reading developed by the Psychological Corporation of San Antonio, Texas, USA and the categories of LMI and LI.

Sternberg developed the second well-known intelligence theory. According to Sternberg (1999a, 1999b), individuals show their intelligence when they apply the information-processing components of intelligence to cope with relatively novel tasks and situations. Within this approach to intelligence, Sternberg (1985) proposed the triarchic theory of intelligence, according to which there are three different, but interrelated, aspects of intellect: (a) analytic intelligence, (b) creative intelligence, and (c) practical intelligence. Individuals highly skilled in analytical intelligence are adept at analytical thinking, which involves applying the components of thinking to abstract, and often academic, problems. Individuals who have a high degree of creative intelligence are skilled at discovering, creating, and inventing ideas and products. People who have a high level of practical intelligence are good at using, implementing, and applying ideas and products. Sternberg (1997) developed an instrument, the Sternberg Triarchic Abilities Test (STAT), to evaluate triarchically based intelligence. In this instrument each aspect of intelligence is tested through three modes of presentation of problems: verbal, quantitative, and figural. A number of previous researchers have established the construct validity of the STAT (Sternberg, Castejón, Prieto, Hautamäki, & Grigorenko, 2001; Sternberg, Ferrari, Clinkenbeard, & Grigorenko, 1996). Although Sternberg did not intend the STAT to be a measure of general intelligence, as assessed by conventional intelligence tests, in related literature (Brody, 2003) there are contradictory results and opinions on this issue. Sternberg (2000a, 2000b) has claimed that the STAT is independent of measures of general intelligence and a more accurate predictor of academic achievement. However, Gottfredson (2002) pointed out that the data obtained to support this claim are sparse and suggested that the data collected by Sternberg et al. (1996) support the conclusion that the STAT is related to other measures of intelligence and may, in fact, be a measure of general intelligence. The triarchic abilities are related to different intelligence tests scores (e.g., Concept Mastery Test, Watson Glaser Critical Thinking Appraisal, Cattle Culture-Fait Test of g; Sternberg et al., 1996). However, Brody (2003) suggested that although these correlations are substantial, it is likely that they underestimate general intelligence because they were obtained from a sample of high school students who were predominately categorized as gifted, as determined by IQ scores, and these students were, therefore, likely to record a restricted range of scores on the tests.

In the present study I hypothesized that both multiple intelligences total scores and STAT total scores would be predictors of academic achievement. Specifically, I hypothesized that the LI and LMI, and the analytical STAT, would be predictors of student success in the subject areas of mathematics, science, social science, and foreign- language learning.

Method

Participants

Participants were 174 randomly selected fifth- and sixth-grade students (81 girls and 93 boys) attending primary school in Istanbul, Turkey. Students' ages ranged from 11 to 12 years old.

Instruments

The students completed the Turkish version of Gardner's Multiple Intelligences Inventory (MII; Saban, 2002) to assess participants' preferred intelligence within one of the eight categories: LI, LMI, SI, MI, KI, INPI, INTI, and NI. The possible score for the MII ranges from 0 to 80. The individual category in which a student has the highest score is considered to be the type of intelligence in which that student is most skilled. The overall Cronbach's alpha reliability coefficient in this study was .96, denoting high reliability; .89 for LI; .83 for LMI; .89 for SI; .88 for MI; .78 for KI; .85 for INPI; .85 for INTI; and .84 for NI.

The second instrument that I used in this study was Sternberg's Triarchic Abilities Test (STAT). The test comprises 81 items divided across three subsections designed to measure analytical, creative, and practical abilities. I translated this test into Turkish using the back-translation technique. In order to ensure that the back-translation retained the meaning of the original form, I conducted validity and reliability checks. The Turkish and the English versions of the test were given to 80 bilingual Turkish- and English-speaking students to complete within two weeks. Analyses of scores for the Turkish and English versions of test completed by these students yielded high correlation values (.85 for analytical, .79 for practical, and .81 for creative subsections). The overall alpha reliability coefficient of this test was .89, and for the subsections it was .80 for analytical, .77 for practical, and .78 for creative.

Procedure

The students completed the instruments during class time and in their classrooms. There was no time limit for completion. Each test session lasted approximately 60 minutes. The parents of the participating children gave permission for the researcher to access the students' grade point average for mathematics, science, social science, and foreign language courses at the end of the year during which the study was conducted. Each participant received a pen and pencil as a thank-you gift for his/her participation in this study.

Data Analysis

The data were analyzed using SPSS version 15 to conduct correlation analysis and multiple regression analysis.

Results

As shown in Table 1, the children's STAT total scores ($M = 35.34$, $SD = 9.09$) were significantly and positively related to LI ($M = 28.98$; $SD = 7.59$), LMI ($M = 30.12$, $SD = 6.87$), and INTI ($M = 29.10$, $SD = 7.15$) scores ($p < .01$). Analytical subsection STAT scores ($M = 13.76$, $SD = 3.96$) were significantly related to LM intelligence scores ($p < .01$). STAT practical subsection scores ($M = 10.37$, $SD = 3.06$) were significantly correlated only with INTI scores ($p < .01$).

Table 1. Relationships Among STAT Total Scores, Analytical, Practical, and Creative Ability Scores, and Multiple Intelligences Scores

	LI	LMI	SI	MI	KI	INPI	INTI	NI
Analytical	.303	.413**	-.057	.093	.036	.021	.281	-.102
Practical	.274	.268	.003	.113	.041	.095	.434**	-.109
Creative	.291	.540**	-.062	.103	.004	-.049	.361*	-.098
Total	.351*	.506**	-.051	.123	.031	.019	.425**	-.124

Note. ** $p < .01$, * $p < .05$. LI = linguistic intelligence, LMI = logical-mathematical intelligence, SI = spatial intelligence, MI = musical intelligence, KI = bodily-kinesthetic intelligence, INPI = interpersonal intelligence, INTI = intrapersonal intelligence, NI = naturalist intelligence.

Mathematics course grades ($M = 3.78$; $SD = 1.20$) were significantly related to the STAT total ($p < .001$) and to the STAT analytical ($p < .001$), practical ($p < .01$), and creative ($p < .01$) subsections. Similarly, social science ($M = 3.78$, $SD = 1.10$) and science course grades ($M = 3.51$, $SD = 1.40$) were significantly related to the STAT total ($p < .01$) and to the STAT analytical ($p < .01$) and creative ($p < .01$) subsections. However, foreign language course grades ($M = 3.57$, $SD = 1.16$) were significantly related to all of the subsection scores of the STAT ($p < .001$; see Table 2).

Table 2. Relationships Among STAT Total Scores, Analytical, Practical, and Creative Subsection Scores, and Academic Success

	Mathematics	Science	Social science	Foreign language
Analytical	.536*	.395**	.304**	.454*
Practical	.461**	.264	.269	.451*
Creative	.491*	.378**	.307**	.442*
Total	.588*	.415**	.347**	.527*

Note. * $p < .001$, ** $p < .01$.

Mathematics grades of the participants were significantly related to LI ($p < .01$), LMI ($p < .01$), INPI ($p < .05$), and INTI ($p < .01$) scores. Similarly, students' course grades for science were significantly related to LI ($p < .05$), LMI ($p < .01$), and INTI ($p < .05$) scores; students' social science course grades were significantly related to LI ($p < .05$), LMI ($p < .01$), and INTI ($p < .05$) scores; and students' course grades for foreign languages were significantly related to LI ($p < .01$), LMI ($p < .01$) and INTI ($p < .01$) scores (see Table 3).

Table 3. *Relationships Between Multiple Intelligences Scores and Academic Success*

	LI	LMI	SI	MI	KI	INPI	INTI	NI
Mathematics	.458**	.695**	.080	.174	.285	.356*	.522**	.140
Science	.340*	.575**	.007	.070	.239	.312	.379*	.085
Social science	.359*	.598**	.125	.118	.217	.319	.356*	.139
Foreign language	.484**	.718**	.211	.201	.260	.316	.495**	.227

Note. ** $p < .01$, * $p < .05$. LI = linguistic intelligence, LMI = logical-mathematical intelligence, SI = spatial intelligence, MI = musical intelligence, KI = bodily-kinesthetic intelligence, INPI = interpersonal intelligence, INTI = intrapersonal intelligence, NI = naturalist intelligence.

Multiple regression analyses were conducted in which the variance caused by the MII was removed, and partial correlations were computed between course grades and children's STAT total and subsection scores. Separate analyses were conducted for each subject area using first the STAT subsections and then using just the STAT total scores. Analyses regarding mathematics course grades yielded significant partial correlations for the creative subsection score ($P_r = .44$, $p < .01$) and for the total STAT score ($P_r = .62$, $p < .01$), but the partial correlations were not significant for the analytical ($P_r = .14$) and practical ($P_r = .05$) STAT scores. Similarly, the regression analyses predicting students' science course grades yielded significant partial correlations for STAT total scores ($P_r = .53$, $p < .01$) and for the creative subsection score ($P_r = .42$, $p < .01$), but not for the analytical ($P_r = .14$) or practical ($P_r = .06$) STAT scores. Additionally, when I performed the same analyses of social science course grades these yielded significant partial correlations with STAT total scores ($P_r = .54$, $p < .01$) and creative subsection scores ($P_r = .34$, $p < .05$) but not with analytical ($P_r = .19$) or creative ($P_r = .04$) STAT scores. Finally, analyses yielded the same pattern for foreign language course grades and STAT total and subsection scores. Regression analyses yielded significant partial correlations for practical subsection scores ($P_r = .41$, $p < .02$) and for total STAT scores ($P_r = .61$, $p < .01$). Thus, the total STAT scores and creative subsection scores significantly predicted academic achievement in mathematics, science, social science, and foreign language courses, independent of multiple intelligences scores; however, the analytical and practical subsection scores did not. Correspondingly, the partial correlations

between course grade (for mathematics, social science, science, and foreign language) and the MII subsection scores, with the variation caused by the STAT removed, were significant only for LMI ($P_r = .70$, $p < .01$) scores. This finding indicates that, independent of the STAT, only LMI scores predicted achievement in any subject area.

Discussion

The results in this study showed that STAT total scores were significantly related to LI, LMI, and INTI scores. Analytical subsection STAT scores were significantly related to LMI scores. Practical STAT subsection scores were significantly correlated only with INTI scores. These results are based on the partial correlations between multiple intelligences and STAT scores. However, I limited the scope of this study to the students' own preferences in regard to their multiple intelligences. In future studies students' intelligence types should be assessed together with the performances of students on related intelligences for different age groups and different subject areas. In the present study mathematics course grades were significantly related to STAT total scores and to scores for the STAT analytical, practical, and creative abilities subsections. Similarly, science, social science, and foreign language course grades were significantly related to the LI, LM, and INTI scores of the participants.

Results of multiple regression analyses indicated that total STAT scores and creative ability scores significantly predicted academic achievement in mathematics, social science, science, and foreign language learning, independent of multiple intelligences scores; however, the analytical and practical ability scores did not. These results are consistent with those reported by Sternberg et al. (2001), who found that total STAT and creative ability scores significantly predicted academic achievement. However, contrary to the findings reported by Sternberg et al., in my study the analytical and practical ability scores did not relate significantly to academic achievement. On the other hand, Koke and Vernon (2003) reported that total STAT scores and only practical ability scores predicted psychology course midterm grades of university students. All these results might indicate that there may be cultural differences within the dominant cognitive abilities represented in the national education systems of various countries.

My results in this study also revealed that the partial correlation between course grades for all of the subject areas and each of the MII subsection scores, with the variation caused by the STAT removed, was significant for only the LMI score. This indicates that, independent of the STAT, only LMI scores predicted achievement in any subject area. It should also be noted that in this study the students' multiple intelligences scores were based on their own preferences for

the items representing various kinds of intelligences. In other words, the multiple intelligences scores did not indicate the actual performance of the children in each type of intelligence. I believe that it would be of value for future researchers to test how well the STAT would predict academic achievement for scores on a test in which students' multiple intelligences scores were each taken into account separately. The relationship between other tests and STAT scores could also be examined with more heterogeneous sample groups.

References

- Al-Balhan, E. M. (2006). Multiple intelligence styles in relation to improved academic performance in Kuwaiti middle school reading. *Digest of Middle East Studies*, 15, 18-34. <http://doi.org/cd8zdh>
- Armstrong, T. (2009). *Multiple intelligences in the classroom*. Alexandria, VA: ASCD.
- Brody, N. (2003). Construct validation of the Sternberg Triarchic Abilities Test: Comment and reanalysis. *Intelligence*, 31, 319-329. <http://doi.org/ffgmzb>
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Castejón, J. L., Gilar, R., & Perez, N. (2008). From "g factor" to multiple intelligences: Theoretical foundations and implications for classroom practice. In E. P. Velliotis (Ed.), *Classroom culture and dynamics* (pp. 101-127). New York: Nova Science.
- Douglas, O., Burton, K. S., & Reese-Durham, N. R. (2008). The effects of the multiple intelligence teaching strategy on the academic achievement of eighth grade math students. *Journal of Instructional Psychology*, 35, 182-187.
- Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences*. New York: Basic.
- Gottfredson, L. S. (2002). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general intelligence factor: How general is it?* (pp. 331-380). Mahwah, NJ: Erlbaum.
- Greenhawk, J. (1997). Multiple intelligences meet standards. *Educational Leadership*, 55, 62-64.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger/Greenwood.
- Koke, L. C., & Vernon, P. A. (2003). The Sternberg Triarchic Abilities Test (STAT) as a measure of academic achievement and general intelligence. *Personality and Individual Differences*, 35, 1803-1807. <http://doi.org/fmfpqb>
- McMahon, S. D., Rose, D., & Parks, M. (2004). Multiple intelligences and reading achievement: An examination of the Teele Inventory of Multiple Intelligences. *The Journal of Experimental Education*, 73, 41-52. <http://doi.org/bwptfs>
- Mettetal, G., Jordan, C., & Harper, S. (1997). Attitude toward a multiple intelligences curriculum. *Journal of Educational Research*, 91, 115-122. <http://doi.org/dmsgds>
- Özdermir, P., Güneysu, S., & Tekkaya, C. (2006). Enhancing learning through multiple intelligences. *Journal of Biological Education*, 40, 74-78. <http://doi.org/fn2x6h>
- Reid, C., Romanoff, B., Algozzine, B., & Udall, A. (2000). An evaluation of alternative screening procedures. *Journal for the Education of the Gifted*, 23, 378-396.
- Saban, A. (2002). *Öğrenme ve öğretim* [Learning and teaching: New theories and approaches]. Ankara: Nobel.
- Sternberg, R. J. (1985). Implicit theories of intelligence, creativity, and wisdom. *Journal of Personality and Social Psychology*, 49, 607-627. <http://doi.org/cstvmp>
- Sternberg, R. J. (1993). *The Sternberg Triarchic Abilities Test*. Unpublished manuscript.

- Sternberg, R. J. (1997). The concept of intelligence and its role in lifelong learning and success. *American Psychologist*, 52, 1030-1037. <http://doi.org/dzxj2p>
- Sternberg, R. J. (1999a). Intelligence as developing expertise. *Contemporary Educational Psychology*, 24, 359-375. <http://doi.org/dzvjsj>
- Sternberg, R. J. (1999b). The theory of successful intelligence. *Review of General Psychology*, 3, 292-316. <http://doi.org/cqrkxh>
- Sternberg, R. J. (2000). The concept of intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 3-13). New York: Cambridge University Press.
- Sternberg, R. J. (2000). *Practical intelligence in everyday life*. New York: Cambridge University Press.
- Sternberg, R. J., Castejón, J. L., Prieto, M. D., Hautamäki, J., & Grigorenko, E. L. (2001). Confirmatory factor analysis of the Sternberg Triarchic Abilities Test in three international samples: An empirical test of the triarchic theory of intelligence. *European Journal of Psychological Assessment*, 17, 1-16. <http://doi.org/cn7tjp>
- Sternberg, R. J., Ferrari, M., Clinkenbeard, P. R., & Grigorenko, E. L. (1996). Identification, instruction, and assessment of gifted children: A construct validation of a triarchic model. *Gifted Child Quarterly*, 40, 129-137. <http://doi.org/d3rf9w>
- Snyder, R. F. (1999). The relationship between learning styles/multiple intelligences and academic achievement of high school students. *High School Journal*, 83, 11-20.

Copyright of Social Behavior & Personality: an international journal is the property of Society for Personality Research and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Resource: Gregory, B.J. (2014). Psychological testing: History, Principles, and applications (7th ed.). Boston, MA: Pearson

CHAPTER 5

Theories and Individual Tests of Intelligence and Achievement

TOPIC 5A Theories of Intelligence and Factor Analysis

5.1 Definitions of Intelligence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec1#ch05lev1sec1>)

Case Exhibit 5.1 Learning and Adaptation as Core Functions of Intelligence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec1#ch05box2>)

5.2 A Primer of Factor Analysis

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05lev1sec2>)

5.3 Galton and Sensory Keeness

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec3#ch05lev1sec3>)

5.4 Spearman and the *g* Factor

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec4#ch05lev1sec4>)

5.5 Thurstone and the Primary Mental Abilities

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec5#ch05lev1sec5>)

5.6 Cattell-Horn-Carroll (CHC) Theory

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec6#ch05lev1sec6>)

5.7 Guilford and the Structure-of-Intellect Model

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec7#ch05lev1sec7>)

5.8 Planning, Attention, Simultaneous, and Successive (Pass) Theory

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec8#ch05lev1sec8>)

5.9 Information Processing Theories of Intelligence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec9#ch05lev1sec9>)

5.10 Gardner and the Theory of Multiple Intelligences

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec10#ch05lev1sec10>)

5.11 Sternberg and the Triarchic Theory of Successful Intelligence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec11#ch05lev1sec11>)

This chapter opens an extended discussion of intelligence and achievement testing, a topic so important and immense that we devote the next two chapters to it as well. In order to understand contemporary cognitive testing, the reader will need to assimilate certain definitions, theories, and mainstream assessment practices. The goal of **Topic 5A** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05#ch05box1>), Theories of Intelligence and Factor Analysis, is to investigate the various meanings given to the term intelligence and to discuss how definitions and theories have influenced the structure and content of intelligence tests. An important justification for this topic is that an understanding of theories of intelligence is crucial for establishing the construct validity of IQ measures. Furthermore, because the statistical tools of factor analysis are so vital to many theories of intelligence, we provide a primer of the topic here. In **Topic 5B** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec11#ch05box3>), Individual Tests of Intelligence and Achievement, we summarize a number of noteworthy approaches to individual assessment and focus on one important application, the evaluation of learning disabilities. We begin with a foundational question: How is intelligence defined?

Intelligence is one of the most highly researched topics in psychology. Thousands of research articles are published each year on the nature and measurement of intelligence. New journals such as *Intelligence* and *The Journal of Psychoeducational Assessment* have flourished in response to the scholarly interest in this topic. Despite this burgeoning research literature, the definition of intelligence remains elusive, wrapped in controversy and mystery. In fact, the discussion that follows will illustrate a major paradox of modern testing: Psychometricians are better at measuring intelligence than conceptualizing it!

Even though defining intelligence has proved to be a frustrating endeavor, there is much to be gained by reviewing historical and contemporary efforts to clarify its meaning. After all, intelligence tests did not materialize out of thin air. Most tests are grounded in a specific theory of intelligence and most test developers offer a definition of the construct as a starting point for their endeavors. For these reasons, we can better understand and evaluate the multifaceted character of contemporary tests if we first review prominent definitions and theories of intelligence.

5.1 DEFINITIONS OF INTELLIGENCE

Before we discuss definitions of intelligence, we need to clarify the nature of definition itself. Sternberg (1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)) makes a distinction between operational and “real” definitions that is important in this context. An **operational definition** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss229>) defines a concept in terms of the way it is measured. Boring (1923 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib179>)) carried this viewpoint to its extreme when he defined intelligence as “what the tests test.” Believe it or not, this was a serious proposal, designed largely to short-circuit rampant and divisive disagreements about the definition of intelligence.

Operational definitions of intelligence suffer from two dangerous shortcomings (Sternberg, 1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)). First, they are circular. Intelligence tests were invented to measure intelligence, not to define it. The test designers never intended for their instruments to define intelligence. Second, operational definitions block further progress in understanding the nature of intelligence, because they foreclose discussion on the adequacy of theories of intelligence.

This second problem—the potentially stultifying effects of relying on operational definitions of intelligence—casts doubt on the common practice of affirming the concurrent validity of new tests by correlating them with old tests. If established tests serve as the principal criterion against which new tests are assessed, then the new tests will be viewed as valid only to the extent that they correlate with the old ones. Such a conservative practice drastically curtails innovation. The operational definition of intelligence does not allow for the possibility that new tests or conceptions of intelligence may be superior to the existing ones.

We must conclude, then, that operational definitions of intelligence leave much to be desired. In contrast, a **real definition** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss272>) is one that seeks to tell us the true nature of the thing being defined (Robinson, 1950; Sternberg, 1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)). Perhaps the most common way—but by no means the only way—of producing real definitions of intelligence is to ask experts in the field to define it.

Expert Definitions of Intelligence

Intelligence has been given many real definitions by prominent researchers in the field. In the following, we list several examples, paraphrased slightly for editorial consistency. The reader will note that many of these definitions appeared in an early but still influential symposium, “Intelligence and Its Measurement,” published in the *Journal of Educational Psychology* (Thorndike, 1921 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1640>)). Other definitions stem from a modern update of this early symposium, *What Is Intelligence?*, edited by Sternberg and Detterman (1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)). Intelligence has been defined as the following:

- Spearman (1904 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1526>), 1923 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1527>)): a general ability that involves mainly the education of relations and correlates.

- Binet and Simon (**1905** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib160>)): the ability to judge well, to understand well, to reason well.
- Terman (**1916** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1629>)): the capacity to form concepts and to grasp their significance.
- Pintner (**1921** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1303>)): the ability of the individual to adapt adequately to relatively new situations in life.
- Thorndike (**1921** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1640>)): the power of good responses from the point of view of truth or fact.
- Thurstone (**1921** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1643>)): the capacity to inhibit instinctive adjustments, flexibly imagine different responses, and realize modified instinctive adjustments into overt behavior.
- Wechsler (**1939** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1725>)): The aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with the environment.
- Humphreys (**1971** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib796>)): the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature that are available at any one period of time.
- Piaget (**1972** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1292>)): a generic term to indicate the superior forms of organization or equilibrium of cognitive structuring used for adaptation to the physical and social environment.
- Sternberg (**1985a** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1560>), **1986** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)): the mental capacity to automatize information processing and to emit contextually appropriate behavior in response to novelty; intelligence also includes metacomponents, performance components, and knowledge-acquisition components (discussed later).
- Eysenck (**1986** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib475>)): error-free transmission of information through the cortex.
- Gardner (**1986** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib564>)): the ability or skill to solve problems or to fashion products that are valued within one or more cultural settings.
- Ceci (**1994** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib289>)): multiple innate abilities that serve as a range of possibilities; these abilities develop (or fail to develop, or develop and later atrophy) depending upon motivation and exposure to relevant educational experiences.
- Sattler (**2001** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1438>)): intelligent behavior reflects the survival skills of the species, beyond those associated with basic physiological processes.

The preceding list of definitions is representative although definitely not exhaustive. For one thing, the list is exclusively Western and omits several cross-cultural conceptions of intelligence. Eastern conceptions of intelligence, for example, emphasize benevolence, humility, freedom from conventional standards of judgment, and doing what is right as essential to intelligence. Many African conceptions of intelligence place heavy emphasis on social aspects of intelligence such as maintaining harmonious and stable intergroup relations (**Sternberg & Kaufman, 1998** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1569>)). The reader can consult Bracken and Fagan (1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib194>)), Sternberg (1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1564>)), and Sternberg and Detterman (1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)) for additional ideas. Certainly, this sampling of views is sufficient to demonstrate that there appear to be as many definitions of intelligence as there are experts willing to define it!

In spite of this diversity of viewpoints, two themes recur again and again in expert definitions of intelligence. Broadly speaking, the experts tend to agree that intelligence is (1) the capacity to learn from experience and (2) the capacity to adapt to one's environment. That learning and adaptation are both crucial to intelligence stands out with poignancy in certain cases of mental disability in which persons fail to possess one or the other capacity in sufficient degree (**Case Exhibit 5.1** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec1#ch05box2>)).

CASE EXHIBIT 5.1

Learning and Adaptation as Core Functions of Intelligence

Persons with mental disability often demonstrate the importance of experiential learning and environmental adaptation as key ingredients of intelligence. Consider the case history of a 61-year-old newspaper vendor with moderate mental retardation well known to local mental health specialists. He was an interesting if not eccentric gentleman who stored canned goods in his freezer and cursed at welfare workers who stopped by to see how he was doing. In spite of his need for financial support from a state agency, he was fiercely independent and managed his own household with minimal supervision from case workers. Thus, in some respects he maintained a tenuous adaptation to his environment. To earn much-needed extra income, he sold a local 25-cent newspaper from a streetside newsstand. He recognized that a quarter was proper payment and had learned to give three quarters in change for a dollar bill. He refused all other forms of payment, an arrangement that his customers could accept. But one day the price of the newspaper was increased to 35 cents, and the newspaper vendor was forced to deal with nickels and dimes as well as quarters and dollar bills. The amount of learning required by this slight shift in environmental demands exceeded his intellectual abilities, and, sadly, he was soon out of business. His failed efforts highlight the essential ingredients of intelligence: learning from experience and adaptation to the environment.

How well do intelligence tests capture the experts' view that intelligence consists of learning from experience and adaptation to the environment? The reader should keep this question in mind as we proceed to review major intelligence tests in the topics that follow. Certainly, there is cause for concern: Very few contemporary intelligence tests appear to require the examinee to learn something new or to adapt to a new situation as part and parcel of the examination process. At best, prominent modern tests provide indirect measures of the capacities to learn and adapt. How well they capture these dimensions is an empirical question that must be demonstrated through validation research.

Layperson and Expert Conceptions of Intelligence

Another approach to understanding a construct is to study its popular meaning. This method is more scientific than it may appear. Words have a common meaning to the extent that they help provide an effective portrayal of everyday transactions. If laypersons can agree on its meaning, a construct such as intelligence is in some sense “real” and, therefore, potentially useful. Thus, asking persons on the street, “What does intelligence mean to you?” has much to recommend it.

Sternberg, Conway, Ketron, and Bernstein (1981

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1559>) conducted a series of studies to investigate conceptions of intelligence held by American adults. In the first study, people in a train station, entering a supermarket, and studying in a college library were asked to list behaviors characteristic of different kinds of intelligence. In a second study—the only one discussed here—both laypersons and experts (mainly academic psychologists) rated the importance of these behaviors to their concept of an “ideally intelligent” person.

The behaviors central to expert and lay conceptions of intelligence turned out to be very similar, although not identical. In order of importance, experts saw verbal intelligence, problem-solving ability, and practical intelligence as crucial to intelligence. Laypersons regarded practical problemsolving ability, verbal ability, and social competence to be the key ingredients in intelligence. Of course, opinions were not unanimous; these conceptions represent the consensus view of each group. In their conception of intelligence, experts place more emphasis on verbal ability than problem solving, whereas laypersons reverse these priorities. Nonetheless, experts and laypersons alike consider verbal ability and problem solving to be essential aspects of intelligence. As the reader will see, most intelligence tests also accent these two competencies. Prototypical examples would be vocabulary (verbal ability) and block design (problem solving) from the Wechsler scales, discussed later. We see then that everyday conceptions of intelligence are, in part, mirrored quite faithfully by the content of modern intelligence tests.

Some disagreement between experts and laypersons is also evident. Experts consider practical intelligence (sizing up situations, determining how to achieve goals, awareness and interest in the world) an essential constituent of intelligence, whereas laypersons identify social competence (accepting others for what they are, admitting mistakes, punctuality, and interest in the world) as a third component. Yet, these two nominations do share one property in common: Contemporary tests generally make no attempt to measure either practical intelligence or social competence. Partly, this reflects the psychometric difficulties encountered in devising test items relevant to these content areas. However, the more influential reason intelligence tests do not measure practical intelligence or social competence is inertia: Test developers have blindly accepted historically incomplete conceptions of intelligence. Until recently, the development of intelligence testing has been a conservative affair, little changed since the days of Binet and the Army Alpha and Beta tests for World War I recruits. There are some signs that testing practices may soon evolve, however, with the development of innovative instruments. For example, Sternberg and colleagues have proposed innovative tests based on his model of intelligence. Another interesting instrument based on a new model of intelligence is the Everyday Problem Solving Inventory (Cornelius & Caspi, 1987

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib351>). In this test, examinees must indicate their typical response to everyday problems such as failing to bring money, checkbook, or credit card when taking a friend to lunch.

Many theorists in the field of intelligence have relied on factor analysis for the derivation or validation of their theories. In fact, it is not an overstatement to say that perhaps the majority of the theories in this area have been impacted by the statistical tools of factor analysis, which provide ways to portion

intelligence into its subcomponents. One of the most compelling theories of intelligence, the Cattell-Horn-Carroll theory reviewed later, would not exist without factor analysis. Thus, before summarizing theories, we provide a brief review of this essential statistical tool.

5.2 A PRIMER OF FACTOR ANALYSIS

Broadly speaking, there are two forms of factor analysis: confirmatory and exploratory. In confirmatory factor analysis, the purpose is to confirm that test scores and variables fit a certain pattern predicted by a theory. For example, if the theory underlying a certain intelligence test prescribed that the subtests belong to three factors (e.g., verbal, performance, and attention factors), then a confirmatory factor analysis could be undertaken to evaluate the accuracy of this prediction. Confirmatory factor analysis is essential to the validation of many ability tests.

The central purpose of exploratory **factor analysis**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss113>) is to summarize the interrelationships among a large number of variables in a concise and accurate manner as an aid in conceptualization (Gorsuch, 1983). For instance, factor analysis may help a researcher discover that a battery of 20 tests represents only four underlying variables, called **factors**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss112>). The smaller set of derived factors can be used to represent the essential constructs that underlie the complete group of variables.

Perhaps a simple analogy will clarify the nature of factors and their relationship to the variables or tests from which they are derived. Consider the track-and-field decathlon, a mixture of 10 diverse events including sprints, hurdles, pole vault, shot put, and distance races, among others. In conceptualizing the capability of the individual decathlete, we do not think exclusively in terms of the participant's skill in specific events. Instead, we think in terms of more basic attributes such as speed, strength, coordination, and endurance, each of which is reflected to a different extent in the individual events. For example, the pole vault requires speed and coordination, while hurdle events demand coordination and endurance. These inferred attributes are analogous to the underlying factors of factor analysis. Just as the results from the 10 events of a decathlon may boil down to a small number of underlying factors (e.g., speed, strength, coordination, and endurance), so too may the results from a battery of 10 or 20 ability tests reflect the operation of a small number of basic cognitive attributes (e.g., verbal skill, visualization, calculation, and attention, to cite a hypothetical list). This example illustrates the goal of factor analysis: to help produce a parsimonious description of large, complex data sets.

We will illustrate the essential concepts of factor analysis by pursuing a classic example concerned with the number and kind of factors that best describe student abilities. Holzinger and Swineford (1939 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib779>)) gave 24 ability-related psychological tests to 145 junior high school students from Forest Park, Illinois. The factor analysis described later was based on methods outlined in Kinnear and Gray (1997 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib893>)).

It should be intuitively obvious to the reader that any large battery of ability tests will reflect a smaller number of basic, underlying abilities (factors). Consider the 24 tests depicted in **Table 5.1** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab1>). Surely some of these tests measure common underlying abilities. For example, we would expect Sentence Completion, Word Classification, and Word Meaning (variables 7, 8, and 9) to assess a factor of general language ability of some kind. In like manner, other groups of tests seem likely to measure common underlying abilities—but how many abilities or factors? And what is the nature of these underlying abilities? Factor analysis is the ideal tool for answering these questions. We follow the factor analysis of the Holzinger and Swineford (1939

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib779>) data from beginning to end.

TABLE 5.1 The 24 Ability Tests Used by Holzinger and Swineford (1939

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib779>)

-
- Visual Perception
 - Cubes
 - Paper Form Board
 - Flags
 - General Information
 - Paragraph Comprehension
 - Sentence Completion
 - Word Classification
 - Word Meaning
 - Add Digits
 - Code (Perceptual Speed)
 - Count Groups of Dots
 - Straight and Curved Capitals
 - Word Recognition
 - Number Recognition
 - Figure Recognition
 - Object-Number
 - Number-Figure
 - Figure-Word
 - Deduction
 - Numerical Puzzles
 - Problem Reasoning
 - Series Completion
 - Arithmetic Problems
-

The Correlation Matrix

The beginning point for every factor analysis is the **correlation matrix**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss74>), a complete table of intercorrelations among all the variables.¹

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fn01>) The correlations between the 24 ability variables discussed here can be found in **Table 5.2**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab2>). The reader will notice that variables 7, 8, and 9 do, indeed, intercorrelate quite strongly (correlations of .62, .69, and .53), as we suspected earlier. This pattern of intercorrelations is presumptive evidence that these variables measure something in common; that is, it appears that these tests reflect a common underlying factor. However, this kind of intuitive factor analysis based on a visual inspection of the correlation matrix is hopelessly limited; there are just too many intercorrelations for the viewer to discern the underlying patterns for all the variables. Here is where factor analysis can be helpful. Although we cannot elucidate the mechanics of the procedure, factor analysis relies on modern high-speed computers to search the correlation matrix according to objective statistical rules and determine the smallest number of factors needed to account for the observed pattern of intercorrelations. The analysis also produces the factor

matrix, a table showing the extent to which each test loads on (correlates with) each of the derived factors, as discussed in the following section.

The Factor Matrix and Factor Loadings

The **factor matrix** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss115>) consists of a table of correlations called factor loadings. The factor loadings (which can take on values from -1.00 to $+1.00$) indicate the weighting of each variable on each factor. For example, the factor matrix in **Table 5.3**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab3>) shows that five factors (labeled I, II, III, IV, and V) were derived from the analysis. Note that the first variable, Series Completion, has a strong positive loading of $.71$ on factor I, indicating that this test is a reasonably good index of factor I. Note also that Series Completion has a modest negative loading of $-.11$ on factor II, indicating that, to a slight extent, it measures the opposite of this factor; that is, high scores on Series Completion tend to signify low scores on factor II, and vice versa.

TABLE 5.2 The Correlation Matrix for 24 Ability Variables

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		
2	32																							
3	40	32																						
4	47	23	31																					
5	32	29	25	23																				
6	34	23	27	33	62																			
7	30	16	22	34	66	72																		
8	33	17	38	39	58	53	62																	
9	33	20	18	33	72	71	69	53																
10	12	06	08	10	31	20	25	29	17															
11	31	15	09	11	34	35	23	30	28	48														
12	31	15	14	16	22	10	18	27	11	59	43													
13	49	24	32	33	34	31	35	40	28	41	54	51												
14	13	10	18	07	28	29	24	25	26	17	35	13	20											
15	24	13	07	13	23	25	17	18	25	15	24	17	14	37										
16	41	27	26	32	19	29	18	30	24	12	31	12	28	41	33									
17	18	01	18	19	21	27	23	26	27	29	36	28	19	34	35	32								
18	37	26	21	25	26	17	16	25	21	32	35	35	32	21	33	34	45							
19	27	11	31	14	19	25	23	27	27	19	29	11	26	21	19	26	32	36						
20	37	29	30	34	40	44	45	43	45	17	20	25	24	30	27	39	26	30	17					
21	37	31	17	35	32	26	31	36	27	41	40	36	43	18	23	35	17	36	33	41				
22	41	23	25	38	44	39	40	36	48	16	30	19	28	24	25	28	27	32	34	46	37			
23	47	35	38	34	44	43	41	50	50	26	25	35	38	24	26	36	29	27	30	51	45	50		
24	28	21	20	25	42	43	44	39	42	53	41	41	36	30	17	26	33	41	37	37	45	38	43	

Note: Decimals omitted.

Source: Reprinted with permission from Holzinger, K., & Harman, H. (1941). *Factor analysis: A synthesis of factorial methods*. Chicago: University of Chicago Press. Copyright © 1941 The University of Chicago Press.

The factors may seem quite mysterious, but in reality they are conceptually quite simple. A factor is nothing more than a weighted linear sum of the variables; that is, each factor is a precise statistical combination of the tests used in the analysis. In a sense, a factor is produced by “adding in” carefully determined portions of some tests and perhaps “subtracting out” fractions of other tests. What makes the factors special is the elegant analytical methods used to derive them. Several different methods exist. These methods differ in subtle ways beyond the scope of this text; the reader can gather a sense of the differences by examining names of procedures: principal components factors, principal axis factors, method of unweighted least squares, maximum-likelihood method, image factoring, and alpha factoring (Tabachnick & Fidell, 1989

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1607>). Most of the methods yield highly similar results.

The factor loadings depicted in **Table 5.3**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab3>) are nothing more than correlation coefficients between variables and factors. These correlations can be interpreted as showing the weight or loading of each factor on each variable. For example, variable 9, the test of Word Meaning, has a very strong loading (.69) on factor I, modest negative loadings (-.45 and -.29) on factors II and III, and negligible loadings (.08 and .00) on factors IV and V.

TABLE 5.3 The Principal Axes Factor Analysis for 24 Variables

	<i>Factors</i>				
	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>
23. Series Completion	.71	-.11	.14	.11	.07
8. Word Classification	.70	-.24	-.15	-.11	-.13
5. General Information	.70	-.32	-.34	-.04	.08
9. Word Meaning	.69	-.45	-.29	.08	.00
6. Paragraph Comprehension	.69	-.42	-.26	.08	-.01
7. Sentence Completion	.68	-.42	-.36	-.05	-.05
24. Arithmetic Problems	.67	.20	-.23	-.04	-.11
20. Deduction	.64	-.19	.13	.06	.28
22. Problem Reasoning	.64	-.15	.11	.05	-.04
21. Numerical Puzzles	.62	.24	.10	-.21	.16
13. Straight and Curved Capitals	.62	.28	.02	-.36	-.07
1. Visual Perception	.62	-.01	.42	-.21	-.01
11. Code (Perceptual Speed)	.57	.44	-.20	.04	.01
18. Number-Figure	.55	.39	.20	.15	-.11
16. Figure Recognition	.53	.08	.40	.31	.19
4. Flags	.51	-.18	.32	-.23	-.02
17. Object-Number	.49	.27	-.03	.47	-.24
2. Cubes	.40	-.08	.39	-.23	.34

	<i>Factors</i>				
	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>
12. Count Groups of Dots	.48	.55	-.14	-.33	.11
10. Add Digits	.47	.55	-.45	-.19	.07
3. Paper Form Board	.44	-.19	.48	-.12	-.36
14. Word Recognition	.45	.09	-.03	.55	.16
15. Number Recognition	.42	.14	.10	.52	.31
19. Figure-Word	.47	.14	.13	.20	-.61

Geometric Representation of Factor Loadings

It is customary to represent the first two or three factors as reference axes in two- or three-dimensional space.² (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fm02>) Within this framework the factor loadings for each variable can be plotted for examination. In our example, five factors were discovered, too many for simple visualization. Nonetheless, we can illustrate the value of geometric representation by oversimplifying somewhat and depicting just the first two factors (**Figure 5.1** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fig1>)). In this graph, each of the 24 tests has been plotted against the two factors that correspond to axes I and II. The reader will notice that the factor loadings on the first factor (I) are uniformly positive, whereas the factor loadings on the second factor (II) consist of a mixture of positive and negative.

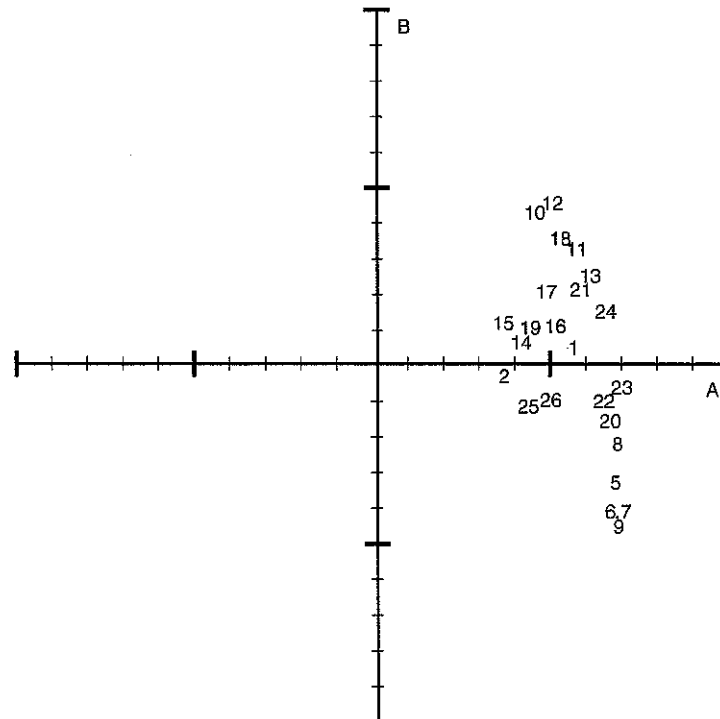


FIGURE 5.1 Geometric Representation of the First Two Factors from 24 Ability Tests

The Rotated Factor Matrix

An important point in this context is that the position of the reference axes is arbitrary. There is nothing to prevent the researcher from rotating the axes so that they produce a more sensible fit with the factor

loadings. For example, the reader will notice in **Figure 5.1**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fig1>) that tests 6, 7, and 9 (all language tests) cluster together. It would certainly clarify the interpretation of factor I if it were to be redirected near the center of this cluster (**Figure 5.2**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fig2>)). This manipulation would also bring factor II alongside interpretable tests 10, 11, and 12 (all number tests).

Although rotation can be conducted manually by visual inspection, it is more typical for researchers to rely on one or more objective statistical criteria to produce the final rotated factor matrix. Thurstone's (1947 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1648>)) criteria of positive manifold and simple structure are commonly applied. In a **rotation to positive manifold** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss281>), the computer program seeks to eliminate as many of the negative factor loadings as possible. Negative factor loadings make little sense in ability testing, because they imply that high scores on a factor are correlated with poor test performance. In a **rotation to simple structure** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss282>), the computer program seeks to simplify the factor loadings so that each test has significant loadings on as few factors as possible. The goal of both criteria is to produce a rotated factor matrix that is as straightforward and unambiguous as possible.

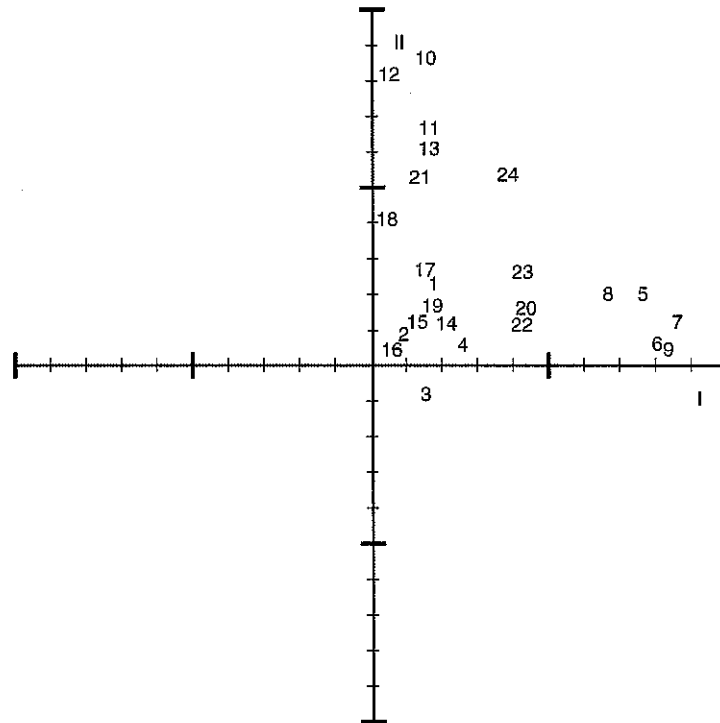


FIGURE 5.2 Geometric Representation of the First Two Rotated Factors from 24 Ability Tests

The rotated factor matrix for this problem is shown in **Table 5.4**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab4>). The particular method of rotation used here is called varimax rotation. Varimax should not be used if the theoretical expectation suggests that a general factor may occur. Should we expect a general factor in the analysis of ability tests? The answer is as much a matter of faith as of science. One researcher may conclude that a general factor is likely and, therefore, pursue a different type of rotation. A second researcher may be

comfortable with a Thurstonian viewpoint and seek multiple ability factors using a varimax rotation. We will explore this issue in more detail later, but it is worth pointing out here that a researcher encounters many choice points in the process of conducting a factor analysis. It is not surprising, then, that different researchers may reach different conclusions from factor analysis, even when they are analyzing the same data set.

The Interpretation of Factors

Table 5.4 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab4>) indicates that five factors underlie the intercorrelations of the 24 ability tests. But what shall we call these factors? The reader may find the answer to this question disquieting, because at this juncture we leave the realm of cold, objective statistics and enter the arena of judgment, insight, and presumption. In order to interpret or name a factor, the researcher must make a reasoned judgment about the common processes and abilities shared by the tests with strong loadings on that factor. For example, in **Table 5.4** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05tab4>) it appears that factor I is verbal ability, because the variables with high loadings stress verbal skill (e.g., Sentence Completion loads .86, Word Meaning loads .84, and Paragraph Comprehension loads .81). The variables with low loadings also help sharpen the meaning of factor I. For example, factor I is not related to numerical skill (Numerical Puzzles loads .18) or spatial skill (Paper Form Board loads .16). Using a similar form of inference, it appears that factor II is mainly numerical ability (Add Digits loads .85, Count Groups of Dots loads .80). Factor III is less certain but appears to be a visual-perceptual capacity, and factor IV appears to be a measure of recognition. We would need to analyze the single test on factor V (Figure-Word) to surmise the meaning of this factor.

TABLE 5.4 The Rotated Varimax Factor Matrix for 24 Ability Variables

	<i>Factors</i>				
	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>
7. Sentence Completion	.86	.15	.13	.03	.07
9. Word Meaning	.84	.06	.15	.18	.08
6. Paragraph Comprehension	.81	.07	.16	.18	.10
5. General Information	.79	.22	.16	.12	-.02
8. Word Classification	.65	.22	.28	.03	.21
22. Problem Reasoning	.43	.12	.38	.23	.22
10. Add Digits	.18	.85	-.10	.09	-.01
12. Count Groups of Dots	.02	.80	.20	.03	.00
11. Code (Perceptual Speed)	.18	.64	.05	.30	.17
13. Straight and Curved Capitals	.19	.60	.40	-.05	.18
24. Arithmetic Problems	.41	.54	.12	.16	.24
21. Numerical Puzzles	.18	.52	.45	.16	.02
18. Number-Figure	.00	.40	.28	.38	.36
1. Visual Perception	.17	.21	.69	.10	.20
2. Cubes	.09	.09	.65	.12	-.18
4. Flags	.26	.07	.60	-.01	.15
3. Paper Form Board	.16	-.09	.57	-.05	.49

	<i>Factors</i>				
	<i>I</i>	<i>II</i>	<i>III</i>	<i>IV</i>	<i>V</i>
23. Series Completion	.42	.24	.52	.18	.11
20. Deduction	.43	.11	.47	.35	-.07
15. Number Recognition	.11	.09	.12	.74	-.02
14. Word Recognition	.23	.10	.00	.69	.10
16. Figure Recognition	.07	.07	.46	.59	.14
17. Object-Number	.15	.25	-.06	.52	.49
19. Figure-Word	.16	.16	.11	.14	.77

Note: Boldfaced entries signify subtests loading strongly on each factor.

These results illustrate a major use of factor analysis, namely, the identification of a small number of marker tests from a large test battery. Rather than using a cumbersome battery of 24 tests, a researcher could gain nearly the same information by carefully selecting several tests with strong loadings on the five factors. For example, the first factor is well represented by test 7, Sentence Completion (.86) and test 9, Word Meaning (.84); the second factor is reflected in test 10, Add Digits (.85), while the third factor is best illustrated by test 1, Visual Perception (.69). The fourth factor is captured by test 15, Number Recognition (.74), and Word Recognition (.69). Of course, the last factor loads well on only test 19, Figure-Word (.77).

Issues in Factor Analysis

Unfortunately, factor analysis is frequently misunderstood and often misused. Some researchers appear to use factor analysis as a kind of divining rod, hoping to find gold hidden underneath tons of dirt. But there is nothing magical about the technique. No amount of statistical analysis can rescue data based on trivial, irrelevant, or haphazard measures. If there is no gold to be found, then none will be found; factor analysis is not alchemy. Factor analysis will yield meaningful results only when the research was meaningful to begin with.

An important point is that a particular kind of factor can emerge from factor analysis only if the tests and measures contain that factor in the first place. For example, a short-term memory factor cannot possibly emerge from a battery of ability tests if none of the tests requires short-term memory. In general, the quality of the output depends upon the quality of the input. We can restate this point as the acronym GIGO, or "garbage in, garbage out."

Sample size is crucial to a stable factor analysis. Comrey (1973 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib331>)) offers the following rough guide:

Sample Size	Rating
50	Very poor
100	Poor
200	Fair
300	Good
500	Very good
1,000	Excellent

In general, it is comforting to have at least five subjects for each test or variable (Tabachnick & Fidell, 1989 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1607>)).

Finally, we cannot overemphasize the extent to which factor analysis is guided by subjective choices and theoretical prejudices. A crucial question in this regard is the choice between orthogonal axes and oblique axes. With **orthogonal axes**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss231>), the factors are at right angles to one another, which means that they are uncorrelated (Figures 5.1

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fig1>) and 5.2

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec2#ch05fig2>) both depict orthogonal axes). In many cases the clusters of factor loadings are situated such that oblique axes

provide a better fit. With **oblique axes**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss225>), the factors are correlated among themselves. Some researchers contend that oblique axes should always be used, whereas others take a more experimental approach. Tabachnick and Fidell (1989

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1607>)) recommend an exploratory strategy based on repeated factor analyses. Their approach is unabashedly opportunistic:

During the next few runs, researchers experiment with different numbers of factors, different extraction techniques, and both orthogonal and oblique rotations. Some number of factors with some combination of extraction and rotation produces the solution with the greatest scientific utility, consistency, and meaning; this is the solution that is interpreted.

With oblique rotations it is also possible to factor analyze the factors themselves. Such a procedure may yield one or more second-order factors. Second-order factors can provide support for the hierarchical organization of traits and may offer a rapprochement between ability theorists who posit a single general factor (e.g., Spearman) and those who promote several group factors (e.g., Thurstone). Perhaps both camps are correct, with the group factors sitting underneath the second-order general factor.

We turn now to a review of major theories of intelligence. A reminder: The justification for reviewing theories is to illustrate how they have influenced the structure and content of intelligence tests. In addition, the construct validity of IQ tests depends on the extent to which they embody specific theories of intelligence, so a review of theories is pertinent to test validation as well.

¹In this example, the variables are tests that produce more or less continuous scores. But the variables in a factor analysis can take other forms, so long as they can be expressed as continuous scores. For example, all of the following could be variables in a factor analysis: height, weight, income, social class, and rating-scale results.

²Technically, it is possible to represent all the factors as reference axes in n -dimensional space, where n is the number of factors. However, when working with more than two or three reference axes, visual representation is no longer feasible.

5.3 GALTON AND SENSORY KEENNESS

The first theories of intelligence were derived in the Brass Instruments era of psychology at the turn of the twentieth century. The reader will recall from **Topic 2A**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch02#ch02box1>) that Sir Francis Galton and his disciple J. McKeen Cattell thought that intelligence was underwritten by keen sensory abilities. This incomplete and misleading assumption was based on a plausible premise:

The only information that reaches us concerning outward events appears to pass through the avenues of our senses; and the more perceptive the senses are of difference, the larger is the field upon which our judgment and intelligence can act. (Galton, 1883

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib559>))

The sensory keenness theory of intelligence promoted by Galton and Cattell proved to be largely a psychometric dead end. However, we do see vestiges of this approach in modern chronometric analyses of intelligence such as the Reaction Time–Movement Time (RT–MT) apparatus, an experimental method favored by Jensen (1980

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib832>)) for the culture-reduced study of intelligence (**Figure 5.3**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec3#ch05fig3>)). In RT–MT studies, the subject is instructed to place the index finger of the preferred hand on the home button; then an auditory warning signal is sounded, followed (in 1 to 4 seconds) by one of the eight green lights going on, which the subject must turn off as quickly as possible by touching the microswitch button directly below it. RT is the time the subject takes to remove his or her finger from the home button after a green light goes on. MT is the interval between removing the finger from the home button and touching the button that turns off the green light. Jensen (1980

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib832>)) reported that indices of RT and MT correlated as high as .50 with traditional psychometric tests of intelligence.³

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec3#ch05fn03>) P. A. Vernon has also reported substantial relationships—as high as .70 for multiple correlations—between speed-of-processing RT-type measures and traditional measures of intelligence (Vernon, 1994). These findings suggest that speed-of-processing measures such as RT might be a useful addition to standardized intelligence test batteries. In general, test developers have resisted the implications of this line of research.

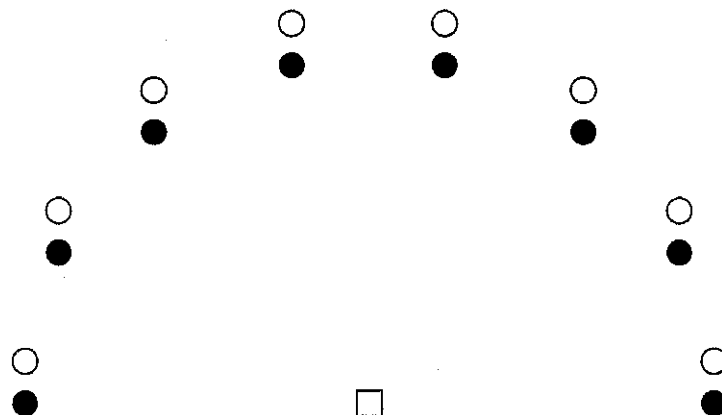


FIGURE 5.3 Schematic Diagram of a Reaction Time—Movement Time Apparatus

Note: The square box □ indicates the starting point; the open circles ○ indicate the signal lights; the dark circles ● indicate the push buttons.

One reason for the lack of ongoing progress in mental chronometry is the absence of standardization in measurement and data analysis. Not all devices for measuring reaction time are the same; consequently, the data from one laboratory cannot be compared to results from another setting. Making matters worse, many “reaction time” devices lump together RT (the time needed to lift the finger off the home button) and MT (the time in transit to the target button), which drastically obscures the relationship between chronometric data and intelligence (**Jensen, 2006**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib836>)). The problem with combining the two is that RT is related to IQ, whereas MT is a motor measure uncorrelated with IQ. In addressing these issues, Jensen (**2011**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib837>)) has commissioned a leading electronics company to create a standard apparatus for administering and recording reaction time and other indices of mental chronometry. Use of a single standard instrument would provide an vital foundation for progress in this area of assessment.

³Actually, the raw correlation coefficient is negative because *faster* reaction times (*lower* numerical scores) are associated with *higher* intelligence scores.

5.4 SPEARMAN AND THE g FACTOR

Based on extensive study of the patterns of correlations between various tests of intellectual and sensory ability, Charles Spearman (1904

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1526>) , 1923

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1527>) , 1927

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1528>)) proposed that intelligence consisted of two kinds of factors: a single **general factor**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss132>) g and numerous **specific factors** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss299>) $s_1, s_2, s_3,$ and so on. As a necessary adjunct to his theory, Spearman helped invent factor analysis to aid his investigation of the nature of intelligence. Spearman used this statistical technique to discern the number of separate underlying factors that must exist to account for the observed correlations between a large number of tests.

In Spearman's view, an examinee's performance on any homogeneous test or subtest of intellectual ability was determined mainly by two influences: g , the pervasive general factor, and s , a factor specific to that test or subtest. (An error factor e could also sway scores, but Spearman sought to minimize this influence by using highly reliable instruments.) Because the specific factor s was different for each intellectual test or subtest and was usually less influential than g in determining performance level, Spearman expressed less interest in studying it. He concentrated mainly on defining the nature of g , which he likened to an "energy" or "power" that serves in common the whole cortex. In contrast, Spearman considered s , the specific factor, to have a physiological substrate localized in the group of neurons serving the particular kind of mental operation demanded by a test or subtest. Spearman (1923 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1527>)) wrote, "These neural groups would thus function as alternative 'engines' into which the common supply of 'energy' could be alternatively distributed."

Spearman reasoned that some tests were heavily loaded with the g factor, whereas other tests—especially purely sensory measures—were representative mainly of a specific factor. Two tests each heavily loaded with g should correlate quite strongly. In contrast, psychological tests not saturated with g should show minimal correlation with one another. Much of Spearman's research was aimed at demonstrating the truth of these basic propositions derived from his theory. We have illustrated these points graphically in **Figure 5.4**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec4#ch05fig4>) . In this figure, each circle represents an intelligence test, and the degree of overlap between circles indicates the strength of correlation. Notice that tests A and B, each heavily loaded on g , correlate quite strongly. Tests C and D have weak loadings on g and subsequently do not correlate well.

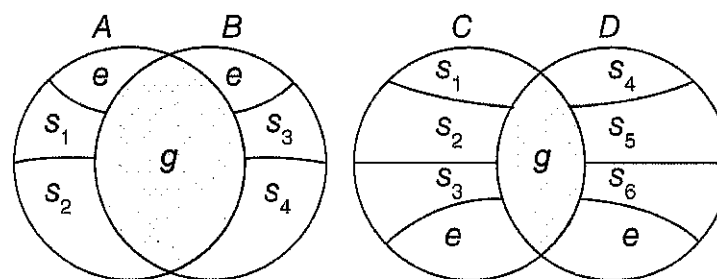


FIGURE 5.4 Spearman's Two-Factor Theory of Intelligence

Note: Tests *A* and *B* correlate strongly, whereas *C* and *D* correlate weakly. See text.

Spearman (1923 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1527>)) believed that individual differences in *g* were most directly reflected in the ability to use three principles of cognition: apprehension of experience, eduction of relations, and eduction of correlations. Incidentally, the little-used term *eduction* refers to the process of figuring things out. These three principles can be explained by examining how we solve analogies of the form $A:B::C:?$ that is, *A* is to *B* as *C* is to? A simple example might be HAMMER:NAIL::SCREWDRIVER:? To solve this analogy, we must first perceive and understand each term based on past experience; that is, we must have apprehension of experience. If we have no idea what a hammer, nail, and screwdriver are, there is little chance we can complete the analogy correctly. Next, we must infer the relation between the first two analogy terms, in this case, HAMMER and NAIL. Using a somewhat stilted phrase, Spearman referred to the ability to infer the relation between two concepts as eduction of relations. The final step, eduction of correlates, refers to the ability to apply the inferred principle to the new domain, in this case, applying the rule inferred to produce the correct response, namely, SCREWDRIVER:SCREW.

Although Spearman's physiological speculations have been largely dismissed, the idea of a general factor has been a central topic in research on intelligence and is still very much alive today (Jensen, 1979 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib831>)). The correctness of the *g* factor viewpoint is more than an academic issue. If it is true that a single, pervasive general factor is the essential wellspring of intelligence, then psychometric efforts to produce factorially pure subtests (e.g., measuring verbal comprehension, perceptual organization, short-term memory, and so on) are largely misguided. To the extent that Spearman is correct, test developers should forgo subtest derivation and concentrate on producing a test that best captures the general factor.

The most difficult issue faced by Spearman's two-factor theory is the existence of group factors. As early as 1906, Spearman and his contemporaries noted that relatively dissimilar tests could have correlations higher than the values predicted from their respective *g* loadings (Brody & Brody, 1976 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib221>)). This finding raised the possibility that a group of diverse measures might share in common a unitary ability other than *g*. For example, several tests might share a common unitary memorization factor that was halfway between the *g* factor and the various *s* factors unique to each test. Of course, the existence of group factors is incompatible with Spearman's meticulous two-factor theory.

5.5 THURSTONE AND THE PRIMARY MENTAL ABILITIES

Thurstone (1931 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1646>)) developed factor-analysis procedures capable of searching correlation matrices for the existence of group factors. His methods permitted a researcher to discover empirically the number of factors present in a matrix and to define each factor in terms of the tests that loaded on it. In his analysis of how scores on different kinds of intellectual tests correlated with each other, Thurstone concluded that several broad group factors—and not a single general factor—could best explain empirical results. At various points in his research career, he proposed approximately a dozen different factors. Only seven of these factors have been frequently corroborated (Thurstone, 1938 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1647>); Thurstone & Thurstone, 1941 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1650>)) and they have been designated **primary mental abilities** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss252>) (PMAs). They are as follows:

- **Verbal Comprehension:** The best measure is vocabulary, but this ability is also involved in reading comprehension and verbal analogies.
- **Word Fluency:** Measured by such tests as anagrams or quickly naming words in a given category (e.g., foods beginning with the letter *S*).
- **Number:** Virtually synonymous with the speed and accuracy of simple arithmetic computation.
- **Space:** Such as the ability to visualize how a three-dimensional object would appear if it was rotated or partially disassembled.
- **Associative Memory:** Skill at rote memory tasks such as learning to associate pairs of unrelated items.
- **Perceptual Speed:** Involved in simple clerical tasks such as checking for similarities and differences in visual details.
- **Inductive Reasoning:** The best measures of this factor involve finding a rule, as in a number series completion test.

Thurstone (1938 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1647>)) published the Primary Mental Abilities Test consisting of separate subtests, each designed to measure one PMA. However, he later acknowledged that his primary mental abilities correlated moderately with each other, proving the existence of one or more second-order factors. Ultimately, Thurstone acknowledged the existence of *g* as a higher-order factor. By this time, Spearman had admitted the existence of group factors representing special abilities, and it became apparent that the differences between Spearman and Thurstone were largely a matter of emphasis (Brody & Brody, 1976 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib221>)). Spearman continued to believe that *g* was the major determinant of correlations between test scores and assigned a minor role to group factors. Thurstone reversed these priorities.

P. E. Vernon (1950 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1697>)) provided a rapprochement between these two viewpoints by proposing a hierarchical group factor theory. In his view, *g* was the single factor at the top of a hierarchy that included two major group factors labeled verbal-educational (V:ed) and practical-mechanical-spatial-physical (k:m). Underneath these two major group factors were several minor group factors resembling the PMAs of Thurstone; specific factors occupied the bottom of the hierarchy.

Thurstone's analysis of PMAs continues to influence test development even today. Schaie (1985 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1452>)) has revised and modified the Primary Mental Abilities Test and used these measures in an enormously influential longitudinal study of adult intelligence. If intelligence were mainly a matter of g , then the group factors should change at about the same rate with aging. In support of the group factor approach to intellectual testing, Schaie (1985 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1452>)) reports that some PMAs show little age-related decrement (Verbal Comprehension, Word Fluency, Inductive Reasoning), whereas other PMAs decline more rapidly in old age (Space, Number). Thus, there may be practical real-world reasons for reporting group factors and not condensing all of intelligence into a single general factor.

5.6 CATTELL-HORN-CARROLL (CHC) THEORY

Raymond Cattell (1941

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib284>), 1971

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib285>) proposed an influential theory of the structure of intelligence that has been revised and extended by John Horn (1968

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib784>), 1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib786>) and John Carroll

(1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>)). Based on

the reanalysis of 461 data sets from hundreds of independent studies published by other researchers,

Carroll's contributions to the theory are especially vital. The ensuing theory, known as Cattell-Horn-

Carroll (CHC) theory, is a taxonomic tour de force that synthesizes the findings from almost a century of

factor-analytic research on intelligence. Many psychometricians consider CHC theory to possess the

strongest empirical foundation of any theory of intelligence and also to provide the most far-reaching

implications for psychological testing (McGrew, 1997

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1095>)). Although the "big

picture" of CHC theory is well established, researchers continue to refine the details. Under the direction

of Kevin McGrew, the Institute for Applied Psychometrics manages an informative website dedicated to

the advancement of CHC theory and applications (www.iapsych.com (<http://www.iapsych.com>)).

<i>Stratum III</i>	<i>Stratum II</i>		<i>Stratum I</i>
	Fluid Intelligence/Reasoning	(<i>Gf</i>)	5 narrow abilities
	Crystallized Intelligence/Knowledge	(<i>Gc</i>)	10 narrow abilities
	Domain-Specific Knowledge	(<i>Gkn</i>)	7 narrow abilities
General	Visual-Spatial Abilities	(<i>Gv</i>)	11 narrow abilities
Intelligence, <i>g</i>	Auditory Processing	(<i>Ga</i>)	13 narrow abilities
	Broad Retrieval [Memory]	(<i>Gr</i>)	13 narrow abilities
	Cognitive Processing Speed	(<i>Gs</i>)	7 narrow abilities
	Decision/Reaction Time or Speed	(<i>Gt</i>)	5 narrow abilities

FIGURE 5.5 Outline of the CHC Three-Stratum Theory of Cognitive Abilities

Source: Based on Carroll, J. B. (1993). *Cognitive abilities: A survey of factor analytic studies*.

New York: Cambridge University Press, and table 3 from www.iapsych.com

(<http://www.iapsych.com>).

According to CHC theory, intelligence consists of pervasive, broad, and narrow abilities that are hierarchically organized. These are known as Stratum III, II, and I, respectively (Figure 5.5

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec6#ch05fig5>)). At the highest and most pervasive level called Stratum III, a single general factor known as little *g* oversees all cognitive activities. Stratum II capacities, which reside beneath general intelligence, include several prominent and well-established abilities. In Figure 5.5

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec6#ch05fig5>), we have depicted eight abilities originally identified by Carroll (1993

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>)), but other researchers have proposed a slightly larger list that includes additional tentative entries such as

psychomotor, olfactory, and kinesthetic abilities. The precise name given to each broad factor differs slightly from one theorist to another, as well as the scale abbreviations. Even so, there is strong consensus for the essential list. These broad factors include “basic constitutional and longstanding characteristics of individuals that can govern or influence a great variety of behaviors in a given domain” (Carroll, 1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>), p. 634). The narrow abilities at Stratum I include approximately 70 abilities identified by Carroll (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>)) in his comprehensive review of factor-analytic studies of intelligence. As might be expected, the list of narrow abilities is continually revised and expanded with ongoing research. These narrow abilities “represent greater specializations of abilities, often in quite specific ways that reflect the effects of experience and learning, or the adoption of particular strategies of performance” (Carroll, 1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>), p. 634).

Definitions of CHC Broad Ability Factors

As noted, the broad factors of CHC are more firmly established than the narrow abilities, which continue to undergo revision and extension. We provide brief definitions of the broad factors, based on Carroll (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>)), McGrew (1997 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1095>)), and www.iapsych.com (<http://www.iapsych.com>).

- Fluid Intelligence/Reasoning (*Gf*): Fluid intelligence encompasses high-level reasoning and is used for novel tasks that cannot be performed automatically. The mental operations of **fluid intelligence** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss121>) may involve drawing inferences, forming concepts, generating and testing hypotheses, understanding implications, inductive reasoning, and deductive reasoning. The classic example of fluid intelligence is found in matrix reasoning tasks such as Raven’s Progressive Matrices (Raven, 2000 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1334>)). The abilities that make up fluid intelligence are largely nonverbal and not heavily dependent on exposure to a specific culture. For these reasons, Cattell (1940) believed that measures of fluid intelligence were culture-free. Based on this assumption, he devised the Culture Fair Intelligence Test (CFIT) in an attempt to eliminate cultural bias in testing. Of course, calling a test culture fair does not make it necessarily so. In fact, the goal of a completely culture-free intelligence test has proved elusive. We discuss the CFIT in more detail in **Topic 6A** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch06#ch06box1>), Group Tests of Ability and Related Concepts.
- Crystallized Intelligence/Knowledge (*Gc*): This form of intelligence is typically defined as an individual’s breadth and depth of acquired cultural knowledge—knowledge of the language, information, and concepts of a person’s culture. The quintessential example is the extent of vocabulary that an individual understands. But **crystallized intelligence** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss86>) also includes the application of verbal and cultural knowledge (e.g., oral production, verbal fluency, and communication ability). Because crystallized intelligence arises when fluid intelligence is applied to cultural products, we would expect these two kinds of cognitive ability to possess a strong correlation. In fact, it is commonly found that measures of crystallized and fluid intelligence possess a healthy relationship ($r = .5$).
- Domain-Specific Knowledge (*Gkn*): Domain-specific knowledge represents a person’s acquired knowledge in one or more specialized domains that do not represent the typical experiences of

individuals in the culture. This might include, for example, knowledge of biology, skill in lip reading, or knowledge of how to use computers.

- **Visual-Spatial Abilities (*Gv*):** This ability has to do with imagining, retaining, and transforming mental representations of visual images. For example, visual-spatial ability involves the capacity to predict how a shape will appear when it is rotated, or to identify quickly a known object from a vague, incomplete picture, or to find an object hidden in a picture. This capacity includes visual memory.
- **Auditory Processing (*Ga*):** This is the ability to perceive auditory information accurately, which involves the capacity to analyze, comprehend, and synthesize patterns or groups of sounds. Auditory processing involves the ability to discriminate speech sounds and to judge and discriminate tonal patterns in music. A key characteristic of *Ga* abilities is the cognitive talent needed to control the perception of auditory information (i.e., to filter signal from noise).
- **Broad Retrieval [Memory] (*Gr*):** Broad retrieval includes the ability to consolidate and store new information in long-term memory and then to retrieve the information later through association. Included in broad retrieval are such narrow abilities as associative memory (e.g., when provided the first part, recalling the second part of a previously learned but unrelated pair of items), ideational fluency (e.g., ability to call up ideas), and naming facility (e.g., rapidly providing the names of familiar faces). Some researchers further divide the broad memory factor into additional subtypes. In addition, some theorists propose a separate broad factor for short-term memory (*Gsm*), the ability to retain awareness of events that have occurred in the last minute or less (**Horn & Masunaga, 2000** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib788>)).
- **Cognitive Processing Speed (*Gs*):** This ability refers to the speed of executing overlearned or automatized cognitive processes, especially when high levels of attention and focused concentration are required. For example, the ability to perform simple arithmetic calculations with lightning speed would indicate a high level of *Gs* ability.
- **Decision/Reaction Time or Speed (*Gt*):** This is the ability to make decisions quickly in response to simple stimuli, typically measured by reaction time. For example, the capacity to quickly press the space bar whenever the letter *X* appears on a computer screen would involve the use of *Gt* ability.

Utility of CHC Theory

CHC theory is unusual in its detail, which permits robust theory testing. A number of lines of evidence support its validity. For example, the structure of intelligence as posited by CHC theory has been shown to be invariant across a number of key variables, including age, ethnicity, and gender (**Bickley, Keith, & Wolfe, 1995** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib158>); **Keith, 1999** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib873>); **Carroll, 1993** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>)). In empirical studies, the broad CHC abilities also reveal theory-confirming relationships with numerous academic and occupational variables (**McGrew & Flanagan, 1998**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1096>)). In one study, for example, measures of CHC broad and narrow cognitive abilities were selectively and appropriately related to mathematics achievement in a representative sample of children and adolescents (**Floyd, Evans, & McGrew, 2003**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib513>)). In general, practitioners praise the CHC approach to partitioning intelligence because the broad and narrow abilities are empirically verified and possess meaningful real-world implications (**Fiorello & Primerano, 2005** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib503>)).

5.7 GUILFORD AND THE STRUCTURE-OF-INTELLECT MODEL

After World War II, J. P. Guilford (1967

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib666>) , 1985

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib667>)) continued the search for the factors of intelligence that had been initiated by Thurstone. Guilford soon concluded that the number of discernible mental abilities was far in excess of the seven proposed by Thurstone. For one thing, Thurstone had ignored the category of creative thinking entirely, an unwarranted oversight in Guilford's view. Guilford also found that if innovative types of tests were included in the large batteries of tests he administered his subjects, then the pattern of correlations between these tests indicated the existence of literally dozens of new factors of intellect. Furthermore, Guilford noticed that some of these new factors had recurring similarities with respect to the kinds of mental processes involved, the kinds of information featured, or the form that the items of information took. As a result of these recurring similarities in the newly discovered factors of intellect, he became convinced that these multitudinous factors could be grouped along a small number of main dimensions. Guilford (1967

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib666>)) proposed an elegant structure-of-intellect (SOI) model to summarize his findings. Visually conceived, Guilford's SOI model classifies intellectual abilities along three dimensions called operations, contents, and products.

By *operations*, Guilford has in mind the kind of intellectual operation required by the test. Most test items emphasize just one of the operations listed here:

Cognition	Discovering, knowing, or comprehending
Memory	Committing items of information to memory, such as a series of numbers
Divergent	Retrieving from memory items of production a specific class, such as naming objects that are both hard and edible
Convergent	Retrieving from memory a correct production item, such as a crossword puzzle word
Evaluation	Determining how well a certain item of information satisfies specific logical requirements

Contents refers to the nature of the materials or information presented to the examinee. The five content categories are as follows:

Visual	Images presented to the eyes
Auditory	Sounds presented to the ears
Symbolic	Such as mathematical symbols that stand for something
Semantic	Meanings, usually of word symbols
Behavioral	The ability to comprehend the mental state and behavior of other persons

The third dimension in Guilford's model, *products*, refers to the different kinds of mental structures that the brain must produce to derive a correct answer. The six kinds of products are as follows:

Unit	A single entity having a unique combination of properties or attributes
Class	What it is that similar units have in common, such as a set of triangles or high-pitched tones
Relation	An observed connection between two items, such as two tones an octave apart
System	Three or more items forming a recognizable whole, such as a melody or a plan for a sequence of actions
Transformation	A change in an item of information, such as a correction of a misspelling
Implication	What an individual item implies, such as to expect thunder following lightning

In total, then, Guilford (1985

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib667>) identified five types of operations, five types of content, and six types of products, for a total of $5 \times 5 \times 6$ or 150 factors of intellect. Each combination of an operation (e.g., memory), a content (e.g., symbolic), and a product (e.g., units) represents a different factor of intellect. Guilford claims to have verified over 100 of these factors in his research.

The SOI model is often lauded on the grounds that it captures the complexities of intelligence. However, this is also a potential Achilles' heel for the theory. Consider one factor of intellect, memory for symbolic units. A test that requires the examinee to recall a series of *spoken* digits (e.g., Digit Span on the WAIS-III) might capture this factor of intellect quite well. But so might a *visual* digit span test and perhaps even an analogous test with *tactile* presentation of symbols, such as vibrating rods applied to the skin. Perhaps we need a separate cube for hearing, vision, and touch; such an expanded model would incorporate 450 factors of intellect, surely an unwieldy number.

Although it seems doubtful that intelligence could involve such a large number of unique abilities, Guilford's atomistic view of intellect nonetheless has caused test developers to rethink and widen their understanding of intelligence. Prior to Guilford's contributions, most tests of intelligence required mainly convergent production—the construction of a single correct answer to a stimulus situation. Guilford raised the intriguing possibility that **divergent production** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss93>)—the creation of numerous appropriate responses to a single stimulus situation—is also an essential element of intelligent behavior. Thus, a question such as “List as many consequences as possible if clouds had strings hanging down from them” (divergent production) might assess an aspect of intelligence not measured by traditional tests.

5.8 PLANNING, ATTENTION, SIMULTANEOUS, AND SUCCESSIVE (PASS) THEORY

Some modern conceptions of intelligence owe a debt to the neuropsychological investigations of the Russian psychologist Aleksandr Luria (1902–1977). Luria (1966 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1017>)) relied primarily on individual case studies and clinical observations of brain-injured soldiers to arrive at a general theory of cognitive processing. The heart of his theory is as follows:

Analysis shows that there is strong evidence for distinguishing two basic forms of integrative activity of the cerebral cortex by which different aspects of the outside world may be reflected. . . . The first of these forms is the integration of the individual stimuli arriving in the brain into simultaneous, and primarily spatial groups, and the second is the integration of individual stimuli arriving consecutively in the brain into temporally organized, successive series. (Luria, 1966 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1017>))

Since this approach focuses upon the mechanics by which information is processed, it is often called an information processing theory.

Luria (1970 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1018>)) proposed three functional units in the brain. Processing of information proceeds from lower units to higher units. The first unit is found in subcortical areas including the brain stem, midbrain, and thalamus. Attentional processes originate here, including selective attention and resistance to distraction. The second unit consists of the rearward sensory portions of the cerebral cortex (parietal, temporal, and occipital lobes). This large unit subserves the simultaneous and successive processes discussed later in this chapter. These processes are to some extent lateralized, with simultaneous processing engaged more with the right hemisphere, and successive processing connected more with the left hemisphere. However, lateralization is relative, not absolute (Springer & Deutsch, 1997 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1542>)). The third unit is located in the frontal lobes. This is primarily where planning occurs and also where motor output initiates.

Naglieri and Das (1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1212>), 2005) have developed the Planning, Attention, Simultaneous, Successive (PASS) theory of intelligence as a modern extension of Luria's work. Planning involves the selection, usage, and monitoring of effective solutions to problems. Anticipation of consequences and use of feedback are essential. Planning also entails impulse control. As noted, the frontal lobes are heavily engaged in this process. Even though it is listed first in the PASS acronym, Planning is actually the last stage of information processing. The first process is Attention, which requires selectively attending to some stimuli while ignoring others. In some cases, attention also entails vigilance over a period of time. Difficulties with this process underlie attention deficit/hyperactivity disorder. As noted, the brain stem and other midline subcortical structures are vital to attentional processes.

Simultaneous processing

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss292>) of information is characterized by the execution of several different mental operations simultaneously. Forms of thinking and perception that require spatial analysis, such as drawing a cube, require simultaneous information

processing. In drawing, the examinee must simultaneously apprehend the overall shape and guide hand and fingers in the execution of the shape. A sequential approach to drawing a cube (if one were even possible) would be horrifically complex. In effect, the examinee would have to draw individual lines of highly specific lengths and angular orientations, and just hope that everything would line up. In the absence of a simultaneous mental gestalt to guide the drawing, a distorted production is almost guaranteed. **Successive processing**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss318>) of information is needed for mental activities in which a proper sequence of operations must be followed. This is in sharp contrast to simultaneous processing (such as drawing), for which sequence is unimportant. Successive processing is needed in remembering a series of digits, repeating a string of words (e.g., *shoe, ball, egg*), and imitating a series of hand movements (fist, palm, fist, fist, palm). Most forms of information processing require an interplay of simultaneous and successive mechanisms. Das (1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib388>)) cites the example of reading an unfamiliar word such as *taciturn*:

The single letters are to be recognized, and that involves simultaneous coding. The reader matches the visual shape of the letter with a mental dictionary and comes up with a name for it. The letter sequences, then, have to be formed (successive coding) and blended together as a syllable (simultaneous). Then the string of syllables has to be made into a word (successive), the word is recognized (simultaneous), and a pronunciation program is then assembled (successive), leading to oral reading (successive and simultaneous).

Das admits that this may be a simplified view of what occurs when a reader is confronted with a word. The essential point is that higher-level information processing relies upon an interplay of specific, anatomically localizable forms of information processing.

The challenge of a simultaneous-successive approach to the assessment of intelligence is to design tasks that tap relatively pure forms of each approach to information processing. Tests that use this strategy are the Kaufman Assessment Battery for Children II (K-ABC-II), discussed in the next topic, and the Das-Naglieri Cognitive Assessment System (Das & Naglieri, 2012). The Das-Naglieri battery includes successive tasks that involve rapid articulation (such as, "Say *can, ball, hot* as fast as you can 10 times") and simultaneous measures of both verbal and nonverbal tasks. The battery also assesses planning and attention, so as to embody the PASS theory (Naglieri & Das, 2005).

5.9 INFORMATION PROCESSING THEORIES OF INTELLIGENCE

Information processing conceptions of intelligence propose models of how individuals mentally represent and process information. Borrowing from Campione and Brown (1978 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib265>)), Borkowski (1985 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib181>))) has put forward a comprehensive theory that bears a loose analogy to the functioning of a computer. The **architectural system** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss16>) (hardware) refers to biologically based properties necessary for information processing, such as memory span and speed of encoding/decoding information. Properties of the architectural system include capacity (e.g., number of slots in short-term memory, capacity of long-term memory), durability (rate of information loss), and efficiency of operation (e.g., rate of memory search). The architectural system is considered to be relatively “hardwired” and impervious to change by the environment.

In addition to the structural component of intelligence, there are various functional components (software). The **executive system** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss103>), which refers to environmentally learned components that steer problem solving, provides overall guidance to the functional components. Elements of the executive system include the knowledge base (retrieval of knowledge from long-term memory), schemes (rules of thinking), control processes (rules and strategies such as self-checking and rehearsal), and metacognition (self-awareness of one’s own thought processes). Metacognition is the process of thinking about thinking. Flavell (1976 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib511>))), who pioneered research on this topic, explained it as follows:

Metacognition refers to one’s knowledge concerning one’s own cognitive processes or anything related to them, e.g., the learning-relevant properties of information or data. For example, I am engaging in metacognition if I notice that I am having more trouble learning A than B; if it strikes me that I should double check C before accepting it as fact. (p. 232)

The information processing approach to intelligence has generated a large body of research, especially on the concept of metacognition. A consistent finding in this literature is that individuals who use metacognitive strategies perform at much higher levels than those who do not (**Montague & Bos, 1990** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1169>))). For example, in a study of 32 Israeli kindergarten children who were taught metacognition related to mathematics, metacognitive skills explained more of the variance in mathematics performance than general ability (**Mevarech, 1995** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1136>))). Metacognition is essential to intelligence and is one of the primary influences on student learning (**Wang, Haertel, & Walberg, 1990** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1716>))).

5.10 GARDNER AND THE THEORY OF MULTIPLE INTELLIGENCES

Howard Gardner (1983

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib563>), 1993

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib566>) has proposed a theory of multiple intelligences based loosely on the study of brain–behavior relationships. He argues for the existence of several relatively independent human intelligences, although he admits that the exact nature, extent, and number of the intelligences have not yet been definitively established. Gardner (1983 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib563>)) outlines the criteria for an autonomous intelligence as follows:

- Potential isolation by brain damage—the faculty can be destroyed, or spared in isolation, by brain injury.
- Existence of exceptional individuals such as savants—the faculty is uniquely spared in the midst of general intellectual mediocrity.
- Identifiable core operations—the faculty relies upon one or more basic information processing operations.
- Distinctive developmental history—the faculty possesses an identifiable developmental history, perhaps including critical periods and milestones.
- Evolutionary plausibility—admittedly speculative, a faculty should have evolutionary antecedents shared with other organisms (e.g., primate social organization).
- Support from experimental psychology—the faculty emerges in laboratory studies in cognitive psychology.
- Support from psychometric findings—the faculty reveals itself in measurement studies and is susceptible to psychometric measurement.
- Susceptibility to symbol encoding—the faculty can be communicated via symbols including (but not limited to) language, picturing, and mathematics.

Based on these criteria, Gardner (1983

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib563>), 1993

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib566>) proposes that the following seven natural intelligences have been substantially confirmed. The seven intelligences are linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal. Three of these seven types of intelligence are well known—linguistic (i.e., verbal) intelligence, logical-mathematical intelligence, spatial intelligence—and numerous formal tests have been devised to measure them, so we will not discuss them further here. The other four variations of intelligence are somewhat novel and, therefore, require more detailed presentation.

Bodily-kinesthetic intelligence includes the types of skills used by athletes, dancers, mime artists, typists, or “primitive” hunters. Although Western cultures are generally loath to consider the body as a form of intelligence, this is not the case in much of the rest of the world, nor was it true in our evolutionary history. Indeed, persons who could skillfully avoid predators, climb trees, hunt animals, and prepare tools were more likely to survive and pass on their genes to succeeding generations.

The personal intelligences include the capacity to have access to one’s own feeling life (intrapersonal) as well as the ability to notice and make distinctions about the moods, temperaments, motivations, and intentions of others (interpersonal). Thus, personal intelligence encompasses both an intrapersonal and an interpersonal version. The former is found in great novelists who can write introspectively about

their feelings, while the latter is often seen in religious and political leaders (e.g., Mahatma Gandhi or Lyndon Johnson) who can fathom the intentions and desires of others and use this information to influence them and form useful alliances.

Musical intelligence is perhaps the least understood of Gardner's intelligences. Persons with good musical intelligence easily learn to perform an instrument or to write their own compositions. Although knowledge of the structural aspects of melody, rhythm, and timbre is important to musical intelligence, Gardner notes that many experts place the affective or feeling aspects of music at its core. He believes that when the neurological underpinnings of music are finally unraveled, we will have "an explanation of how emotional and motivational factors are intertwined with purely perceptual ones" (Gardner, 1983 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib563>)).

The savant phenomenon provides strong support for the existence of separate intelligences, including musical intelligence.⁴ (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec10#ch05fn04>) A **savant** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss285>) is a mentally deficient individual who has a highly developed talent in a single area such as art, rapid calculation, memory, or music. An example is the extraordinary case of Leslie Lemke, who was born blind and with mental retardation and cerebral palsy. He was not supposed to live. His adoptive mother had to coax him to suck milk from a bottle. Later, she strapped him to her back to help him learn to walk. In spite of his severe disabilities, Leslie became enamored of the piano and showed incredible precocity at picking out melodies on it. Within a few years, at the age of 18, he could listen to a piece of classical piano music a single time and then play it back flawlessly (Patton, Payne, & Beirne-Smith, 1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1270>)). The reader can find additional savant case studies in Miller (1989 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1148>)) and Treffert (1989 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1663>)).

Recently, Gardner (1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib567>)) has added three tentative candidates to his list of intelligences. These are naturalistic, spiritual, and existential intelligences. Naturalistic intelligence is the kind shown by people who are able to discern patterns in nature. Charles Darwin would be a prime example of such a person. Gardner believes that the evidence for this kind of intelligence is relatively strong. In contrast, spiritual intelligence (a concern with cosmic and spiritual issues in one's development) and existential intelligence (a concern with ultimate issues, including the meaning of life) are less well proved as independent intelligences. In general, the theory of multiple intelligence is compelling in its simplicity, but there is little empirical investigation of its validity.

⁴Historically, savants have also been called *idiot savants*, which refers, literally, to a person who is both profoundly retarded and yet "wise" at the same time. For obvious reasons, the prefix has been dropped.

5.11 STERNBERG AND THE TRIARCHIC THEORY OF SUCCESSFUL INTELLIGENCE

Sternberg (1985b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1561>), 1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>), 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1565>)) takes a much wider view on the nature of intelligence than most previous theorists. In addition to proposing that certain mental mechanisms are required for intelligent behavior, he also emphasizes that intelligence involves adaptation to the real-world environment. His theory emphasizes what he calls successful intelligence or “the ability to adapt to, shape, and select environments to accomplish one’s goals and those of one’s society and culture” (Sternberg & Kaufman, 1998 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1569>), p. 494).

Sternberg’s theory is called *triarchic* (ruled by three) because it deals with three aspects of intelligence: componential intelligence, experiential intelligence, and contextual intelligence. Each of these types of intelligence has two or more subcomponents. The entire theory is outlined in **Table 5.5** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec11#ch05tab5>).

Componential intelligence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss56>), also known as analytical intelligence, consists of the internal mental mechanisms that are responsible for intelligent behavior. The components of intelligence serve three different functions. *Metacomponents* are the executive processes that direct the activities of all the other components of intelligence. They are responsible for determining the nature of an intellectual problem, selecting a strategy for solving it, and making sure that the task is completed. The metacomponents receive constant feedback as to how things are going in problem solving. Persons who are strong on the metacomponential aspect of intelligence are very good at allocating their intellectual resources.

TABLE 5.5 An Outline of Sternberg’s Triarchic Theory of Intelligence

Componential (Analytical) Intelligence

Metacomponents or executive processes (e.g., planning)

Performance components (e.g., syllogistic reasoning)

Knowledge acquisition components (e.g., ability to acquire vocabulary words)

Experiential (Creative) Intelligence

Ability to deal with novelty

Ability to automatize information processing

Contextual (Practical) Intelligence

Adaptation to real-world environment

Selection of a suitable environment

Shaping of the environment

Source: Summarized from Sternberg, R. J. (1986). *Intelligence applied: Understanding and increasing your intellectual skills*. San Diego, CA: Harcourt Brace Jovanovich.

In a problem-solving study using novel forms of analogies, Sternberg (1981 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1559>)) found that higher

intelligence is associated with spending relatively more time on global or higher-order planning, and relatively less time on local or lower-order planning. For example, consider this analogy problem:

Man: Skin:: (Dog, Tree):(Bark, Cat)

The examinee must choose the two correct terms on the right that will complete the analogy. (The correct choices are Tree and Bark.) Using reaction time measures for a series of such novel or nonentrenched problems, Sternberg (1981

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1559>) found that persons of higher intelligence spend more time in global planning—forming a macrostrategy that applies to this and similar problems—than did persons of lower intelligence. Thus, a crucial aspect of intelligence is knowing when to step back and allocate intellectual effort instead of obtusely attacking a difficult problem.

Performance components are the well-entrenched mental processes that might be used to perform a task or solve a problem. These aspects of intelligence are the ones that are probably measured the best by existing intelligence tests. Examples of performance components include short-term memory and syllogistic reasoning.

Knowledge acquisition components are the processes used in learning. Sternberg has emphasized that in order to understand what makes some people more skilled than others, we must understand their increased capacity to acquire those skills in the first place. A case in point is vocabulary knowledge, which is learned mainly in context rather than through direct instruction. More-intelligent persons are better able to use surrounding contexts to figure out what a word means; that is, they have greater knowledge-acquisition skills. Their increased vocabulary results, in large measure, from their increased ability to “soak up” the meanings of words they see and hear in their environment. Thus, vocabulary is an excellent measure of intelligence because it reflects people’s ability to acquire information in context.

The second aspect of Sternberg’s theory involves experiential intelligence. According to the theory, a person with good **experiential intelligence**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss105>) is able to deal effectively with novel tasks. Experiential intelligence is also known as creative intelligence. This aspect of his theory explains why Sternberg is so critical of most intelligence tests. For the most part, the existing tests measure things already learned by presenting tasks that the subject has already encountered. According to Sternberg, intelligence also involves the capacity to learn and think within new conceptual systems, not just to deal with tasks already encountered. A second aspect of experiential intelligence is the ability to automatize or “make routine” tasks that are encountered repeatedly. An example of automatizing that applies to most of us is reading, which is carried out largely without conscious thought. But any task or mental skill can be automatized, if it is practiced enough. Playing music is an example of an extremely high-level skill that can become automatized with enough practice.

The third aspect of Sternberg’s theory involves contextual intelligence. **Contextual intelligence**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss68>), also known as practical intelligence, is defined as “mental activity involved in purposive adaptation to, shaping of, and selection of real-world environments relevant to one’s life” (Sternberg, 1986 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>), p. 33). This aspect of Sternberg’s theory appears to acknowledge that human behavior has been shaped by selective pressures during our evolutionary history. Contextual intelligence has three parts: adaptation, selection, and shaping.

Adaptation refers to developing skills required by one's particular environment. Successful adaptation will differ from one culture to the next. In the pygmy cultures of Africa, adaptation might involve the ability to track elephants and kill them with poison-tipped spears. In the Western industrial nations, adaptation might involve presenting oneself favorably in a job interview.

Selection might be called niche finding. This aspect of contextual intelligence involves the ability to leave the environment we are in and to select a different environment more suitable to our talents and needs. Feldman (1982 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib493>)) has illustrated how selection can operate in the career choices of gifted children, thereby determining whether they are highly accomplished as adults. She followed up on the Quiz Kids who were featured in radio and television shows of the 1950s. These were extremely bright children by conventional standards, most with IQs of 140 and higher. A few became highly successful as adults. However, most of them led rather ordinary lives, devoid of the spectacular accomplishments that might have been predicted from their childhood precocity. Those who were most successful had found occupations highly suited to their abilities and interests. In sum, they had selected environmental niches that fitted them well. Sternberg would argue that the ability to select such environments is an important aspect of intelligence.

Shaping is another way to improve the fit between oneself and the environment, especially when selection of a new environment is not practical. In this application of contextual intelligence, we shape the environment itself so that it better fits our needs. An employee who convinces the boss to do things differently has used shaping to make the work environment more suited to his or her talents.

Sternberg (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1563>)) has developed a research instrument based on his theory and has used the test to examine the validity of the triarchic approach. The Sternberg Triarchic Abilities Test (STAT) is unique in going beyond the typical questions that invoke analytical intelligence; the test includes creative and practical questions as well. For example, in one subtest examinees are presented with a map of an area, such as an entertainment park, and then must answer questions about navigating effectively through the area shown in the map (practical intelligence). In another subtest examinees are presented with verbal analogies preceded by incorrect, counterfactual premises (e.g., money falls off trees). Examinees must solve the analogies as though the counterfactual premises were true (creative intelligence). In factor-analytic studies of American, Finish, and Spanish samples, the triarchic model was a better fit to the data than the usual outcome of finding a single factor of general intelligence (Sternberg, Castejon, Prieto, Hautamaki, & Grigorenko, 2001 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1572>)).

Although Sternberg's triarchic theory is the most comprehensive and ambitious model yet proposed, not all psychometric researchers have rushed to embrace it. Detterman (1984) cautions that we should investigate the basic cognitive components of intelligence before introducing higher-order constructs that may be unnecessary. Rogoff (1984 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1397>)) questions whether the three subtheories (componential, experiential, contextual) are sufficiently linked. Other comments on the triarchic theory can be found in *Behavioral and Brain Sciences* (1984 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1397>)), pp. 287–304).

Whatever the final verdict on the triarchic theory of intelligence, Sternberg's insistence that intelligence has several components not measured by traditional tests rings true to anyone who has studied or administered these tests. He cites the case of a colleague who was asked to test a number of residents at an institution for those with mental retardation. These residents had just planned and successfully

executed an escape from the security-conscious school, a feat requiring high levels of practical intelligence. Yet, when administered the Porteus Maze Test (**Porteus, 1965** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1318>)), a standardized test reputed to involve planning ability, they could not solve even the simplest maze correctly. Sternberg (**1986** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1562>)) has made it clear that intelligence just has too many components to be measured by any single test.

TOPIC 5B Individual Tests of Intelligence and Achievement

5.12 Orientation to Individual Intelligence Tests

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec12#ch05lev1sec12>)

5.13 The Wechsler Scales of Intelligence

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec13#ch05lev1sec13>)

5.14 The Wechsler Subtests: Description and Analysis

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05lev1sec14>)

5.15 Wechsler Adult Intelligence Scale-IV

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec15#ch05lev1sec15>)

5.16 Wechsler Intelligence Scale for Children-IV

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec16#ch05lev1sec16>)

5.17 Stanford-Binet Intelligence Scales: Fifth Edition

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec17#ch05lev1sec17>)

5.18 Detroit Tests of Learning Aptitude-4

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec18#ch05lev1sec18>)

5.19 The Cognitive Assessment System-II

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec19#ch05lev1sec19>)

5.20 Kaufman Brief Intelligence Test-2 (KBIT-2)

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec20#ch05lev1sec20>)

5.21 Individual Tests of Achievement

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec21#ch05lev1sec21>)

5.22 Nature and Assessment of Learning Disabilities

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec22#ch05lev1sec22>)

Individual intelligence testing is one of the major achievements of psychology since the founding of the discipline. In response to the success of the Binet-Simon scales in the early 1900s, psychologists developed and refined dozens of individual tests of intelligence patterned after this pathbreaking instrument. The explosive growth in group tests of intelligence, fostered by the enthusiastic acceptance of the Army Alpha and Beta tests during and after World War I, also provided impetus to the individual

testing movement. Many contemporary individual tests of intelligence owe their lineage to Binet, Simon, and the Army testing programs.

The successful application of intelligence tests inspired educators and psychologists to look for ways to appraise the academic progress of students with school-based achievement tests. In turn, this led to the puzzling discovery that many children of normal or even superior intelligence lagged far behind in school achievement. From this discovery, the concept of learning disability gradually developed, and a whole new field of assessment was born.

The purpose of this topic is to provide an overview of noteworthy approaches to the testing of individual intelligence and achievement, and to introduce the reader to the essentials of learning disability assessment. However, an exhaustive survey of individual cognitive tests is simply beyond the scope of this or any other basic reference. New and revised tests appear practically every month, and thousands of new research findings are published every year. We have chosen to review tests that are widely used or that illustrate interesting developments in theory or method. Readers can find information on additional tests in the *Mental Measurements Yearbook* series, now published every two or three years by the Buros Institute.

5.12 ORIENTATION TO INDIVIDUAL INTELLIGENCE TESTS

The individual intelligence tests reviewed in this topic include the following:

- Wechsler Adult Intelligence Scale-IV (WAIS-IV)
- Wechsler Intelligence Scale for Children-IV (WISC-IV)
- Stanford-Binet: Fifth Edition (SB5)
- Detroit Tests of Learning Aptitude-4 (DTLA-4)
- Cognitive Assessment System-II (CAS-II)
- Kaufman Brief Intelligence Test-2 (KBIT-2)

Collectively, these instruments probably account for 95 percent of the intellectual assessments conducted in the United States.

The Wechsler scales have dominated intelligence testing in recent years, but they are by no means the only viable choices for individual assessment. Many other instruments measure general intelligence just as well—some would say better. Consider the implications of a now familiar observation: For large, heterogeneous samples, scores on any two mainstream instruments (e.g., Wechsler, Stanford-Binet, McCarthy, Kaufman scales) typically correlate .80 to .90. Often the correlation between two mainstream instruments is nearly as high as the test-retest correlation for either instrument alone. For purposes of producing a global score, it would appear that any well-normed mainstream intelligence test will suffice.

But producing an overall score is not the only goal of assessment. In addition, the examiner usually desires to gain an understanding of the subject's intellectual functioning. For this purpose, the overall IQ is important, but there are instances in which the global score may be irrelevant or even misleading. To understand a referral's intellectual functioning, the examiner should also inspect the subtest scores in search of hypotheses that might explain the unique functioning of that individual. Of course, examiners need to undertake subtest analysis cautiously, armed with research-based findings on the nature and meaning of subtest scatter for the test in use (**Gregory, 1994b** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib647>)).

If the examiner's goal is to understand intellectual functioning and not merely to determine an overall score, the differences between tests become quite real. Every instrument approaches the measurement of intelligence from a different perspective and yields a distinctive set of subtest scores. Furthermore, a test well suited for one referral issue might perform abysmally in another context. For example, the WAIS-IV performs admirably in the testing of mild mental retardation but contains too few simple items for the effective assessment of persons with moderate or severe developmental disability.

A central axiom of assessment is that the choice of a testing instrument should be based on knowledge of its strengths and weaknesses as they pertain to the referral question. Put simply, the skilled examiner does not blindly rely on a single test for every referral! Instead, the skilled examiner flexibly chooses one or more instruments in light of the perceived assessment needs of the examinee. Each of the tests discussed in this topic has its special merits and also its particular shortcomings. The test user must know these strong and weak facets in order to choose the instruments best suited for each unique referral.

5.13 THE WECHSLER SCALES OF INTELLIGENCE

Beginning in the 1930s, David Wechsler, a psychologist at Bellevue Hospital in New York City, conceived a series of elegantly simple instruments that virtually defined intelligence testing in the mid- to late twentieth century. His influence on intelligence testing is exceeded only by the pathbreaking contributions of Binet and Simon. It is fitting that we begin the survey of individual tests with a historical summary of the Wechsler tradition, followed by a discussion of individual instruments.

Origins of the Wechsler Tests

Wechsler began work on his first test in 1932, seeking to devise an instrument suitable for testing the diverse patients referred to the psychiatric section of Bellevue Hospital in New York (**Wechsler, 1932** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1724>)). In describing the development of his first test, he later wrote, "Our aim was not to produce a set of brand new tests but to select, from whatever source available, such a combination of them as would meet the requirements of an effective adult scale" (**Wechsler, 1939**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1725>)). In fact, the content of his scales was largely inspired by earlier efforts such as the Binet scales and the Army Alpha and Beta tests (**Frank, 1983**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib527>)). Readers who peruse *Psychological Examining in the United States Army*, a volume edited by Yerkes (**1921**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1796>)) just after World War I, might be astonished to discover that Wechsler purloined dozens of test items from this source, many of which have survived to the present day in contemporary revisions of the Wechsler tests. Wechsler was not so much a creative talent as a pragmatist who fashioned a new and useful instrument from the spare parts of earlier, discontinued attempts at intelligence testing.

The first of the Wechsler tests, named the Wechsler-Bellevue Intelligence Scales, was published in 1939. In discussing the rationale for his new test, Wechsler (**1941**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1726>)) explained that existing instruments such as the Stanford-Binet were woefully inadequate for assessing adult intelligence. The Wechsler-Bellevue was designed to rectify several flaws noted in previous tests:

- The test items possessed no appeal for adults.
- Too many questions emphasized mere manipulation of words.
- The instructions emphasized speed at the expense of accuracy.
- The reliance on mental age was irrelevant to adult testing.

To correct these shortcomings, Wechsler designed his test specifically for adults, added performance items to balance verbal questions, reduced the emphasis upon speeded questions, and invented a new method for obtaining the IQ. Specifically, he replaced the usual formula

$$IQ = \frac{\text{Mental Age}}{\text{Chronological Age}}$$

with a new age-relative formula

$$IQ = \frac{\text{Attained or Actual Score}}{\text{Expected Mean Score for Age}}$$

This new formula was based on the interesting presumption—stated in the form of an axiom—that IQ remains constant with normal aging, even though raw intellectual ability might shift or even decline. The assumption of **IQ constancy**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss169>) is basic to the Wechsler scales. As Wechsler (1941

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1726>)) put it:

The constancy of the I.Q. is the basic assumption of all scales where relative degrees of intelligence are defined in terms of it. It is not only basic, but absolutely necessary that I.Q.'s be independent of the age at which they are calculated, because unless the assumption holds, no permanent scheme of intelligence classification is possible.

Although Wechsler's view has been largely accepted by contemporary test developers, it is important to stress that the assumption of IQ invariance with age is really a statement of values, a philosophical choice, and not necessarily an inherent characteristic of human nature.

Wechsler also hoped to use his test as an aid in psychiatric diagnosis. In pursuit of this goal, he divided his scale into separate verbal and performance sections. This division allowed the examiner to compare an examinee's facility in using words and symbols (verbal subtests) versus the ability to manipulate objects and perceive visual patterns (performance subtests). Large differences between verbal ability (*V*) and performance ability (*P*) were thought to be of diagnostic significance. Specifically, Wechsler believed that organic brain disease, psychoses, and emotional disorders gave rise to a marked $V > P$ pattern, whereas adolescent psychopaths and persons with mild mental retardation yielded a strong $P < V$ pattern. Subsequent research demonstrated many exceptions to these simple diagnostic rules, and also helped refine the nature of these two major elements of intelligence. For example, verbal intelligence is now better known as verbal comprehension, and performance intelligence is more commonly recognized as perceptual reasoning. Nonetheless, the distinction between verbal and performance skills has proved useful for many purposes, such as studying brain-behavior relationships, and examining age effects on intelligence. Wechsler's armchair division of subtests into verbal and performance sections, even though refined and extended by others, continues to endure as a major contribution to contemporary intelligence testing (Kaufman, Lichtenberger, & McLean, 2001).

General Features of the Wechsler Tests

Including revisions, David Wechsler and his followers have produced more than a dozen intelligence tests in a span of about 70 years. A major reason for the continued success of these instruments has been the faithful adherence to the familiar content and format first introduced in the Wechsler-Bellevue. By sticking with a single successful formula, Wechsler and company ensured that examiners could switch from Wechsler test to another with minimal retraining. This was not only good psychometrics but also shrewd marketing insofar as it guaranteed several generations of faithful test users.

The latest editions of the Wechsler intelligence tests—the WPPSI-IV, WISC-IV, and WAIS-IV—possess the following common features:

- Thirteen to fifteen subtests. The multisubtest approach allows the examiner to analyze intra-individual strengths and weaknesses rather than just to compute a single global score. In addition, it is possible to combine subtest scores in theoretical meaningful ways that provide

useful information on the broad factors of intelligence. As the reader will learn subsequently, the pattern of subtest and factor scores may convey useful information that is hidden in the overall level of performance.

- An empirically based breakdown into composite scores and a full scale IQ. Whereas the original Wechsler intelligence scales provided only two composite scores—Verbal IQ and Performance IQ—the revisions have been moving toward a more sophisticated partitioning into composites confirmed from factor-analytic research. The WISC-IV and WAIS-IV now yield composite or index scores in the same four areas:
 1. Verbal Comprehension
 2. Perceptual Reasoning
 3. Working Memory
 4. Processing Speed

The WPPSI-IV provides five index scores similar to the above (for ages 4:0 to 7:7) but also includes a Fluid Reasoning composite.

- A common metric for IQ and Index scores. The mean for IQ and Index scores is 100 and the standard deviation is 15 for all tests and all age groups. In addition, the scaled scores on each subtest have a mean of 10 and a standard deviation of approximately 3, which permits the examiner to analyze the subtest scores of the examinee for relative strengths and weaknesses.
- Common subtests for the different test versions. For example, the preschool, child, and adult Wechsler tests (WPPSI-IV, WISC-IV, and WAIS-IV) all share a common core of the same six subtests (**Table 5.6** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec13#ch05tab6>)). An examiner who masters the administration of a core subtest on any of the Wechsler tests (such as the Information subtest on the WAIS-IV) easily can transfer this skill within the Wechsler family of intellectual measures.

TABLE 5.6 Subtest Composition of the Wechsler Intelligence Tests

	<i>WPPSI-IV</i>	<i>WISC-IV</i>	<i>WAIS-IV</i>
Similarities	x	x	x
Vocabulary	x	x	x
Comprehension	x	x	x
Information	x	x	x
Word Reasoning		x	
Receptive Vocabulary	x		
Picture Naming	x		
Block Design	x	x	x
Picture Concepts	x	x	
Matrix Reasoning	x	x	x
Picture Completion		x	x
Visual Puzzles			x
Figure Weights			x
Object Assembly	x		

	<i>WPPSI- IV</i>	<i>WISC- IV</i>	<i>WAIS- IV</i>
L-N Sequencing^a (http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec13#ch05fn6)		x	x
Arithmetic		x	x
Digit Span		x	x
Coding		x	x
Symbol Search		x	x
Cancellation		x	x
Picture Memory		x	
Bug Search		x	
Zoo Memory		x	

^aLetter–Number Sequencing.

Note: The subtests common to all Wechsler intelligence tests are in boldface. Some subtests are optional or used as substitutions. See text for details.

5.14 THE WECHSLER SUBTESTS: DESCRIPTION AND ANALYSIS

Wechsler (1939 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1725>)) defined *intelligence* as “the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment.” He also believed that we can only know intelligence by what it enables a person to do. In designing his tests, then, Wechsler selected components to represent a wide array of underlying abilities so as to estimate the global capacity of intelligence. Furthermore, he asked his subjects to do things, not merely to answer questions. The Wechsler subtests are quite diverse and often rely on what Wechsler referred to as “mental productions.”

We present here a description of subtests from the WISC-IV and WAIS-IV. We also analyze the abilities tapped by each subtest and offer research-based comments. The reader is referred to **Topic 7A** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch07#ch07box1>) , Infant and Preschool Assessment, for a description of the subtests unique to the WPPSI-IV.

Information

The Information subtest is found on all three Wechsler intelligence tests. Factual knowledge of persons, places, and common phenomena is tested here. Questions for children are like the following:

- “How many eyes do you have?”
- “Who invented the telephone?”
- “What causes a solar eclipse?”
- “Which is the largest planet?”

Questions for adults are similar but progress to higher levels of difficulty. Difficult questions on the adult Information subtest resemble:

- “Which is the most common element in air?”
- “What is the population of the world?”
- “How does fruit juice get converted to wine?”
- “Who wrote *Madame Bovary*?”

Information items test general knowledge normally available to most persons raised in the cultural institutions and educational systems of Western industrialized nations. Indirectly, this subtest measures learning and memory skills insofar as subjects must retain knowledge gained from formal and informal educational opportunities in order to answer the Information items.

Information is usually regarded as one of the best measures of general ability among the Wechsler subtests (Kaufman, McLean, & Reynolds, 1988 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib867>)). For example, the WAIS-IV manual reveals that Information typically has the second or third highest correlation with Full Scale IQ across the 13 age groups (Wechsler, Coalson, & Raiford, 2008 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1737>)). Information consistently loads strongly on the first factor identified in factor analyses of the WAIS-IV subtest correlations (see the following). The first factor is labeled Verbal Comprehension. However, Information tends to reflect formal education and motivation for academic achievement and may therefore yield spuriously high ability estimates for perpetual students and avid readers.

Digit Span

Digit Span consists of two separate sections, Digits Forward and Digits Backward. In Digits Forward, the examiner reads a series of digits at one per second, then asks the subject to repeat them. If the subject answers correctly on two consecutive trials of the same length, the examiner proceeds to the next series, which is one digit longer, up to a maximum length of nine digits. For Digits Backward, a similar procedure is used, except the examinee must repeat the digits in reverse order, up to a maximum length of eight digits. For example, the examiner reads:

“6-1-3-4-2-8-5”

and the subject tries to repeat the numbers in the reverse order:

“5-8-2-4-3-1-6.”

On the WAIS-IV only, the Digit Span subtest also includes a third section called Digit Sequencing. For this part, the examinee is asked to sort the series of digits into their correct order. For example, if the examiner says:

“1-7-4-9-2”

the examinee should respond:

“1-2-4-7-9.”

Digit Span is a measure of immediate auditory recall for numbers. Facility with numbers, good attention, and freedom from distractibility are required. Performance on this subtest may be affected by anxiety or fatigue, and many clinicians have noted that patients hospitalized for medical or psychiatric reasons frequently perform poorly on Digit Span.

Digits Forward and Digits Backward may assess fundamentally different abilities. Digits Forward seems to require the examinee to access an auditory code in sequential fashion. In contrast, to perform Digits Backward, the examinee must form an internal visual memory trace from the orally presented numerical sequences and then visually scan from end to beginning. Digits Backward is clearly the more complex test; not surprisingly, it loads higher on general intelligence than does Digits Forward (Jensen & Osborne, 1979 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib831>)). Gardner (1981 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib569>)) argues that examiners should supplement standard reporting procedures and list separate subscores for Digit Span. He presents separate means, standard deviations, and percentile ranks on Digits Forward and Backward for children ages 5 to 15.

Vocabulary

The Vocabulary subtest is found on all three Wechsler intelligence tests. The examinee is asked to define up to several dozen words of increasing difficulty while the examiner writes down each response verbatim. For example, on an easy item the examiner might ask, “What is a cup?” and the examinee would get partial credit for answering, “You drink with it” and full credit for answering, “It has a handle, holds liquids, and you drink from it.” For adults and bright children, the advanced items on the Wechsler Vocabulary subtests can be very challenging, on a par with *tincture*, *obstreperous*, and *egregious*.

Vocabulary is learned largely in context from reading books and listening to others. It is a rare individual who picks up vocabulary by reading the dictionary or memorizing word lists from the “Building Your Wordpower” section of popular magazines. In the main, a person’s vocabulary is a measure of sensitivity

to new information and the ability to decipher meanings based on the context in which words are encountered. Precisely because the acquisition of word meaning depends on contextual inference, the Vocabulary subtest turns out to be the single best measure of overall intelligence on the Wechsler scales (Gregory, 1999 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib650>)). This is a surprise to many laypersons who regard vocabulary as merely synonymous with educational exposure and, therefore, a mediocre index of general intelligence. However, there is simply no denying the empirical evidence: Vocabulary has among the highest subtest correlations with Full Scale IQ on both the WISC-IV and also the WAIS-IV.

Arithmetic

Except for the very easiest items for young people or persons who have mental retardation, the Arithmetic subtest consists of orally presented mathematics problems. The examinee must solve the problems without paper or pencil within a time limit (usually 30 to 60 seconds). The simple items stress fundamental operations of addition or subtraction, for example:

“If you have fifteen apples and give seven away, how many are left?”

The more difficult items require proper conceptualization of the problem and the application of two arithmetic operations, for example:

“John bought a stereo that was marked down 15 percent from the original sales price of \$600. How much did John pay for the stereo?”

Although the mathematical requirements of the Arithmetic items are not excessively demanding, the necessity of solving the problems mentally within a time limit makes this subtest quite challenging for most examinees. In addition to rudimentary arithmetic skills, successful performance on Arithmetic requires high levels of concentration and the ability to maintain intermediate calculations in short-term memory. In factor analyses of the WISC-IV and WAIS-IV, Arithmetic often loads on a third factor interpreted as Working Memory.

Comprehension

Found on all three Wechsler intelligence tests, the Comprehension subtest is an eclectic collection of items that require explanation rather than mere factual knowledge. The easy questions stress common sense, whereas the more difficult questions require an understanding of social and cultural conventions. On the WAIS-IV, several of the most difficult questions require the examinee to interpret proverbs.

An easy item on Comprehension is of the form “Why do people wear clothes?” Difficult items resemble the following:

- “What does this saying mean: ‘A bird in the hand is worth two in the bush.’”
- “Why are Supreme Court Judges appointed for life?”

Comprehension would appear to be, in part, a measure of “social intelligence” in that many items tap the examinee’s understanding of social and cultural conventions. Sipps, Berry, and Lynch (1987 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1502>)) found that Comprehension scores were moderately related to measures of social intelligence on the California Psychological Inventory. Of course, a high score signifies only that the examinee is knowledgeable about social and cultural conventions; choosing the right action may or may not flow from this knowledge. However, studies by Campbell and McCord (1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib260>)) and Lipsitz,

Dworkin, and Erlenmeyer-Kimling (1993

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib998>) provide no support for the commonly accepted clinical lore that Comprehension scores are sensitive to social functioning.

Similarities

In this subtest, the examinee is asked questions of the type, "In what way are shirts and socks alike?" The Similarities subtest evaluates the examinee's ability to distinguish important from unimportant resemblances in objects, facts, and ideas. Indirectly, these questions assess the assimilation of the concept of likeness. The examinee must also possess the ability to judge when a likeness is important rather than trivial. For example, "shirts" and "socks" are alike in that both begin with the letter *s*, but this is not the essential similarity between these two items. The important similarity is that shirts and socks are both exemplars of a concept, namely, "clothes." As this example illustrates, Similarities can be thought of as a test of verbal concept formation and is found on all three Wechsler intelligence tests.

Letter-Number Sequencing

The examiner orally presents a series of letters and numbers that are in random order. The examinee must reorder and repeat the list by saying the numbers in ascending order and then the letters in alphabetical order. For example, if the examiner says "R-3-B-5-Z-1-C," the examinee should respond "1-3-5-B-C-R-Z." This test measures attention, concentration, and freedom from distractibility. Together with Arithmetic and Digit Span, this subtest contributes to the Working Memory Index score on the WAIS-IV (see the following). Donders, Tulskey, and Zhu (2001

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib430>) found the Letter-Number Sequencing subtest to be highly sensitive to the effects of moderate and severe traumatic brain injury.

Picture Completion

For this subtest, the examiner asks the examinee to identify the "important part" that is missing from a picture. For example, a simple item might be of this type: a picture of a table with one leg missing. The items get harder and harder; testing continues until the examinee misses several in a row. **Figure 5.6** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05fig6>) depicts an item similar to those found on the WAIS-IV. The Picture Completion subtest presupposes that the examinee has been exposed to the object or situation represented. For this reason, Picture Completion may be inappropriate for culturally disadvantaged persons.

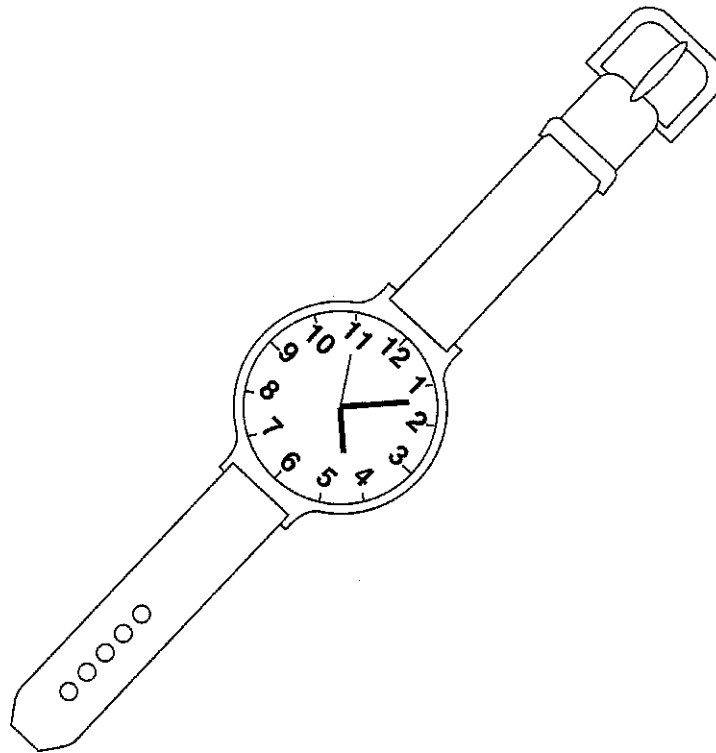


FIGURE 5.6 Picture Completion Item Similar to Those Found on the WAIS-IV

Picture Concepts

This subtest is found on the WPPSI-IV and the WISC-IV. For each item, the child is shown a card with two or three rows of pictures and instructed to choose one picture from each row to form a group with a common characteristic. This is a recent subtest designed to measure abstract, categorical reasoning. The 28 items reflect increasingly more difficult levels of abstraction. For example, for an easy item the commonality might be that a fruit is found in each row, whereas for a more difficult item the commonality might be that a device used for signaling (bell, flashlight, flags) is found in each row.

Block Design

On the Block Design subtest, the examinee must reproduce two-dimensional geometric designs by proper rotation and placement of three-dimensional colored blocks. For all of the Wechsler scales, the first few Block Design items can be solved through trial and error. However, the more difficult items require the analysis of spatial relations, visual-motor coordination, and the rigid application of logic. Block Design demands much more problem-solving and reasoning ability than most of the Performance subtests in which memory and prior experience are more heavily weighted.

Block Design is a strongly speeded test. Consider the WAIS-IV version, which consists of 14 designs of increasing difficulty. To obtain a high score on this subtest, adults must not only reproduce each of the designs correctly, but they must also earn bonus points on the last six designs by completing them quickly. An examinee who solves all the designs within the time limit but who fails to garner any bonus points will test out at just slightly above average on this subtest. Block Design scores may be misleading for examinees who do not value speeded performance.

Matrix Reasoning

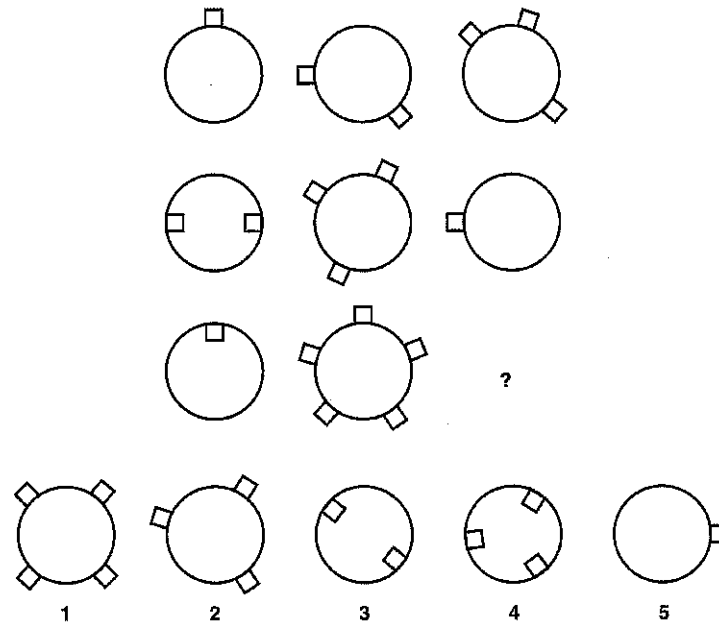


FIGURE 5.7 Matrix Reasoning Item Similar to Those Found on the WAIS-IV

Matrix Reasoning is included on all of the Wechsler intelligence tests. The subtest consists of figural reasoning problems arranged in increasing order of difficulty (**Figure 5.7** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05fig7>)). Finding the correct answer requires the examinee to identify a recurring pattern or relationship between figural stimuli drawn along a straight line (simple items) or in a 3×3 grid (hard items) in which the last item is missing. Based on nonverbal reasoning about the patterns and relationships, the examinee must infer the missing stimulus and select it from five choices provided at the bottom of the card.

Matrix Reasoning was designed to be a measure of fluid intelligence, which is the capacity to perform mental operations such as manipulation of abstract symbols. The items tap pattern completion, reasoning by analogy, and serial reasoning. Overall, the subtest is an excellent measure of inductive reasoning based on figural stimuli. Matrix Reasoning is not timed. Interestingly, Donders et al. (2001 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib430>)) report that the Matrix Reasoning subtest is relatively unaffected by moderate and severe traumatic brain injury.

Object Assembly

This subtest is found only on the WPPSI-III. For each item, the examinee must assemble the pieces of a jigsaw puzzle to form a common object (**Figure 5.8** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05fig8>)). The examiner does not identify the items, so the examinee must first discern the identity of each item from its disarranged parts. Success on this subtest requires high levels of perceptual organization; that is, the examinee must grasp a larger pattern or gestalt based on perception of the relationships among the individual parts.

Object Assembly is one of the least reliable of the Wechsler subtests. The modest reliability of Object Assembly may reflect, in part, the small number of items as well as the role of chance factors in solving jigsaw puzzles.

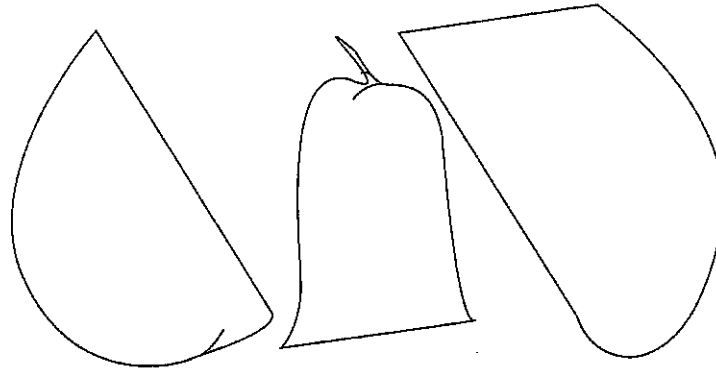


FIGURE 5.8 Object Assembly Item Similar to Those Found on the WPPSI-III

Coding

The WISC-IV version consists of two separate and distinct parts, one for examinees under age 8 (Coding A) and another for those 8 years of age and over (Coding B). In Coding A, the child must draw the correct symbol inside a series of randomly sequenced shapes. The task utilizes five shapes (star, circle, triangle, cross, and square), and each shape is assigned a unique symbol (vertical line, two horizontal lines, single horizontal line, circle, and two vertical lines, respectively). After a brief practice session, the child is told to draw the correct symbol inside 43 of the randomly sequenced shapes. However, since there is a two-minute time limit, high scores require rapid performance.

Coding B on the WISC-IV and Coding on the WAIS-IV are identical in format (**Figure 5.9** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05fig9>)). For both subtests, the examinee must associate one symbol with each of the digits 0 through 9 and quickly draw the appropriate symbol underneath a long series of random digits. The time limit for both versions is two minutes. Very few examinees manage to code all the stimuli in this amount of time.

1	2	3	4	5	6	7	8	9
⊖	⊥	⊥	=	∨	⊂	×	⊥	L
6	2	5	9	1	3	2	6	4

FIGURE 5.9 Digit Symbol Items Similar to Those Found on the WAIS-IV

Estes (1974 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib465>)) analyzed the Coding subtest from the standpoint of learning theory and concluded that efficient performance requires the ability to quickly produce distinctive verbal codes to represent each of the symbols in memory. For example, in **Figure 5.9**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05fig9>), the examinee might code the symbol underneath the number 2 as an “inverted T.” Verbal coding mediates quick performance by simplifying a difficult task. Efficient performance also demands immediate learning of the digit-symbol pairings so that the examinee need not look from each digit to the reference table to

determine the correct response. In this regard, Coding is unique: It is the only Wechsler subtest that necessitates on-the-spot learning of an unfamiliar task.

Coding scores show a steep decrement with advancing age. In cross-sectional studies, raw scores on Coding decline by as much as 50 percent from age 20 to age 70 (**Wechsler, 1981** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1732>)). The decrement is approximately linear and not easily explained by superficial references to motivational differences or motor slowing. Of course, cross-sectional results are not necessarily synonymous with longitudinal trends. However, the age decrement on Coding is so steep that it must indicate, in part, a real age change in the speed of basic information processing skills. Coding is one of the most sensitive subtests to the effects of organic impairment (**Donders et al., 2001** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib430>); **Lezak, 1995** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib981>)).

Symbol Search

This is a highly speeded subtest in which the examinee looks at a target group of symbols, then quickly examines a search group of symbols, and finally marks a “YES” or “NO” box to indicate whether one or more of the symbols in the target group occurred within the search group. A Symbol Search item is depicted in **Figure 5.10**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec14#ch05fig10>). This subtest would appear to be a measure of processing speed. Symbol Search is highly sensitive to the impact of traumatic brain injury (**Donders et al., 2001** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib430>)).

Cancellation

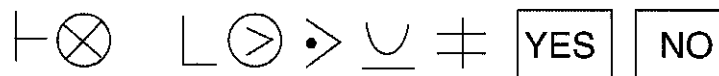


FIGURE 5.10 Symbol Search Item Similar to Those Found on the WISC-IV

Note: The examinee’s task is to determine whether either shape at the left occurs among the five shapes to the right.

On the WISC-IV, this is a timed subtest in which the child is instructed to draw a line through or “cancel” drawings of animals placed randomly among drawings of inanimate objects (e.g., umbrella, car, hydrant, lightbulb). For example, on a standard-sized sheet of paper, about 160 stimuli are pictured, including 30 animals (horse, bear, seal, fish, chicken). Cancellation consists of two trials: one with a random arrangement of visual stimuli, and one with clearly structured rows and columns of stimuli. In addition to a total subtest score, separate process scores for the random and the structured trials are available for comparison. This subtest is similar to existing cancellation tasks designed to measure processing speed, vigilance, and visual attention. It is well established that examinees with neuropsychological impairments perform poorly, especially on the random trial (e.g., **Bate, Mathias, & Crawford, 2001** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib106>); Geldmacher, 1996). On the WAIS-IV, Cancellation is somewhat more complex, involving *two* target stimuli consisting of geometric shapes. The examinee is told, for example, to cancel “red squares and yellow triangles” among an array of red and yellow squares and red and yellow triangles. A second trial involves stars and circles in orange and blue. This timed task (45 seconds per trial) is much more difficult than it seems.

Visual Puzzles

Visual Puzzles is found only on the WAIS-IV. The examinee is shown a picture of a completed shape such as a rectangle, and asked to select from six smaller shapes the three that could be used to assemble the larger completed shape. Successful performance requires visual-spatial analysis and the mental rotation of shapes. According to the WAIS-IV Technical Manual, this subtest taps for “visual perception, broad visual intelligence, fluid intelligence, simultaneous processing, spatial visualization and manipulation, and the ability to anticipate relationships among parts (Wechsler, 2008b, p. 14). The 26 items have strict time limits of 20 seconds for the initial easy items and 30 seconds for the remaining items. Visual Puzzles is a core subtest that contributes to the Perceptual Reasoning Index of the WAIS-IV.

Figure Weights

Figure Weights is found only on the WAIS-IV. It is a supplemental subtest that contributes to the Perceptual Reasoning Index. The examinee is shown a picture of an old-fashioned fulcrum scale that is missing weight(s) on one side. The task is to select from six options the response that would bring the scale into balance. This subtest is a measure of quantitative and analogical reasoning; inductive and deductive logic are essential for success. Easy items provide a time limit of 20 seconds, hard items allow 40 seconds.

5.15 WECHSLER ADULT INTELLIGENCE SCALE-IV

The WAIS-IV is a significant revision of the WAIS-III, even though many of the previous items were retained (Wechsler, 2008

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1737>)). The most significant changes include the addition of two subtests, a simplified test structure, and an emphasis on index scores that provide a sharper demarcation of discrete domains of cognitive functions. In addition, the WAIS-IV abandons the familiar (but psychometrically indefensible) bifurcation of intelligence into Verbal IQ and Performance IQ, preferring instead the fourfold breakdown discussed below. In addition to traditional approaches to scoring the WAIS-IV subtests, the new edition also provides neuropsychologically relevant process scores for four of the subtests. These scores are useful mainly for advanced forms of test interpretation in the context of a comprehensive test battery. We do not discuss process scores in this section. Because of improvements in the WAIS-IV protocol forms (e.g., prominent display of discontinue rules), this test is somewhat easier to administer than its predecessor.

Lichtenberger and Kaufman (2009

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib986>)) provide an outstanding overview of the WAIS-IV in clinical practice.

The WAIS-IV is comprised of 15 subtests, but only 10 of the subtests, known as core subtests, are needed to obtain the traditional IQ score and the component index scores. The other five subtests are deemed supplemental. These are often used to provide additional clinical information; in specific instances, supplemental subtests may be used as acceptable substitutes for core subtests.

In addition to the traditional Full Scale IQ score, normed to a mean of 100 and standard deviation of 15, the WAIS-IV is scored for four index scores, each based on 2 or 3 of the 10 core subtests. These are derived from factor analysis of the subtests, which revealed four domains: Verbal Comprehension, Perceptual Reasoning, Working Memory, and Processing Speed. The index scores are also based on the familiar mean of 100 and standard deviation of 15. The breakdown of subtests for the four index scores is as follows:

Verbal Comprehension Index

- Similarities
- Vocabulary
- Information

Perceptual Reasoning Index

- Block Design
- Matrix Reasoning
- Visual Puzzles

Working Memory Index

- Digit Span
- Arithmetic

Processing Speed Index

- Symbol Search
- Coding

The Verbal Comprehension Index (VCI) is similar to the outdated notion (used on the WAIS-III) of Verbal IQ or VIQ. However, from a psychometric standpoint, VCI is a cleaner and more direct measure of verbal comprehension than VIQ, hence it is now the preferred index. Likewise, the Perceptual Reasoning Index

(PRI) is similar to the former notion (from the WAIS-III) of Performance IQ or PIQ. Yet, as a more refined measure of perceptual reasoning, PRI is therefore the preferred index. Put simply, VCI and PRI fit the factor analytic data better. Long-held conventions tend to persist, but it is time to let the outdated notions of Verbal IQ and Performance IQ fade into oblivion.

The Working Memory Index (WMI) is comprised of subtests sensitive to attention and immediate memory (Digit Span and Arithmetic). A relatively low score on this index may signify that the examinee has an attentional or memory problem, especially with orally presented materials. The Processing Speed Index (PSI) comprises subtest that require the *highly* speeded process of visual information (Symbol Search and Coding). The PSI is sensitive to a wide variety of neurological and neuropsychological conditions (Tulsky, Zhu, & Ledbetter, 1997).

WAIS-IV Standardization

The standardization of the WAIS-IV was undertaken with great care and based on data gathered by the U.S. Bureau of the Census in 2005. The total sample of 2,200 adults (ages 16 to 91) was carefully stratified on these variables: gender, race/ethnicity, education level, and geographic region. Census figures from 2005 were used as the target values for the stratification variables. For example, of persons in the 55- to 64-year-old range, the Census Bureau found that 3.35 percent are African Americans with high school education. In like manner, 3.00 percent of the standardization participants were African Americans with high school education.

The standardization sample was divided into 13 age bands: 16–17, 18–19, 20–24, 25–29, 30–34, 35–44, 45–54, 55–64, 65–69, 70–74, 75–79, 80–84, 85–90. Except for the four oldest age groups, each sample included 200 participants carefully stratified on the demographic variables noted earlier; the last four age groups included 100 participants each. The resulting sample bears a very close correspondence to the U.S. Census proportions. However, persons suspected of even mild cognitive impairment were excluded, so that the standardizations sample likely is healthier than its census counterparts. Specifically, several exclusionary criteria were used in the standardization sample, including: uncorrected visual or hearing impairment, current hospitalization, evidence of drug/alcohol problems, upper extremity impairment, use of certain prescription drugs such as anticonvulsants, and a variety of potentially brain-impairing conditions (e.g., head injury, stroke, epilepsy, dementia, and mood disorder). Uncooperative participants and those for whom English was a second language also were excluded. In sum, the standardization sample was restricted to cooperative, reasonably healthy, English-speaking individuals who did not manifest significant brain-impairing conditions.

Although the WAIS-IV is similar to the WAIS-III and has a substantial item overlap, the two tests do not yield analogous IQs. In counterbalanced studies comparing scores of 240 adults on the two tests, WAIS-IV IQ scores are lower by 3 points. In sum, the WAIS-IV is a harder test than the WAIS-III. There is a troubling enigma here: Why does the normative sample for the WAIS-IV appear to be smarter than the normative sample for the WAIS-III? We take up this point in more detail in **Topic 6B**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch06lev1sec6#ch06lev2sec21>), Test Bias and Other Controversies.

Reliability

The reliability of the WAIS-IV is exceptionally good. Composite split-half reliabilities averaged across all age groups for the Index scores and IQ are: VCI .96, PCI .95, WMI .94, PSI .90, and Full Scale IQ 98. Further supporting the reliability of the WAIS-IV, reliability estimates for subtest scores of special groups (e.g., persons with intellectual disability, probable Alzheimer's disease, traumatic brain injury, major depression, autism) are equal to or higher than reliability estimates found in the general population

(Wechsler, 2008b). This suggests that the WAIS-IV is a reliable tool not just with the general population but also with the special populations who are more likely to be the focus of assessment.

For Full Scale IQ, the standard error of measurement is 2.6 points for the youngest examinees (ages 16 and 17), but even smaller at 2.1 points for all other age groups. Consider what this means: 95 percent of the time, an examinee's *true* Full Scale IQ will be within ± 4 points (2 standard errors of measure) of the obtained value. In common parlance, psychometrists would say that WAIS-IV IQ has an 8-point band of error, that is, IQ scores are accurate within about ± 4 points. In contrast to the strong reliabilities found for IQ and Index scores, the reliabilities of the 15 individual subtests are generally much weaker. The only subtests with stability coefficients in excess of .90 are Information (.90) and Vocabulary (.91). For the remaining subtests, reliability values range from the low .70s to the mid .80s. The most important implication of these weaker reliability findings is that examiners should approach subtest profile analysis with extreme caution. Subtest scores that appear discrepantly high (or low) for an individual examinee might be a consequence of the generally weak reliability of certain subtests rather than indicating true cognitive strengths or weaknesses. Some reviewers conclude that profile analysis (the identification of specific cognitive strengths and weaknesses based on analysis of peaks and valleys in the subtest scores) is not justified by the evidence.

Validity

The developers of the WAIS-IV provide a number of different lines of evidence to support the validity of this instrument (Wechsler, 2008b). Good content validity was built in from the beginning through comprehensive literature review and consultation with experts to assure that items and subtests tap the relevant range of cognitive processes. Good criterion-related validity was demonstrated in several studies correlating the WAIS-IV with mainstream intelligence tests and other measures. For example, WAIS-IV Full Scale IQ correlates strongly with global scores on other mainstream measures: .94 with the WAIS-III, .91 with the WISC-IV (for 16-year-olds in the overlapping age group), and .88 with the Wechsler Individual Achievement Test-II. The WAIS-IV also reveals appropriate convergent and discriminant validity in the patternings of strong and weak correlations with a wide variety of other instruments, including measures of attention deficit disorder, executive functions, and memory. As a generalization, correlations are appropriately strong among similar subtests and constructs from the WAIS-IV and other tests, and appropriately weak among dissimilar subtests and constructs.

Studies with special groups also provide theory-confirming results that speak to the validity of the WAIS-IV. The multiplicity of these studies is such that we can only provide a few examples here. Specifically, when 41 young adults with diagnosed Mathematics Disorder were compared to matched controls on WAIS-IV subtests, the most substantial difference by far was found on the Arithmetic subtest, where the clinical group averaged 6.6 compared to 8.8 for the matched controls (a subtest score of 10 is average in the general population). This corroborates the sensitivity of the instrument to the elements of one specific learning disability. In like manner, when 22 individuals with a history of moderate or severe brain injury were compared to matched controls, the largest difference among the four index scores was found on the Processing Speed Index (mean of 80.5 versus mean of 97.6), whereas the smallest difference among the four index scores was found on the Verbal Comprehension Index (mean of 92.1 versus mean of 100.8). These findings are exactly what would be predicted from a wide body of research on the impact of traumatic brain injury (e.g., **Lezak, Howieson, & Loring, 2004** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib984>)).

The construct validity of the WAIS-IV is also supported by confirmatory factor analyses of the subtest scores from the standardization sample, as detailed in the technical manual (Wechsler, 2008b). These complex analyses were designed to determine if the relations among observed subtest scores support

the existence of the hypothesized factors of intelligence measured by the four index scores of VCI, PRI, WMI, and PSI. The goodness-of-fit of the four factor hierarchical model of intelligence (Full Scale IQ at the top, sitting above the four index scores, each sitting above two or three constituent subtest scores) turns out to be exceptionally strong, although difficult to summarize in visual form. A simple way to depict the strong confirmatory fit is through a 4×10 table that shows the correlations among the four index scores and the 10 core subtest scores (**Table 5.7**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec15#ch05tab7>). Where appropriate, these correlations are corrected for overlap between the subtest scores and the index scores. For example, Similarities is a component of VCI, so the simple correlation between these two variables is artificially inflated. The values shown in **Table 5.7**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec15#ch05tab7>) are corrected for this kind of overlap. The reader will notice that with only a single exception, the subtests that compose each index score reveal their highest correlations with that index score. The only exception is the Arithmetic subtest, which is factorially more complex than other subtests, showing an almost identical relationship with VCI, PRI, and WMI.

Finally, the validity of the WAIS-IV is also buttressed by its strong overlap with the previous three editions of the test, for which there is an impressive array of validity data. For a full review of these findings the reader can consult Matarazzo (1972

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1058>)) and Kaufman (1990 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib862>)).

TABLE 5.7 Correlations Among WAIS-IV Subtests and Index Scores

	VCI	PRI	WMI	PSI
Verbal comprehension Subtests				
Similarities	74	57	57	42
Vocabulary	81	55	60	41
Information	63	54	56	37
Perceptual Reasoning Subtests				
Block Design	51	67	53	45
Matrix Reasoning	56	59	55	46
Visual Puzzles	48	66	49	41
Working Memory Subtests				
Digit Span	53	52	60	47
Arithmetic	63	59	60	44
Processing Speed Subtests				
Symbol Search	38	47	43	65
Coding	43	48	49	65

Source: Based on data in Wechsler, D. (2008). *WAIS-IV technical and interpretive manual*. San Antonio, TX: Pearson.

Note: Decimals have been omitted. Where appropriate, these correlations are corrected for overlap. For example, because Similarities is a component of VCI, the simple uncorrected correlation between these two variables would be artificially inflated. The values above are corrected for any componential overlap between subtests and index scores.

5.16 WECHSLER INTELLIGENCE SCALE FOR CHILDREN-IV

The Wechsler Intelligence Scale for Children (WISC) was published in 1949 as a downward extension of the original Wechsler-Bellevue. Although used widely in the next two decades, psychometricians perceived a number of flaws in the WISC: absence of nonwhites in the standardization sample, ambiguities of scoring, inappropriate items for children (e.g., reference to “cigars”), and absence of females and African Americans in the pictorial content of items. The WISC-R, WISC-III, and WISC-IV corrected these flaws.

The WISC-IV consists of 15 subtests, 10 of which are designated as core subtests used in the computation of composite scores and Full Scale IQ and five of which are designated as supplemental:

Core Subtests

Block Design	Vocabulary
Similarities	Letter–Number Sequencing
Digit Span	Matrix Reasoning
Picture Concepts	Comprehension
Coding	Symbol Search

Supplemental Subtests

Picture Completion
Cancellation
Information
Arithmetic
Word Reasoning

Although the supplemental subtests are not required for the computation of Full Scale IQ and composite scores (discussed later), careful examiners nonetheless may choose to administer them because of the important diagnostic information they often provide. For example, the Cancellation subtest is supplemental but affords important information about vigilance and visual attention; hence, many examiners use it. The Arithmetic subtest also is supplemental but often chosen by examiners because it is helpful in the assessment of auditory attention (the questions are presented orally).

Another function of the supplementary subtests is suitable substitution for a core subtest. In well-defined circumstances, an examiner may elect to give a supplemental subtest in place of a core subtest. For example, when testing a child with fine motor problems—such as might be observed in a child with cerebral palsy—an examiner would be well advised to use Cancellation in place of Coding, and Picture Completion in place of Block Design. Both of these supplementary tests (Cancellation and Picture Completion) are relatively unaffected by fine motor difficulties. In contrast, the core subtests (Coding and Block Design) would be severely impacted by fine motor difficulties and, therefore, could yield unfair assessments of cognitive functioning. Substitutions also are allowed when a core subtest accidentally is invalidated. However, an examiner may not elect to substitute a supplemental subtest merely because a child has performed poorly on a core subtest.

The standardization of the WISC-IV is first class, based on 100 boys and girls at each year of age from 6½ through 16½ (total $N = 2,200$). These cases were carefully selected and stratified on the basis of the

2000 U.S. Census with respect to gender, race/ethnicity (white, African American, Hispanic, and Asian), geographic region, and parent educational level. A desirable feature of the standardization sample is that 5.7 percent of the sample consisted of children with defined characteristics such as giftedness, learning disability, expressive language disorder, head injury, autism, and motor impairment. The purpose of adding these children was to ensure that the normative sample accurately represented the population of children attending school. The correspondence between the standardization sample and the U.S. Census data on essential stratification variables was nearly perfect (**Wechsler, 2003** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1736>) , p. 40).

The reliability of the WISC-IV is strong and comparable to previous editions of the test. For example, the IQ and composite scores show split-half and test-retest reliabilities in the .90s, whereas the individual subtests possess somewhat lower reliability coefficients, ranging from .79 (Cancellation and Symbol Search) to .90 (Letter-Number Sequencing). Most reliabilities are in the high .80s, for example, Block Design and Similarities at .86, and Vocabulary and Matrix Reasoning at .89. Test-retest reliabilities tend to be slightly lower.

The validity of the WISC-IV rests, in part, on its overlap with the WISC-III, for which dozens of supportive studies could be cited. We do not want to overwhelm with excessive detail, so we refer the interested reader to Sattler (**2001**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1438>)) for a good review of earlier studies. The WISC-IV manual cites an impressive array of validity studies, which we summarize here. First, we discuss correlations of WISC-IV test scores with its predecessor and with other Wechsler intelligence tests. The preliminary findings indicate strong correlations with comparable WISC-III subtests, most in the high .70s or low .80s. The correlation for Full Scale IQ is much higher, $r = .89$. Likewise, correlations with the WPPSI-III are strong for comparable subtests, and, again, exceptionally strong for Full Scale IQ, $r = .89$. A similar pattern is found with 16-year-old examinees, who can be tested legitimately with both the WISC-IV and the WAIS-III. In a sample of 198 children tested in counterbalanced order over a period of about three weeks, correlations were strong for comparable subtests and exceptionally strong for composite and Full Scale IQ scores ($r = .89$). Overall, these are remarkable correlations, nearly as strong as the reliabilities of the respective scales would allow. An interesting finding is that WISC-IV IQs are an average of 2.5 points lower than WISC-III IQs and 3 points lower than WAIS-III IQs. This is a consistent finding in the history of individual intelligence tests; namely, newer tests almost invariably yield lower Q scores in comparison to older tests. We discuss this intriguing result, called the Flynn effect, in the next chapter.

Factor-analytic studies of the standardization sample provided additional evidence for the utility of the WISC-IV in the diagnostic assessment of children. The results of numerous factor analyses, including separate analyses for four age groups (6-7, 8-10, 11-13, 14-16) strongly confirmed a four-factor solution that was used to define the composite scores, called Index scores, for the test (**Wechsler, 2003** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1736>)). The factors and the core subtests assigned to them were as follows:

Verbal Comprehension Index

Similarities

Vocabulary

Comprehension

Perceptual Reasoning Index

Block Design

Picture Concepts

Matrix Reasoning
Working Memory Index
Digit Span
Letter–Number Sequencing
Processing Speed Index
Coding
Symbol Search

The four Index scores are based on the familiar mean of 100 and standard deviation of 15. Thus, the WISC-IV provides substantial detail about the nuances of intellectual functioning—up to 15 subtest scores, four Index scores, and the Full Scale IQ. The robust findings of the four-factor solution to the WISC-IV provided the rationale for abandoning Wechsler’s original two-factor division of Verbal IQ and Performance IQ. In fact, there is no longer any method on the WISC-IV to obtain a Verbal IQ or a Performance IQ—precisely because these partitions no longer fit with the emerging consensus about the nature of intelligence.

The WISC-IV also revealed theory-confirming correlations with a variety of cognitive, ability, and achievement tests (**Wechsler, 2003** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1736>)). In general, correlations with other measures were appropriately high for similar constructs and predictably low for dissimilar constructs—these are the prerequisites for convergent validity and discriminant validity, respectively. For example, in a sample of 550 children aged 6–16, reading achievement subtest scores from the Wechsler Individual Achievement Test-II correlated more strongly with Verbal Comprehension Index scores from the WISC-IV than with the other Index scores. Likewise, in a sample of 126 children aged 6–16, the Attention/Concentration subtest from the Children’s Memory Scale (**Cohen, 1997** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib318>)) correlated substantially ($r = .74$) with Working Memory Index scores from the WISC-IV but less robustly with the other Index scores. These and other findings indicate general support for the convergent validity of the WISC-IV Index scores. Discriminant validity was confirmed by the negligible relationships among WISC-IV Index scores and measures of emotional intelligence from the BarOn Emotional Quotient Inventory (**BarOn EQI, Bar-On & Parker, 2000** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib94>)). For the most part, research has shown that emotional intelligence is independent of cognitive intelligence. Thus, relationships among Index scores from the WISC-IV and subtest scores from the BarOn EQI should bear out as insignificant. In fact, the correlations were negligible, in the range of .06 to .20. The only exceptions were sensible ones. For example, scores on the Adaptability subscale from the BarOn EQI correlated .34 with WISC-IV Full Scale IQ. Certainly, it is plausible that adaptability as measured by the BarOn EQI is rooted, in part, in a foundation of cognitive skills, as mirrored in IQ, thus illuminating the modest correlation between these two measures.

5.17 STANFORD-BINET INTELLIGENCE SCALES: FIFTH EDITION

With a lineage that goes back to the Binet-Simon scale of 1905, the Stanford-Binet: Fifth Edition (SB5) has the oldest and perhaps the most prestigious pedigree of any individual intelligence test. In **Table 5.8** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec17#ch05tab8>), we outline some important milestones in the development of the SB5 and its predecessors. Released in 2003, the SB5 is a very new test (**Roid, 2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1398>), 2003 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1399>)). For this reason, evaluation of this instrument is based, in part, on its resemblance in content and subtests to the SB4, for which a large body of independent research literature has been amassed.

The SB5 Model of Intelligence

In early editions of the Stanford-Binet, the examiner obtained only a composite IQ. Although the pattern of right and wrong answers could be analyzed qualitatively, the earlier Stanford-Binet tests (prior to the fourth edition) did not provide a basis for quantitative analysis of the subcomponents of the entire scale. The fourth and fifth editions corrected this shortcoming.

TABLE 5.8 Milestones in the Development of the Stanford-Binet and Predecessor Tests

<i>Year</i>	<i>Test/Authors</i>	<i>Comment</i>
1905	Binet and Simon	Simple 30-item test
1908	Binet and Simon	Introduced the mental age concept
1911	Binet and Simon	Expanded to include adults
1916	Stanford-Binet Terman and Merrill	Introduced the IQ concept
1937	Stanford-Binet-2 Terman and Merrill	First use of parallel forms (L and M)
1960	Stanford-Binet-3 Terman and Merrill	Modern item-analysis methods used
1972	Stanford-Binet-3 Terman and Merrill	SB-3 restandardized on 2,100 persons
1986	Stanford-Binet-4 Thorndike, Hagen, and Sattler	Complete restructuring into 15 subsets
2003	Stanford-Binet-5 Roid	Five factors of intelligence

The organization of the SB5 was guided by the principle that each of five factors of intelligence can be assessed in two distinct domains—nonverbal and verbal. The five factors—derived from modern cognitive theories such as Carroll (**1993** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib273>)) and Baddeley (**1986** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib73>))—are fluid reasoning, knowledge, quantitative reasoning, visual-spatial processing, and working memory. When these five factors of intelligence are “crossed” with the two domains (nonverbal and verbal), the result is an instrument with 10 subtests (**Figure 5.11**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec17#ch05fig11>)). Thus, the SB5 provides a number of different perspectives on the cognitive functioning of an examinee: 10 subtest scores (mean of 10, SD of 3), three IQ scores (the familiar Full Scale IQ, Verbal IQ, and Nonverbal IQ), as well as five factor scores (Fluid Reasoning, Knowledge, Quantitative Reasoning, Visual-Spatial Processing, and Working Memory). The IQ and factor scores are normed to a mean of 100 and SD of 15.

Routing Procedure and Tailored Testing

The SB5 maintains the historical tradition of this instrument by using a **routing procedure** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss283>) to estimate the general cognitive ability of the examinee before proceeding to the remainder of the test. The purpose of the routing procedure is to identify the appropriate starting points for subsequent subtests. The routing items are both nonverbal (object series and matrices) and verbal (vocabulary). These items also provide the Abbreviated IQ, sometimes used for screening purposes. Roid (2002) (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1398>) describes the advantages of using a routing procedure:

This tailored approach to assessment provides greater richness of factor measurement within a shorter, efficient test administration. The use of modern item response theory in the design of SB5 allows for greater precision of measurement due to the adaption of the test to the functional level of the examinee in an efficient time frame.

		DOMAINS	
		Nonverbal	Verbal
FACTORS	Fluid Reasoning	Nonverbal Fluid Reasoning	Verbal Fluid Reasoning
	Knowledge	Nonverbal Knowledge	Verbal Knowledge
	Quantitative Reasoning	Nonverbal Quantitative Reasoning	Verbal Quantitative Reasoning
	Visual-Spatial Reasoning	Nonverbal Visual-Spatial Processing	Verbal Visual-Spatial Processing
	Working Memory	Nonverbal Working Memory	Verbal Working Memory
		Nonverbal IQ	Verbal IQ
		FULL SCALE IQ	

FIGURE 5.11 Structure of the Stanford-Binet: Fifth Edition

Thus, the purpose of the routing procedure is not just to reduce the number of items administered (and, therefore, save time), but to do so without loss of measurement precision. This is possible because the SB5 was constructed according to the principles of item response theory (Embretson, 1996) (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib458>)). When a test is constructed within the framework of item response theory, item difficulty levels and other parameters are precisely calibrated during the development phase.

Special Features of the SB5

In addition to providing a more familiar partition of intelligence into Full Scale IQ, Verbal IQ, and Nonverbal IQ, the SB5 also features a number of other improvements over its predecessor, the SB4. The

test now includes extensive high-end items, designed to assess the highest level of gifted performance. Many of these items are updates from very early editions of the Stanford-Binet, when the instrument was renowned for its very high ceiling. At the other extreme, improved low-end items provide better assessment for very young children (as young as age 2) and adults with mental retardation. In addition, the items and subtests that contribute to the Nonverbal IQ do not require expressive language, which makes this part of the test ideal for assessing individuals with limited English, deafness, or communication disorders. The developers of the SB5 also screened test items for fairness based on religious as well as traditional concerns. Expert panels examined the entire test on fairness issues related to the standard variables (gender, race, ethnicity, and disability) and religious tradition (Christian, Jewish, Muslim, Hindu, and Buddhist backgrounds). This is the first time in the history of intelligence testing that religious tradition has been considered in test development. Finally, the Working Memory factor, consisting of both verbal and nonverbal subtests, shows promise in helping to assess and understand children with attention-deficit/hyperactivity disorder.

Standardization and Psychometric Properties of the SB5

The SB5 is suitable for children age 2 through adults age 85 and older, and the standardization sample consists of 4,800 individuals stratified by gender, ethnic, regional, and educational levels in the United States, based on the year 2000 census. In part because item selection was determined by modern item response theory, the reliability of subtests, indices, and IQ scores is very strong and comparable to other mainstream individual intelligence tests. For example, the Verbal IQ, Nonverbal IQ, and Full Scale IQ each have reliabilities in the .90s, and the individual subtests are in the range of .70 to .85 (**Roid, 2002** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1398>)).

As is typical in the release of a new test, the manual for the SB5 (**Roid, 2003** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1399>)) reports on numerous affirming correlational studies (e.g., with the Wechsler scales, the SB4, the UNIT) that provide strong support for criterion-related validity. The validity of the test as a measure of general intelligence is also supported by its resemblance to the SB4, about which a large body of research can be cited. For example, Lamp and Krohn (**2001** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib946>)) studied the longitudinal predictive validity of the SB4 in a sample of 89 Head Start children (39 African American and 50 white) from impoverished backgrounds who ranged in age from about 4 to 6½. These children were retested several times over an eight-year period on both the SB4 and the Metropolitan Achievement Test. The correlations between the initial SB4 score and the subsequent achievement scores were very strong (mainly in the .50s), and the test was equally good at predicting outcome for African American and white children. In another study (**Atkinson, Bevc, Dickens, & Blackwell, 1992** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib68>)), the concurrent validity of the SB4 was tested against the Leiter International Performance Scale and the Vineland Adaptive Behavior Scales in a sample of 24 children with developmental delays. The correlations were very robust (.78 and .70, respectively). These and many other studies strongly support the validity of the SB4 as a measure of general intelligence. As new research is reported on the SB5, it is likely that this recent edition also will prove to be highly valid and even more useful than its predecessor as a measure of intelligence.

In summary, the SB5 is a very promising new test that is especially useful at both ends of the cognitive spectrum—the very young or those with developmental delays, and very gifted persons. Based on the care with which the instrument was constructed, the test is likely to become a mainstay of individual intelligence testing in a wide variety of settings.

5.18 DETROIT TESTS OF LEARNING APTITUDE-4

The Detroit Tests of Learning Aptitude-4 (**DTLA-4; Hammill, 1999**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib690>) is a recent revision of an instrument first published in 1935. The test is individually administered and designed for schoolchildren from 6 through 17 years of age. The DTLA-4 consists of 10 subtests that form the basis for computing 16 composites, including general intelligence, optimal level, and 14 ability areas. The subtests are largely within the Binet-Wechsler tradition, although there are a few surprises such as the inclusion of Story Construction, a measure of storytelling ability (**Table 5.9** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec18#ch05tab9>)).

The General Mental Ability composite is formed by combining standard scores for all 10 subtests in the battery. The Optimal Level composite is based on the highest four standard scores earned by the examinee and is thought to represent how well the examinee might perform under optimal circumstances. Each of the remaining 14 composite scores is derived from a combination of several subtests thought to measure a common attribute. For example, subtests that involve knowledge of words and their use are combined to form the Verbal Composite, whereas subtests that do not involve reading, writing, or speech comprise the Nonverbal Composite. Several of the composite scores are designed to represent major constructs within contemporary theories of intelligence. In addition to the General Mental Ability composite and the Optimal Level composite, the remaining 14 DTLA-4 composite scores are as follows:

Verbal	Nonverbal	(Linguistic)
Attention-enhanced	Attention-reduced	(Attentional)
Motor-enhanced	Motor-reduced	(Motoric)
Fluid	Crystallized	(Horn & Cattell)
Simultaneous	Successive	(Das)
Associative	Cognitive	(Jensen)
Verbal	Performance	(Wechsler)

The 16 composite scores are based on the familiar mean of 100 and standard deviation of 15. The 10 subtests are normed for a mean of 10 and standard deviation of 3.

The composites were designed to offer contrasting assessments such that a difference between scores may be of diagnostic significance. For example, an examinee who scored well on Attention-Reduced aptitude but poorly on Attention-Enhanced aptitude (in the Attentional domain) presumably experiences difficulty with immediate recall, short-term memory, or focused concentration.

The DTLA-4 was standardized on 1,350 students whose backgrounds closely matched census data for sex, race, urban/rural residence, family income, educational attainment of parents, and geographic area. The reliability of this instrument is similar to other individual tests of intelligence, with internal consistency coefficients generally exceeding .80 for the subtests and .90 for the composites, and test-retest coefficients for the subtests and the composites in the .80s and .90s. Criterion-related validity is well established through correlational studies with other mainstream instruments such as the WISC-III, K-ABC, and Woodcock-Johnson.

TABLE 5.9 Brief Description of the DTLA-4 Subtests

<i>Subtest</i>	<i>Task</i>
Word Opposites	Provide antonyms—word opposites.
Design Sequences	Discriminate and remember nonsensical graphic material.
Sentence Imitation	Repeat orally presented sentences.
Reversed Letters	Short-term visual memory and attention.
Story Construction	Create a logical story from several pictures.
Design Reproduction	Copy designs from memory.
Basic Information	Knowledge of everyday facts and information.
Symbolic Relations	Select from a series of designs the part that was missing from a previous design.
Word Sequences	Repeat a series of unrelated words.
Story Sequences	Organize pictorial material into meaningful sequences.

A concern with the DTLA-4 is that the conceptual breakdown into composites is not sufficiently supported by empirical evidence. For example, while it may be true that the Simultaneous composite does measure the simultaneous cognitive processes proposed by Das, Kirby, and Jarman (1979 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib391>)), there is scant empirical support to buttress this claim. Another problem with this instrument is that there are more composites than there are subtests! Inevitably, the composites will be highly intercorrelated, because each subtest occurs in several composites. In sum, DTLA-4 may be a good measure of general intelligence, but the use of composite scores for purposes of psychoeducational planning requires additional empirical study. Smith (2001 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1511>))) provides a thorough review of the DTLA-4.

5.19 THE COGNITIVE ASSESSMENT SYSTEM-II

The Cognitive Assessment System-II (CAS-II) is an individually administered test of cognitive abilities designed for children and adolescents ages 5 through 17 (Naglieri, Das, & Goldstein, 2012 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1209>)). The CAS-II was explicitly constructed to embody the Planning, Attention, Simultaneous, and Successive (PASS) theory of intelligence discussed at the beginning of the chapter (Das, Kirby, & Jarman, 1979 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib391>)); Das, Naglieri, & Kirby, 1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib388>)). The Standard Battery consists of 12 subtests and takes about 60 minutes to complete (Figure 5.12 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec19#ch05fig12>)). A shorter version of eight subtests is available, but most practitioners recommend the full battery because it provides a better picture for diagnosis and intervention.

The CAS-II provides a standard score (mean of 100, *SD* of 15) for each of the four process scales (Planning, Attention, Simultaneous, and Successive), as well as a Full Scale standard score. The 12 subtests are normed to a mean of 10 and *SD* of 3. The Planning Scale is a measure of the ability to develop strategies for task completion. For example, in the *Matched Numbers* subtest, the child views rows of six numbers and is instructed to underline the two numbers in each row that are identical. The numbers increase in length from one digit to seven digits. The subtest score is based on a combination of time to completion and number correct. In the *Planned Codes* subtest, the task is to learn a code depicted at the top of the page (such as A goes with X-O, B goes with O-O, C goes with X-X, D goes with O-X) and then fill in missing codes in the remainder of the page (for example, A _ _ , C _ _ , B _ _ , A _ _ , D _ _ , etc.). In the *Planned Connections* subtest (a variation of the Trail Making Test, part B, Reitan & Wolfson, 1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1350>))), the child draws a pencil line to connect randomly placed numbers and letters in sequential order, alternating between numbers and letters (1-A-2-B-3-C, etc.). The Planning subtests involve cognitive control and self-regulation.

<i>Scale</i>	<i>Subtests</i>
Planning	Matched Numbers
	Planned Codes
	Planned Connections
Attention	Expressive Attention
	Number Detection
	Receptive Attention
Simultaneous	Nonverbal Matrices
	Verbal Spatial Relations
	Figure Memory
Successive	Word Series
	Sentence Repetition
	Sentence Questions or
	Speech Rates (ages 5–7)

FIGURE 5.12 Cognitive Assessment System-II Scales and Subtests

The Attention Scale is a measure of the mental processes involved in resistance to distraction and focused attention over time. For example, in the *Expressive Attention* subtest, a variation of the Stroop procedure (**Stroop, 1935**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1595>), the child first reads a long list of color words (Blue, Yellow, Red, Green) repeated in random order, then quickly names blocks of color printed in these four colors. These tasks are preamble to the final task, the only part that is scored. In the final section of the Expressive Attention subtest, a lengthy list of the color words (Blue, Yellow, Red, Green) is presented, *each word printed in a competing color* (e.g., the word *Blue* printed in *red* ink), with instructions to name the colors, not read the words. The raw score is the ratio of the total number correct to the time needed for completion of the last section. In the *Number Detection* subtest, the child is required to underline specific digits in particular fonts, for example, the task might be to detect the numbers 1, 2, and 3 among random digits, but only when printed in bold font. In the *Receptive Attention* subtest, the child first underlines letter pairs that are physically the same (e.g., TT but not Tt) and then underlines letter pairs that are the same name (e.g., Bb but not Ba). The score is based on accuracy and total time.

The Simultaneous Scale is a measure of the ability to organize information into coherent wholes. Both nonverbal and verbal processes are utilized to analyze and synthesize spatial and verbal relationships. *Nonverbal Matrices* is a variation on the familiar matrix reasoning task first employed in the Raven Progressive Matrices (**Raven, 1938**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1335>) and found in many intelligence tests. A 3 × 3 matrix of geometric shapes is shown, with a missing shape in the lower right-hand corner. Below the matrix are six shapes, one of which completes the rules of progression in

the matrix from left to right and top to bottom. Based on inference, the task is to choose the correct shape. In the *Verbal Spatial Relations* subtest, the child views six drawings, each depicting a particular spatial relationship between shapes, and then encounters a series of printed question such as *Show me the square to the right of the circle*. The task is to choose the one drawing among six that depicts the relationship. In the *Figure Memory* subtest, the child views a two- or three-dimensional drawing for five seconds, and then must correctly locate the original drawing embedded within a larger, more complex drawing. The Simultaneous subtests involve the perception of stimuli as a whole, in contrast to what is needed in successive processing.

The Successive Scale involves mental processes needed to remember and complete a task in a specific order or sequence. In *Word Series*, the task is to recall in correct order a series of two to nine words orally presented at one word per second. This task is similar to measures of digit span, except words are used instead of digits. The same nine words (one-syllable, high-frequency words such as Car, Dog, Shoe) are used. In the *Sentence Repetition* subtest, the child reads 20 sentences aloud, one by one. After each sentence is read, the child is asked to repeat it exactly, word for word, after the sentence is withdrawn from view. Color words are used so as to minimize meaning (e.g., *The green is yellowing*). The sentences are of varying lengths. The raw score is the number of words correctly recalled. For younger children (ages 5 to 7), the child repeats a specific three-word combination (like cat-book-ball) 10 times in quick succession. The raw score is the total time required. In the *Sentence Questions* subtest (ages 8 to 17), the child answers questions about orally presented sentences similar to those used in *Sentence Repetition* (e.g., *The green is yellowing. Who is yellowing?*). For younger children (ages 5 to 7), *Speech Rates* is administered instead. This subtest requires the repetition of a one-syllable and two-syllable word combination 10 times as quickly as possible. The raw score is the total time needed to complete the repetitions. Correct sequencing of stimuli or activities is essential to the Successive subtests.

In addition to 12 subtest scores and 4 process scores, The CAS also yields a Full Scale score based on the familiar mean of 100 and *SD* of 15. Psychometric properties of the test are excellent. The average internal consistency reliabilities are: Planning (.88), Attention (.88), Simultaneous (.93), Successive (.93), and Full Scale (.96). The standardization sample consisted of 2,200 children and adolescents, stratified on demographic variables to closely match the U.S. population (Naglieri, Das, & Goldstein, 2012 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1209>)). The validity of the CAS-II rests in large measure on its similarity to the first edition, the CAS, which stands up well in factor analytic studies and yields meaningful results for special groups. For example, using multigroup confirmatory factor analysis, Naglieri, Taddei, and Williams (2012 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1209>))) found that the factorial structure of the CAS was highly similar in two cross-cultural samples, one comprised of 1,174 U.S. children and the other consisting of 809 Italian children. Further, results for both samples were broadly supportive of the four factors of the PASS theory embodied in the CAS.

In a study of 60 children meeting the criteria for Attention-Deficit Hyperactivity Disorder (ADHD), Naglieri and colleagues found that subtest and process scores were theoretically consistent with current understandings of ADHD. Specifically, average scores on the four process scales were: Planning 89.1, Attention 92.3, Simultaneous 101.2, and Successive 101.7 (Naglieri & Paolitto, n.d. (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1204>))). These findings fit well with the hypothesis that children with ADHD manifest problems with goal-directed planning and show difficulties with attention due to distractibility (Barkley, 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib90>))).

An intriguing result with the CAS is that differences between Black and White children on the Full Scale score are minimal when key demographic variables such as socioeconomic status are controlled.

Naglieri, Rojahn, Matto, and Aquilino (2005

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1210>) found an estimated CAS Full Scale mean score difference of 4.8 points between Black ($N = 298$) and White ($N = 1,691$) children, smaller than typically reported with traditional IQ tests. The relationships between CAS scores and school achievement were strongly positive and highly similar for both groups as well. Overall, these results indicate that the CAS is useful for assessment in special education. On a similar note, Naglieri and Rojahn (2001

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1207>) found that CAS scores classified a smaller proportion of Blacks as having intellectual disability than did WISC-III scores. They argued that the problem of disproportionate representation of Blacks in special education classes might be mitigated if the CAS were used for this assessment purpose. The CAS-II is a promising test that deserves to see wider use in assessment and research.

5.20 KAUFMAN BRIEF INTELLIGENCE TEST-2 (KBIT-2)

The individual intelligence tests previously discussed in this and the preceding topic are excellent measures of intellectual ability, but they are not without their drawbacks. One problem is the time required to administer them. Testing sessions with the Wechsler scales, Kaufman Assessment Battery for Children, and the Stanford-Binet easily can last one hour, and two hours is not unusual if the examinee is bright and highly verbal. A second disadvantage to these mainstream tests is the amount of training required to administer them. Proper administration of most individual intelligence tests is based upon the assumption that the examiner has an advanced degree in psychology or a related field and has received extensive supervised experience with the instruments in question.

Alan Kaufman responded to the need for a brief, easily administered screening measure of intelligence by developing the Kaufman Brief Intelligence Test (K-BIT), recently released in a second edition, the KBIT-2 (Kaufman & Kaufman, 2004). The KBIT-2 consists of a Verbal or Crystallized scale that includes two types of items (Verbal Knowledge and Riddles) and a Nonverbal or Fluid Scale that consists of Matrices items (2 × 2 and 3 × 3 figural analogies).

The KBIT-2 is normed for examinees ages 4 to 90 and can be administered in approximately 20 minutes. The test yields standard scores with means of 100 and standard deviation of 15 for Verbal, Nonverbal, and combined scores. In spite of the comparability of these scoring dimensions with well-known intelligence tests, the KBIT-2 authors make it clear that their instrument is not intended as a substitute for traditional approaches (e.g., WPPSI-III, KABC-2, WISC-IV, or SB5). The KBIT-2 is mainly a screening test useful in signaling the need for more extensive assessment. The brevity of this test makes it a natural choice for research on intelligence.

The test authors suggest a number of uses for the instrument, including the following:

- Provide a quick estimate of intelligence where accuracy is not essential
- Estimate verbal versus nonverbal intelligence in children or adults
- Reevaluate intellectual status of previously tested examinees
- Screen students who may benefit from placement in gifted programs
- Screen high-risk students who may need further assessment
- Obtain a quick estimate of intelligence in adult treatment or institutional settings

The KBIT-2 manual reports highly supportive validity data from numerous correlational studies. However, the most compelling evidence for the validity of the instrument is its strong resemblance to the K-BIT, for which a substantial body of research has been published. For example, Naugle, Chelune, and Tucker (1993 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1217>)) compared K-BIT results and WAIS-R scores for 200 referrals to a neuropsychological assessment center.

The patient sample included persons with seizure disorders, head injuries, substance abuse, psychiatric disturbance, stroke, dementia, and other neurological conditions. The heterogeneity of the referral sample guaranteed a wide range of functional ability, a desirable feature in a validation study. Although the K-BIT scores tended to be about 5 points higher than their WAIS-R counterparts, the correlations between these two instruments were extremely high and theory-confirming. Vocabulary IQ (K-BIT) and Verbal IQ (WAIS-R) correlated .83; Matrices IQ (K-BIT) and Performance IQ (WAIS-R) correlated .77; and overall IQs from the two instruments correlated an amazing .88. In a study comparing the K-BIT and the WISC-III scores for 50 referred students, Prewett (1995 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1323>)) also reported

strong correlations ($r = .78$ for overall scores) and also discovered that the K-BIT scores tended to be about 5 points higher than their WISC-III counterparts. In a sample of 65 children with reading disability, Chin, Ledesma, Cirino, and others (2001

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib303>)) also found that the K-BIT overestimated WISC-III IQs by 1.2 to 5.0 points, on average. However, their study also showed that, in individual cases, K-BIT scores can underestimate or overestimate WISC-III scores by as much as 25 points, reaffirming that the K-BIT is not appropriate for placement and diagnostic purposes. Canivez (1995) found comparable scores between the K-BIT and the WISC-III for 137 elementary and middle school children and also reported very strong correlations between the two tests, especially for overall scores ($r = .87$). Eisenstein and Engelhart (1997

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib451>)) found that the K-BIT performed well in estimating IQs in adult neuropsychology referrals, but Donders (1995

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib428>)) recommends caution when using the test with brain-injured children. The reason for caution is that K-BIT scores show a negligible relationship with length of coma; that is, the test is not a good index of neuropsychological status in children. In spite of these cautions about its predecessor, the KBIT-2 is an outstanding screening measure of general intelligence for use in research or in those situations listed earlier in which time constraints preclude use of a longer instrument.

5.21 INDIVIDUAL TESTS OF ACHIEVEMENT

Whereas intelligence tests are designed to measure the broad mental abilities of the individual, achievement tests are intended to appraise what a person has learned in school or some other course of study. Group achievement tests are paper-and-pencil measures given to dozens of students at a time. These kinds of measures are discussed in **Topic 6A**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch06#ch06box1>), Group Tests of Ability and Related Concepts. Our focus here is on *individual* achievement tests administered one-on-one and, therefore, better suited for the appraisal of learning problems.

Of course, scores on intelligence and achievement tests should bear a strong relationship to one another—brighter children likely are capable of higher achievement. In fact, as we shall see, the notion that intelligence and achievement typically parallel one another is at the very heart of the concept of learning disability—which commonly involves a discrepancy between the two. We introduce the reader here to the makeup of individual achievement tests as a backdrop to the final topic in this chapter, the assessment of learning disabilities.

More than a dozen individually administered intelligence tests exist, but only a few are widely used in clinical and educational assessment. A number of prominent individual achievement tests are summarized in **Table 5.10**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec21#ch05tab10>). Owing to limitations of space, we have selected one test, the Kaufman Test of Educational Achievement-II (KTEA-II), for more detailed presentation (**Kaufman & Kaufman, 2004b**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib865>)). Readers who seek further information are encouraged to consult Sattler (2001

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1438>), Chapter 17) or the *Mental Measurements Yearbook* series.

Kaufman Test of Educational Achievement-II (KTEA-II)

The KTEA-II is an untimed test of educational achievement for children ages 4½ through 25. A brief, three-subtest version exists and extends the age range to 90+, but for diagnostic assessment of learning difficulties the Comprehensive Form is preferred. The core of the KTEA-II Comprehensive Form consists of eight subtests in four areas:

Reading

- Letter and Word Recognition
- Reading Comprehension

Mathematics

- Math Concepts and Applications
- Math Computation

Written Language

- Written Expression
- Spelling

Oral Language

- Listening Comprehension
- Oral Expression

In addition to yielding scores on each subtest, the battery provides three composite scores (Reading, Mathematics, and Written Language) and a Total Battery Composite. For diagnostic purposes, a number of supplemental subtests designed to evaluate reading skills (e.g., Phonological Awareness) are also available. For older children, the test takes about 80 minutes to administer; for younger children about 30 minutes are needed. The KTEA-II is co-normed with the KABC-II.

Brief examples of KTEA-II-like items are shown in **Table 5.11**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/ch05lev1sec21#ch05tab11>) . These examples would be at the upper end of the subtests, suitable for high school students. The KTEA-II utilizes entry and exit rules for each subtest to ensure that students only encounter items of appropriate difficulty. Scoring is objective and highly reliable. Raw scores are converted to standard scores (mean of 100, SD of 15) for each subtest, the composite scores, and the Total Battery Composite.

TABLE 5.10 Survey of Widely Used Individual Achievement Tests

Diagnostic Achievement Battery-3 (DAB-3) (Newcomer, 2001)

Suitable for ages 6 through 14, the DAB-3 consists of 14 subtests used to compute eight diagnostic composites. The composite scores include Listening, Speaking, Reading, Writing, Mathematics, Spoken Language, Written Language, and Total Achievement. More comprehensive than most achievement tests, the DAB-3 takes up to two hours to administer. The test was carefully normed on 1,534 children nationwide.

Kaufman Test of Educational Achievement (KTEA-II) (Kaufman & Kaufman, 2004b)

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib865>))

A well-normed individual test of educational achievement, a special feature of the KTEA-II is the detailed error analysis (see text). Currently, norms extend from age 4½ through age 25. A separate brief form that can be administered in 30 minutes or less is useful for screening purposes.

Mini-Battery of Achievement (MBA) (Woodcock, McGrew, & Werder, 1994)

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1786>))

Assesses four broad achievement areas—reading, writing, mathematics, and factual knowledge—for persons ages 4 through 90+. The complete battery can be administered in 30 minutes. The MBA provides a more extensive coverage of basic and applied skills than any other brief battery. For example, the reading component assesses letter-word identification, vocabulary, and comprehension.

Peabody Individual Achievement Test-Revised-Normative Update (PIAT-R/NU) (Markwardt, 1997)

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1044>))

For ages 5 through 22, this 60-minute test includes subtests of general information, reading recognition, reading comprehension, mathematics, and spelling. A new subtest, written expression, is now offered for screening written language skills. Administration of the PIAT-R/NU requires minimal training; the test can be administered by properly trained classroom teachers.

Wechsler Individual Achievement Test-II (WIAT-III) (Wechsler, 2009)

The WIAT-III consists of 16 subtests organized into seven composites. The composites are Oral Language, Total Reading, Basic Reading, Reading Comprehension and Fluency, Written Expression, Mathematics, and Math Fluency. The test is suitable for children ages 4 through adults age 50, and is empirically linked with all of the Wechsler intelligence scales. The feature of linkage allows for direct comparisons of achievement and intelligence, which facilitates the assessment of learning disabilities. A new element of the third edition is the Essay Composition subtest, which requires the examinee to write

Kaufman and Kaufman (2004b

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib865>)) stress that the error analysis provides the diagnostician with a source of information from which instructional objectives can be developed. For example, a weakness in vowel digraphs and diphthongs on the Spelling subtest translates directly to classroom objectives: practice in the spelling and reading of these elements in isolation, progressing to spelling and pronouncing words containing digraphs and diphthongs, and ending in writing and reading sentences containing words with vowel digraphs and diphthongs. The KTEA-II manual contains many useful clinical insights with educational ramifications.

The content validity of the KTEA-II appears to be very strong, but this point may vary from one school system to another. After all, individual school systems may choose to emphasize different domains of achievement. Salvia and Ysseldyke (1991) warn that users must be sensitive to the correspondence of test content with the students' curriculum. As with any achievement test, the user should verify that the content of the KTEA-II is appropriate within the curricular setting. Nonetheless, Kaufman and Kaufman (2004b (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib865>)) offer sufficient evidence for the validity of the test to make a case for general adequacy.

5.22 NATURE AND ASSESSMENT OF LEARNING DISABILITIES

Because individual intelligence and achievement tests are foundational to the assessment of learning disabilities, we close this chapter with brief review of this topic. The learning disability (LD) field is one of the fastest growing areas within assessment. Paradoxically, it is also one of the most controversial and perplexing domains of psychological testing. Some background is needed to understand the role of intelligence and achievement tests in the evaluation of learning disabilities. We begin by asking a seemingly simple question that turns out to have a complicated answer: What is a **learning disability** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss183>) ?

Definitions of learning disability have gone through at least three phases in the last several decades. Early views were influenced heavily by federal legislation and relied on a discrepancy between intelligence and achievement as the defining characteristic. These ideas were followed by a model that featured intra-individual weakness in one or more core psychological processes as the essential attribute. Most recently, responsiveness to intervention has been featured as the prevailing quality. We turn now to a survey of these shifting paradigms in the history of LD assessment.

The Federal Definition of Learning Disabilities

For decades the essential nature of learning disabilities was understood in terms of a definition embedded in federal law. In 1975, Congress passed Public Law 94-142, the Education for All Handicapped Children Act. One of the provisions of this act was a definition of learning disabilities as follows:

The term "specific learning disability" means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in imperfect ability to listen, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning disabilities which are primarily the result of visual, hearing, or motor handicaps, of mental retardation, or emotional disturbance, or of environmental, cultural, or economic disadvantage. (USDE, 1977 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1675>), p. 65083)

The commitment to a federally mandated definition was reaffirmed in 1990 by passage of Public Law 101-476, the Individuals with Disabilities Education Act (IDEA).

The federal definition embodied in IDEA also stipulated an operational approach to the identification of children with learning disabilities. Specifically, candidates for an LD diagnosis had to demonstrate a severe discrepancy between general ability (intelligence) and specific achievement in one or more of these seven areas:

- Oral expression
- Listening comprehension
- Written expression
- Basic reading skill
- Reading comprehension
- Mathematics calculation
- Mathematics reasoning

The discrepancy model for the identification of LD children functioned as a directive for school psychologists. In effect, the model mandated that psychologists should administer an individual intelligence test (general ability measure) and an individual achievement test (specific achievement measure) and then look for a discrepancy between Full Scale IQ and one or more areas of school achievement (e.g., reading, mathematics, written expression).

In practical terms, a severe discrepancy was defined as a difference of one standard deviation or more between general intelligence and specific achievement. A common practice in identification of LD children was to compare Full Scale IQ on an individual intelligence test such as the WISC-III with specific achievement scores on an individual achievement test such as the WIAT (Wechsler Individual Achievement Test) or similar instrument that has subtests normed with a mean of 100 and a standard deviation of 15. A difference of 15 points or more between Full Scale IQ and specific achievement in any of the previously listed areas would then raise the suspicion of learning disability.

Unfortunately, the federal definition did not serve its intended purposes, and, increasingly, school psychologists and other professionals looked to other approaches for understanding and assessing learning disabilities in children. The fundamental problem was that many, many children who exhibit serious learning problems in school and who would benefit from services for LD simply did not meet the psychometric criteria of a severe discrepancy.

The National Joint Committee on Learning Disabilities Definition

After a lengthy period of confusion and struggle over the definition of learning disabilities, specialists and educators began to rally around a consensus view in the early 1990s. The new definition was proposed by the National Joint Committee on Learning Disabilities (NJCLD), a group of representatives from eight national organizations with a special interest in learning disabilities. Although similar to the federal definition, the new approach contains important contrasts:

Learning disabilities is a general term that refers to a heterogeneous group of disorders manifested by significant difficulties in the acquisition and use of listening, speaking, reading, writing, reasoning, or mathematical abilities. These disorders are intrinsic to the individual, presumed to be due to central nervous system dysfunction, and may occur across the life span. Problems in self-regulatory behaviors, social perception and social interaction may exist with learning disabilities but do not by themselves constitute a learning disability. Although learning disabilities may occur concomitantly with other handicapping conditions (for example, sensory impairment, mental retardation [MR], serious emotional disturbance [ED]) or with extrinsic influences (such as cultural differences, insufficient or inappropriate instruction), they are not the result of those conditions or influences. (NJCLD, 1988 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1216>), p. 1)

The new definition avoided vague reference to “basic psychological processes,” specifies that the disorder is intrinsic to the individual, identifies central nervous system dysfunction as the origin of LD problems, and states explicitly that learning disabilities may extend into adulthood.

Perhaps most important of all, the NJCLD approach abandoned the excessive reliance upon discrepancy between ability and achievement as the hallmark of LD. Instead, the new model specified that the necessary (but not sufficient) condition of LD was that the individual (child or adult) exhibit an intraindividual weakness in one or more of the core areas of academic functioning (listening, speaking, reading, writing, reasoning, or mathematical abilities). Shaw et al. (1995 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1487>)) described how

the NJCLD model might look in practice. In this approach, the first task is to identify one or more intraindividual weaknesses in the core areas. These are always relative to strengths in several other core areas. In other words, persons who are slow learners in all areas do not meet the criteria of LD. The second step is to trace the learning difficulties to central nervous system dysfunction, which may manifest as problems with information processing. For example, a young adult with a severe weakness in listening (as judged by her inability to learn from the traditional lecture approach to teaching) might exhibit a deficit on a test of verbal memory—confirming that an information processing problem was at the heart of her disability. The purpose of the third step (examining psychosocial skills, physical and sensory abilities) is to specify additional problems that may need to be addressed for program-planning purposes. Finally, in the fourth step the examiner rules out non-LD explanations for the learning difficulties (since these explanations would mandate a different strategy for remediation).

The New Face of Learning Disabilities: Response to Intervention

In 2004, Congress reauthorized the Individuals with Disabilities Education Act (IDEA), which is the ongoing legislation governing special services, including the assessment of LD, in school systems that receive federal funding. IDEA 2004 changed the law about how to identify children with specific learning disabilities by moving away from the discrepancy model that had reigned supreme since the 1970s.

Instead, the new law recommended **response to intervention (RTI)**

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm01#bm01gloss278>) as the preferred method for identifying children with learning disabilities. In particular, IDEA 2004 says that a school “may use a process that determines if the child responds to scientific, research-based intervention as part of the evaluation procedures . . .” in its evaluation for LD.

RTI is a broader concept than LD and refers both to (1) methods for increasing the capacity of school systems to respond effectively to the diverse academic needs of students and (2) approaches for identifying LD children who need special education services. The RTI approach specifically deemphasizes cognitive discrepancies in the diagnostic process, focusing instead on low age-based achievement levels and failure to respond to evidence-based instructional approaches (**Fletcher & Vaughn, 2009** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib512>); **Torgerson, 2009** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1657>)).

The implementation of RTI is complicated and multifaceted. The process involves multiple feedback loops and decision points. Yet, proponents of RTI view it as an improvement because it provides for early, preventive intervention in contrast to the “wait to fail” approach of the discrepancy model. Fuchs and Fuchs (**2005** (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib548>)) describe a systematic approach to using RTI in a school system. The first step is school-wide screening in the first weeks of the school year to identify children “at risk” for school failure. Those scoring below a certain prescribed cut-off (perhaps the 25th percentile in reading or math) would be noted. Teachers would then implement empirically validated curricular interventions for these children, who would be monitored for progress after eight weeks. Those who do not respond would receive another interval of supplementary instruction for an additional eight weeks. Those who still do not respond would receive a comprehensive, individualized evaluation to rule out sources of underachievement such as intellectual disability, visual problems, or emotional disturbance. Finally, with the involvement of parents, the child would receive a designation of LD and become eligible for special education placement.

In sum, RTI is a shift in perspective that focuses on early results and outcomes with at-risk children instead of later spending excessive time and resources on questions of discrepancy-based eligibility after children are already failing because of their LD. The hope is that the RTI perspective will catch at-risk children earlier and thereby reduce the number of children needing special education services.

Essential Features of Learning Disabilities

Even though the definition of LD remains a point of contention, we can cite several features of these disorders that are less controversial. As the reader will discover, the features discussed in the following dictate, to some extent, the nature of testing practices in the assessment of learning disabilities. There is general agreement—with occasional dissenting votes—on five features of learning disabilities.

First, a learning disability involves an intraindividual discrepancy in cognitive functioning. The child (or adult) with LD reveals a relative weakness in one area compared to strengths in most other areas. According to the federal definition followed within many school systems, the discrepancy is between general ability (intelligence) and specific achievement. We have described previously some of the pitfalls of this definition and prefer the NJCLD approach in which the discrepancy is not rigidly tied to a difference between IQ and achievement test scores.

Second, an exclusionary clause is included in most definitions of learning disability. If the academic difficulties are primarily caused by other disabling conditions (mental retardation, emotional disturbance, visual or hearing impairment, cultural or social disadvantage), then a diagnosis of learning disability is typically ruled out. This clause is often misinterpreted. A person can be both learning disabled and impaired in other ways (e.g., have mental retardation). The important point is that the coexisting condition must not be the primary cause of the learning difficulties.

Third, Learning disabilities are heterogeneous; that is, there are many different varieties. Research on the identification of subtypes is still in its infancy, but most researchers express optimism that meaningful subgroups of persons with learning disabilities can be identified. Pending further research and refinement, only two broad categories of learning disability are recognized currently. These two types are dyslexia or verbal learning disability, and right hemisphere or nonverbal learning disability. Our coverage here is based on Forster (1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib525>). The primary manifestation of dyslexia is an unexpected difficulty in learning to read or spell. The fundamental deficiency is thought to be a problem with phonological coding, which is the ability to automatically associate sounds with specific letter combinations. Verbal learning disability constitutes about 90 percent of all LD cases, and is much more common in boys than girls. In contrast, right hemisphere or nonverbal learning disability manifests as poor skills in mathematics, handwriting, and, often, social cognition. The fundamental problem is thought to be a problem in spatial cognition, which is the visuospatial perception of relationships. The problem likely originates in right cerebral hemisphere dysfunction, and constitutes about 10 percent of all LD cases. Boys and girls are equally affected.

Fourth, a learning disability is a developmental phenomenon that is usually evident in early childhood that may persist into adulthood. Even though remediation efforts should be based upon optimism—so as to avoid self-fulfilling prophecies—a dose of realism is needed, too. Longitudinal studies of children with severe learning disabilities suggest that marked improvement in academic achievement is the exception, not the rule, even when these subjects receive intensive educational intervention. For example, Frauenheim and Heckerl (1983

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib536>) re-tested 11 adults diagnosed as having learning disabilities in childhood. All the participants had received special help for reading; nine had graduated from high school, and two completed the 10th grade. Full Scale IQs were typically in the low 90s, with Verbal IQ below average (mean of 85) and Performance IQ above average (mean of 104). In spite of the remedial intervention, when retested as adults on *exactly* the same achievement test (Wide Range Achievement Test), these examinees were scarcely improved from their elementary school results. These findings are corroborated by several other follow-up studies (see Kolb

& Whishaw, 1990, chap. 29, for a review). Such results indicate that specialists who work with children with learning disabilities should not become fixated solely on academic concerns. Social and emotional problems—which may be more amenable to intervention—also cry out for notice.

Fifth, individuals with learning disabilities frequently experience social and emotional difficulties that are as pervasive and consequential as the deficits in academic achievement. These problems may persist into adolescence and adulthood. In fact, the socioemotional sequelae often become the primary presenting complaint, which can complicate the testing process and obscure the diagnosis. For example, in a needs assessment study of 381 adults with learning disabilities, Hoffman, Sheldon, Minskoff, and others (1987 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib753>)) identified several crucial nonacademic areas meriting intervention by service providers. These adults self-endorsed several social and emotional problems with high frequency: feeling frustrated (40 percent), talking or acting before thinking (33 percent), being shy (31 percent), no self-confidence (28 percent), controlling emotions and temper (28 percent), and dating (27 percent). Many other problems were also endorsed, but by less than 25 percent of the sample. These findings indicate that learning disability assessments should incorporate measures of social and emotional functioning. Vaughn and Haager (1994 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1691>)) provide an excellent overview on the measurement of social skills in persons with learning disability.

Causes and Correlates of Learning Disabilities

Approximately 4 to 5 percent of all school-aged children receive a diagnosis of LD, so this is not a rare problem (Lyon, 1996). The most common form of LD is dyslexia, and boys outnumber girls by about 3:1 or 4:1 (Forster, 1994

(<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib525>)). In a minority of cases, the etiology is clear and can be attributed to a specific cause such as a known brain injury. Left hemisphere impairment is especially likely to result in verbal difficulties, whereas right hemisphere impairment may lead to problems with spatial thinking or other nonverbal skills. Thus, head injury or other neurological problems can be the proximate cause of a child receiving an LD diagnosis.

However, in the majority of cases the direct etiology of LD problems is unclear. A number of possibilities have been proposed and these may explain some but not all cases of LD. For example, pathological neurodevelopmental processes have been identified in some persons with severe dyslexia (Culbertson & Edmonds, 1996 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib377>)). Individuals with this disorder appear to have alterations in brain structures such as the planum temporale (the flat surface on the top of the temporal lobes) known to be important for language processing. Whereas in normal individuals the planum temporale is much larger in the left temporal lobe than in the right, persons with severe dyslexia do not show this pattern of asymmetry (tending toward symmetry instead). Moreover, researchers have identified microscopic cortical malformations called polymicrogyria (numerous small convolutions) that parallel these structural differences. Several postmortem studies of persons with severe dyslexia have revealed these deviations at the cellular level. Spreen (2001 (<http://content.thuzelearning.com/books/Gregory.8055.17.1/sections/bm02#bm02bib1540>)) provides an outstanding review of the possible neurological substrates of learning disabilities. Dyslexia also appears to show a significant genetic component for some persons such that the idea of familial dyslexia needs to be taken seriously. However, what must be emphasized is that for most individuals the etiology of LD (whether dyslexia or other forms) remains a mystery.

Achievement Tests in LD Assessment: A Final Word

Learning disabilities manifest primarily as academic problems; that is, a child with LD is typically unable to master skills important for school success such as reading, mathematics, or written communication.

Because school-based accomplishment is at the heart of the problem, an evaluation for LD must include relevant measures of academic achievement. Furthermore, the evaluation of school achievement—one small part of an LD assessment—must be based on an *individual* test of achievement. Even though a group achievement test might raise the suspicion of a learning disability, practitioners must rely on individual achievement tests for definitive assessment.

Individual achievement tests typically are administered one-on-one with the examiner sitting across from the respondent and posing structured questions and problems. Of course, any well-standardized achievement test will yield normative data about the functioning of a schoolchild. But the special virtue of individual achievement tests is that the examiner can observe the clinical details of deficient (or superior) performance and form hypotheses about the cognitive capacities of the examinee.

Consider the problem of poor spelling, widely observed in children and adults with verbal LD. Any good spelling achievement test will document the disability; however, little insight is gained from mere scores. What the examiner should seek to know is the qualitative nature of the problem, not just its quantitative dimensions. Individual achievement tests are invaluable in this regard. By observing the details of deficient performance, an astute examiner can form hypotheses about the origin of an achievement problem. For example, a child whose spelling is phonetically correct is at least *hearing* the words correctly, whereas a child with nonphonetic spelling might very well display a problem with auditory processing of speech sounds.

Please use the Chapter 5 Resource For This Assignment Also - Assignment 2

In this interactive assignment, you will create a story for children to educate them about psychological assessment. To begin, select a targeted developmental or reading level pre-K through grade 6. Review the elements required for each section of your storybook below. Visit the [My Storybook website](#) (Links to an external site.) to familiarize yourself with this technology. Please review the instructions on the home page for Take a Writing Lesson, Build your Storybook, and Keep your Stories by selecting the link for [Try it Now](#). After you create your free account and develop your story book, make sure you select Publish Public so that you can share your story book. You will be provided a link that you can share in the Week Two Interactive Assignment page so that other students can review your story. Do not select "share" or "download" as those require you to pay, which is not necessary for this assignment. As an alternative, you can develop a PowerPoint presentation with a screenshot of each page of your book for each slide. Please contact your instructor if you have any questions.

Your username will become the professional author name for your book. Therefore, when you register for your My Storybook account it is recommended that you use the following format for your username: first initial followed by last name (e.g., JSmith). If you receive a message that your username has been taken, it is recommended you include your middle initial (e.g., JASmith). Do not use Internet handles and/or other unprofessional appearing pseudonyms.

Review the information in Chapter 5 of your textbook corresponding to the assessments appropriate for the age group you selected and review the [Mental Health Assessment](#) (Links to an external site.) (2013) article for examples of information provided to the public about psychological testing. You may choose any appropriate title for your story. Be sure to address each of the following questions in your storybook in an age-appropriate manner:

- Why is the character in the story being referred for testing?
- Who will conduct the assessment?
- What is being measured?
- How long will testing take?
- Who will be present during the assessment process? If not in the room, where will parents and/or guardians be while the character in the story is being tested?
- How will the results be used? Who will have access to the results (e.g., medical doctor, family, the court, teachers), and why? This will vary depending on the character and plot in your story.
- How will the tests be taken?
- What will be the outcome of the assessment? How will the information be used? How might this information impact the life of the character in your story?

Include content to address any developmentally appropriate fears that individuals of the age group you selected may have. For example, young children commonly associate going to the doctor with getting a shot.

Be sure to include all the required material from the instructions above in your online storybook. Once you have created your storybook, include the link in your initial post. In your initial post, note the age or grade level for the target audience of the story. Briefly

analyze and comment on the challenges and benefits related to explaining psychological assessment concepts using language that is developmentally appropriate for children. Compare at least two assessment instruments used to assess intelligence or achievement for the age group you selected. Include in-text citations and references for all sources used.

Note: It is highly recommended you complete all written work in a separate document first and then cut and paste the required content into your online storybook. This will allow you to edit and save your work separately from the online storybook, should any technical failures occur. This approach will also allow you to work on your content without having to remain connected to the Internet, which will make it easier to develop and edit your content prior to publishing it to your online storybook.

If you experience any technical difficulties, please visit the [My Storybook Help Center](#) (Links to an external site.). The technical support offered through your Student Portal will not be able to assist you with the My Storybook website.



HEALTH
A-Z

DRUGS &
SUPPLEMENTS
LIVING
HEALTHY

FAMILY &
PREGNANCY

NEWS &
EXPERTS

SEARCH



ADVERTISEMENT

Mental Health Assessment

Mental Health > Reference >

MENU

MENTAL HEALTH

- Anxiety & Panic Disorders
- Bipolar Disorder
- Depression
- Eating Disorders
- Schizophrenia
- Substance Abuse & Addiction
- News & Features
- Reference
- Quizzes
- Slideshows
- Videos
- Questions & Answers
- Message Board
- Find a Psychiatrist

RELATED TO MENTAL HEALTH

- ADHD
- Crisis Assistance
- Health & Balance
- Living Healthy
- Military Families Support
- Personality Disorders
- PTSD
- Stress Management
- More Related Topics

What Are Mental Health Assessments?



A mental health assessment is when a professional -- like your family doctor, a psychologist, or a psychiatrist -- checks to see if you might have a mental problem and what type of treatment may help.

Everyone goes through tough times. But sometimes, the negative way someone feels inside -- depressed, anxious, wanting to avoid people, having trouble thinking -- may be more than the ups and downs most people feel now and then. If symptoms like these start to get in the way of your life, or that of a loved one, it's important to take action. Research shows that getting help early can prevent

TODAY ON WEBMD



Medications Disorder
Which drug is b

FROM OUR ADVERTISER
Uncontrollable Limb Movements
Learn about Tardive Dyskinesia
Neurocrine Biosciences



Body-Focus Behaviors
Not just a bad h



Warning Sign Disorder
When it's not "ju



What You Should Know About Bipolar Disorder
Behind the extreme swings.



Diagnosing Mental Illnes...

SUBSCRIBE

SEARCH



The first step is to get a mental health assessment. It usually involves a couple of different things. You may answer questions verbally, get physical tests, and fill out a questionnaire.

What to Expect

Physical exam. Sometimes a physical illness can cause symptoms that mimic those of a mental illness. A physical exam can help find if something else, such as a thyroid disorder or a neurologic problem, may be at play. Tell your doctor about any physical or mental health conditions that you already know you have, any prescription or over-the-counter drugs you take, and any supplements you use.

CONTINUE READING BELOW

YOU MIGHT LIKE

FROM OUR ADVERTISER

Identifying Involuntary Movements

Learn about Tardive Dyskinesia
Neurocrine Biosciences

WEBMD

Diagnosing Prostate Cancer

WebMD explains which tests are used to help diagnose prostate cancer.

Lab tests. Your doctor may order bloodwork, a urine test, a brain scan, or other tests to rule out a physical condition. You will probably also answer questions about drug and alcohol use.

ADVERTISEMENT

RECOMMENDED FOR YOU



ARTICLE
Schizophrenia Exams and Tests



ARTICLE
At-Home Bipolar Disorder Test: Help or Hindrance?

Choices



ARTICLE
Alzheimer's Disease Stages



ARTICLE
Medication for Bipolar Disorder



SLIDESHOW
Conditions That Can Change Your Personality

ADVERTISEMENT



Diagnosing Mental Illnes...

SUBSCRIBE

SEARCH

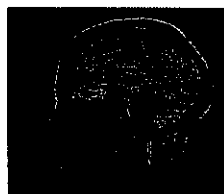
your symptoms, your personal or family history of mental health issues, and any psychiatric treatment you've had.

Personal history. Your doctor may also ask questions about your lifestyle or personal history: Are you married? What sort of work do you do? Did you ever serve in the military? Have you ever been arrested? What was your upbringing like? Your doctor may ask you to list the biggest sources of stress in your life or any major traumas you've had.

Sponsor Content



Current MS Research in Grey Matter



[Learn More](#)

Research shows that changes in grey matter play a bigger role in multiple sclerosis (MS) than once believed.

Mental evaluation. You'll answer questions about your thoughts, feelings, and behaviors. You may be asked about your symptoms in more detail, such as how they affect your day-to-day life, what makes them better or worse, and whether and how you've tried to manage them on your own. Your doctor will also observe your appearance and behavior: Are you irritable, shy, or aggressive? Do you make eye contact? Are you talkative? How do you appear, compared with others your age?

Cognitive evaluation. During the assessment, your doctor will gauge your ability to think clearly, recall information, and use mental reasoning. You may take

Cancer



ARTICLE
Allergic to Milk? How to Spot Hidden Ingredients



ARTICLE
5 Diagnoses That Call for a Second Opinion

ADVERTISEMENT

TOOLS & RESOURCES

[Symptoms of Mental Illness](#)

[Types of Mental Illness](#)

[Schizophrenia Exams and Tests](#)

[Causes of Mental Illness](#)

[What Is Dissociation?](#)

[Dissociative Identity Disorder \(Multiple Personality Disorder\)](#)

SUBSCRIBE TO WEBMD NEWSLETTERS

- Food & Fitness
- Women's Health
- Men's Health
- Good Health

Enter email address



Diagnosing Mental Illnes...

SUBSCRIBE

SEARCH

recognizing common shapes or objects, or solving simple math problems. You may answer questions about your ability to do daily responsibilities, like caring for yourself or going to work.

By clicking "Submit," I agree to the WebMD Terms and Conditions and Privacy Policy. I also agree to receive emails from WebMD and I understand that I may opt out of WebMD subscriptions at any time.

1 2

View All



> NEXT PAGE

SPONSORED ADS

TOP PICKS

Do You Have Trouble Concentrating?

Ways to Talk to Loved Ones About Your Depression

MS and Depression: Tips for Mental Health

Cognitive Therapy: How It Helps Mental Health

Meet Your Acute Myeloid Leukemia Care Team

Getting the Support You Need When You're Depressed




Diagnosing Mental Illness...

SUBSCRIBE

SEARCH



Sponsor Content 

Aspirin: no longer recommended to reduce AFib-related stroke risk.



[Learn More](#)

Health Solutions FROM OUR SPONSORS

First Aid 101
MS Tips
Opioid Addiction
Myths About Epilepsy

Fight Psoriasis Flare-ups
Missing Teeth?
The Fasting Mimicking Diet
Is My Penis Normal?

The Fight Against Germs
Aging & Addiction
Mammograms Save Lives
Life After Cancer Diagnosis

Epilepsy Explained
Medicare Personal Consultations
Avoid Colds This Winter
Bent Fingers?

More from WebMD

MS: Tools to Keep Your Mind Sharp
How MS Affects Your Mind
What Is Endometriosis?
Life With Migraine

First Psoriatic Arthritis Flare
A Personal Story of RA
Beat Crohn's Flares
Psoriasis in the Salon

Immunotherapy for Lung Cancer
Common Psoriasis Triggers
Have Hives? Things You Need to Know
How Beta Thalassemia Is Treated

Psoriatic Arthritis and Your Sleep
Finding the Best MS Care Team
Multiple Myeloma Explained
Where Breast Cancer Spreads



Policies

- [Privacy Policy](#)
- [Cookie Policy](#)
- [Editorial Policy](#)
- [Advertising Policy](#)
- [Correction Policy](#)
- [Terms of Use](#)

About

- [Contact Us](#)
- [About WebMD](#)
- [Careers](#)
- [Newsletter](#)
- [Corporate](#)
- [WebMD Health Services](#)
- [Site Map](#)
- [Accessibility](#)

WebMD Network

- [Medscape](#)
- [Medscape Reference](#)
- [MedicineNet](#)
- [eMedicineHealth](#)
- [RxList](#)
- [OnHealth](#)
- [WebMDRx](#)
- [First Aid](#)
- [WebMD Magazine](#)
- [WebMD Health Record](#)
- [Dictionary](#)
- [Physician Directory](#)

Our Apps

- [WebMD Mobile](#)
- [WebMD App](#)
- [Pregnancy](#)
- [Baby](#)
- [Allergy](#)
- [Medscape](#)

For Advertisers

- [Advertise with Us](#)
- [Advertising Policy](#)

© 2005 - 2020 WebMD LLC. All rights reserved.

WebMD does not provide medical advice, diagnosis or treatment.



Diagnosing Mental Illnes...

SUBSCRIBE

SEARCH



