13 Verbal Functions and Language Skills

The most prominent disorders of verbal functions are the aphasias and associated difficulties in verbal production such as *dysarthria* (defective articulation) and apraxias of speech. Other aspects of verbal functions that are usually affected when there is an aphasic disorder, such as fluency and reading and writing abilities, may be impaired without aphasia being present. Assessment of the latter functions will therefore be discussed separately from aphasia testing.

APHASIA

Aphasic disorders can be mistakenly diagnosed when the problem actually results from a global confusional state, a dysarthric condition, or elective mutism. The reverse can also occur when mild deficits in language comprehension and production are attributed to generalized cognitive impairment or to a memory or attentional disorder. Defective auditory comprehension, in particular, whether due to a hearing disorder or to impaired language comprehension, can result in unresponsive or socially inappropriate behavior that is mistaken for negativism, dementia, or a psychiatric condition. In fact, aphasia occurs as part of the behavioral picture in many brain disorders (Mendez and Cummings, 2002) so that often the question is not whether the patient has aphasia, but rather how (much) the aphasia contributes to the patient's behavioral deficits. Questions concerning the presence of aphasia can usually be answered by careful observation in the course of an informal but systematic review of the patient's capacity to perceive, comprehend, remember, and respond with both spoken and written material, or by using an aphasia screening test. A review of language and speech functions that will indicate whether communication problems are present will include examination of the following aspects of verbal behavior:

1. *Spontaneous speech.*
2. *Repetition* of words, phrases, sentences. "Methodist Episcopal" and similar tongue-twisters elicit disorders of articulation and sound sequencing. "No ifs, ands, or buts" tests for the integrity of connections between the center for expressive speech (Broca's area) and the receptive speech center (Wernicke's area).
3. *Speech comprehension.* a. Give the subject simple commands (e.g., "Show me your chin." "Put your left hand on your right ear."). b. Ask "yes-no" questions (e.g., "Is a ball square?"). c. Ask the subject to point to specific objects.

The wife of a patient diagnosed as a global aphasic (expression and comprehension severely defective in all modalities) insisted that her husband understood what she told him and that he communicated appropriate responses to her by gestures. I examined him in front of her, asking him—in the tone of voice she used when anticipating a "yes" response—"Is your name John?" "Is your name Bill?" etc. Only when she saw him eagerly nod assent to each question could she begin to appreciate the severity of his comprehension deficit [mdl].

An inpatient with new onset global aphasia nodded enthusiastically and said "yes" to all questions, causing his physicians to believe that he had consented to a surgical procedure because they had not asked him a question in which "no" was the appropriate answer [dbh].

4. *Naming.* The examiner points to various objects and their parts asking, "What is this?" (e.g., glasses, frame, nose piece, lens; thus asking for object names in the general order of their frequency of occurrence in normal conversation). Ease and accuracy of naming in other categories, such as colors, letters, numbers, and actions, should also be examined (Goodglass, 1980; Strub and Black, 2000).
5. *Reading.* To examine for accuracy, have the subject read aloud. For comprehension, have the subject follow written directions (e.g., "Tap three times on the table"), explain a passage just read.
6. *Writing.* Have the subject copy, write to dictation, and compose a sentence or two.

When evaluating speech, Goodglass (1986) pointed out the importance of attending to such aspects as the ease and quantity of production (*fluency*), articulatory error, speech rhythms and intonation (*prosody*)), grammar and syntax, and the presence of misspoken words (paraphasias). Although lapses in some of these aspects of speech are almost always associated with aphasia, others—such as articulatory disorders—may occur as speech problems unrelated to aphasia. The examiner should also be aware that familiar and, particularly, personally relevant stimuli will elicit the patient's best responses (Van Lancker and Nicklay, 1992). Thus, a patient examined only on standardized tests may actually communicate better at home and with friends than test scores suggest, particularly when patients augment their communication at home with gestures.

Formal aphasia testing should be undertaken when aphasia is known to be present or is strongly suspected. It may be done for any of the following purposes:

> (1) diagnosis of presence and type of aphasic syndrome, leading to inferences concerning cerebral localization; (2) measurement of the level of performance over a wide range, for both initial determination and detection of change over time; (3) comprehensive assessment of the assets and liabilities of the patient in all language areas as a guide to therapy. (Goodglass and Kaplan, 1983a, p. 1)

The purpose of the examination should determine the kind of examination (screening, symptom focused, or comprehensive?) and the kinds of tests required (Mazaux, Boisson, et Daverat, 1989; Spreen and Risser, 1991).

Aphasia tests differ from other verbal tests in that they focus on disorders of symbol formulation and associated apraxias and agnosias. They are usually designed to elicit samples of behavior in each communication modality—listening, speaking, reading, writing, and gesturing. The examination of the central "linguistic processing of verbal symbols" is their common denominator (Wepman and Jones, 1967). Aphasia tests also differ in that many involve tasks that most adults would complete with few, if any, errors.

Aphasia Tests and Batteries

The most widely used aphasia tests are actually test batteries comprising numerous tests of many discrete verbal functions. Their product may be a score or index for diagnostic purposes or an orderly description of the patient's communication disabilities. Most aphasia tests involve lengthy, precise, and well-controlled procedures. They are best administered by persons, such as speech pathologists, who have more than a passing acquaintance with aphasiology and are trained in the specialized techniques of aphasia examinations.

Aphasia test batteries always include a wide range of tasks so that the nature and severity of the language problem and associated deficits may be determined. Because aphasia tests concern disordered language functions in themselves and not their cognitive ramifications, test items typically present very simple and concrete tasks most children in the lower grades can pass. Common aphasia test questions ask the patient (1) to name simple objects ("What is this?" asks the examiner, pointing to a cup, a pen, or the picture of a boy or a clock); (2) to recognize simple spoken words ("Put the spoon in the cup"); (3) to act on serial commands; (4) to repeat words and phrases; (5) to recognize simple printed letters, numbers, words, primary level arithmetic problems, and common symbols; (6) to give verbal and gestural answers to simple printed questions; and (7) to print or write letters, words, numbers, etc. In addition, some aphasia tests and examination protocols ask the patient to tell a story or draw. Some examine articulatory disorders and apraxias as well (Goodglass, 1986; Stringer, 1996).

Aphasia test batteries differ primarily in their terminology, internal organization, the number of modality combinations they test, and the levels of difficulty and complexity to which the examination is carried. The tests discussed here are both representative of the different kinds of aphasia tests and among the best known. Some clinicians devise their own batteries, taking parts from other tests and adding their own. Detailed reviews of many batteries and tests for aphasia can be found in A.G. Davis, *A Survey of Adult Aphasia* (1993); Spreen and Risser, *Assessment of Aphasia* (2003); and Spreen and Strauss, A *Compendium of Neuropsychological Tests* (1998).

Boston Diagnostic Aphasia Examination (BDAE-2) (Goodglass and Kaplan, 1983a,b), *Boston Diagnostic Aphasia Examination (BDAE-3)* (Goodglass, Kaplan, and Barresi, 2000)

This test battery was devised to examine the "components of language" that would aid in diagnosis and treatment and in the advancement of knowledge about the neuroanatomic correlates of aphasia. It provides for a systematic assessment of communication and communication-related functions in 12 areas defined by factor analysis, with a total of 34 subtests. Time is the price paid for such thorough coverage, for a complete examination takes from one to four hours. As a result many examiners use portions of this test selectively, often in combination with other neuropsychological tests. The BDAE-3 has a new short form that takes only an hour or less. A number of "supplementary language tests" are also provided, to enable discrimination of such aspects of psycholinguistic behavior as grammar and syntax and to examine for disconnection syndromes (see below). The extended version of the BDAE-3 contains instructions for examining the praxis problems which may accompany aphasia.

Evaluation of the patient is based on three kinds of observations. The score for the *Aphasia Severity Rating Scale* has a 6-point range for the 1983 BDAE and a 5-point range for the BDAE-3, based on examiner ratings of patient responses to a semistructured interview and free conversation. Subtests are scored for number correct and converted into percentiles derived from a normative study of aphasic patients, many presenting with relatively selective deficits but, unlike the original 1972 standardization, also including the most severely impaired. These scores are registered on the *Subtest Summary Profile* sheet, permitting the examiner to see at a glance the patient's deficit pattern. In addition, this battery yields a "Rating Scale Profile" for qualitative speech characteristics that, the authors point out, "are not satisfactorily measured by objective scores" but can be judged on seven 7-point scales, each referring to a particular feature of speech production. For some of these scales requiring examiner judgment, relatively low interrater reliability coefficients have been reported (Kertesz, 1989). However, interrater agreement correlations typically run above .75, and percent agreement measures also indicate generally satisfactory agreement levels (A.G. Davis, 1993). Based on his review of BDAE research, Davis suggested that BDAE scores predict performance on other aphasia tests better than patient functioning in "natural circumstances." Data from a 1980 (Borod, Goodglass, and Kaplan) normative study of the BDAE and the supplementary spatial-quantitative tests (see below) contributed to the 1983 norms. The 1999 standardization sample includes 85 adults with aphasia and 15 normal elderly persons. Subjects with low education have lower scores (Borod, Goodglass, and Kaplan, 1980; Pineda, et al., 2000).

Supplementing the verbal BDAE as part of the comprehensive examination for aphasics is a Spatial Quantitative Battery (called the Parietal Lobe Battery [PLB]) (Goodglass and Kaplan, 1983a). This set of tests includes constructional and drawing tasks, finger identification, directional orientation, arithmetic, and clock drawing tasks. While sensitive to parietal lobe lesions, patients with both frontal and parietal damage are most likely to be impaired on this battery (Borod, Carper, Goodglass, and Naeser, 1984).

The range and sensitivity of the "Boston" battery makes it an excellent tool for the description of aphasic disorders and for treatment planning. However, an examiner must be experienced to use it diagnostically. Normative data for the individual tests allow examiners to give them separately as needed, which may account for some of this battery's popularity. Of course, not least of its advantages are the attractiveness and evident face validity of many of the subtests (e.g., the Cookie Theft picture; a sentence repetition format that distinguishes between phrases with high or low probability of occurrence in natural speech).

Two translations of this battery are available. Rosselli, Ardila and their coworkers (1990) provide norms for a Spanish language version (Goodglass and Kaplan, 1986). A French version developed by Mazaux and Orgogozo (1985) has retained the *z*-score profiling of the BDAE first edition.

Communication Abilities in Daily Living, 2nd ed. (CADL-2) (Holland, Frattali, and Fromm, 1999)

The disparity between scores that patients obtain on the usual formal tests of language competency and their communicative competency in real life led to the development of an instrument that might reduce this disparity by presenting patients with language tasks in familiar, practical contexts. The original CADL (Holland, 1980) examined how patients might handle daily life activities by engaging them in role-playing in a series of simulated situations such as "the doctor's office," encouraging the examiner to carry out a dual role as examiner/play-acting participant with such props as a toy stethoscope.

The CADL-2 revision eliminated items that require role playing and most props. This reduced the number of items from the original 68 to 50 but retained the focus on naturalistic everyday communications (e.g., with a telephone, with real money). The number of communication categories was reduced from ten to seven in the CADL-2: (1) reading, writing, and using numbers; (2) communication sequences; (3) social interactions; (4) response to misinformation or proverbs; (5) nonverbal communication; (6) contextual communication; (7) recognition of humor, metaphor. Examination informality is encouraged.

A series of evaluations of CADL performances of 130 aphasic patients demonstrated that this test was sensitive to aphasia, age, and institutionalization (unspecified) but not sex or social background (Holland, 1980). The CADL differentiated patients with the major types of aphasia on the single dimension of severity of communicative disability based on the summation score. The ten category scores also identified aphasia subtypes. The CADL-2 normative sample includes 175 adults with communication disorders, primarily from stroke or TBI. Test-retest reliability for CADL-2 was .85, and interrater reliability for stanine scores was .99.

Because responses need not be vocalized to earn credits, this test tends to be more sensitive to the communication strengths of many speech-impaired (e.g., Broca's aphasia) patients than are traditional testing instruments. Spreen and Risser (2003) recommend the

CADL to provide the descriptive information about functional communication that is lacking in all the larger, comprehensive, batteries: "it allows an estimate of the patient's communication ability rather than . . . accuracy of language" (Spreen and Strauss, 1998). Yet, A.G. Davis (1993) warned, CADL findings cannot be interpreted as representing naturalistic behavior as it "is still a test" and, as such, "does not provide for observing natural interactions."

Functional Communication Profile (FCP) (M.T. Sarno, 1969)

This is a 45-item inventory that takes 20 to 40 minutes to administer. It permits serial scaled ratings of a patient's practical language behavior elicited "in an informal setting," as distinguished from language on more formal testing instruments since "improvement as measured by higher (formal) test scores does not always reflect improvement" in the patient's day-to-day activities (J.E. Sarno et al., 1971). Like battery type aphasia tests, the Functional Communication Profile also requires an experienced clinician to apply it reliably and sensitively. Evaluation proceeds in five different performance areas: "Movement," "Speaking," "Understanding," "Reading," and "Other," not exclusively verbal, adaptive behaviors. The test has no sex bias (M.T. Sarno, Buonagura, and Levita, 1985). Scoring is on a 9-point scale, and ratings are assigned on the basis of the examiner's estimate of the patient's premorbid ability in that area. Scores are recorded on a histogram. Sarno (1969) recommended color coding to differentiate the initial evaluation from subsequent reevaluations for easy visual review. She also offered a rather loose method of converting the item grades into percentages that may be too subjective for research purposes or for comparisons with clinical evaluations made by different examiners. However, this test is of practical value in predicting functional communication (Spreen and Risser, 2003) and for documenting poststroke improvement (M.T. Sarno, 1976).

Multilingual Aphasia Examination (MAE) (Benton and Hamsher, 1989; Benton, Hamsher, and Sivan, 1994)

A seven-part battery was developed from its parent battery, the Neurosensory Center Comprehensive Examination of Aphasia (see below), to provide for a systematic, graded examination of receptive, expressive, and immediate memory components of speech and language functions. The Token Test and Controlled Oral Word Association are variations of tests in general use; others, for instance the three forms of the Spelling test (Oral, Written, and Block—using large metal or plastic letters), were developed for this battery. Most of the tests have two or three forms, thus reducing practice effects on repeated administrations. For each test, age and education effects are dealt with by means of a *Correction Score,* which, when added to the raw score gives an *Adjusted Score.* Percentile conversions for each adjusted score and their corresponding classification have been worked out so that scores on each test are psychometrically comparable. This means of scoring and evaluating subtest performances has the additional virtue of allowing each test to be used separately as, for instance, when an examiner wishes to study verbal fluency or verbal memory in a patient who is not aphasic and for whom administration of many of the other subtests would be a waste of time. A Spanish version of this test (MAE-S) is available (G.J. Rey and Benton, 1991). Most of these tests are both age and education sensitive; the effects of age and education have been reported for many of them (Ivnik, Malec, Smith, et al., 1996; Mitrushina, Boone, and D'Elia, 1999; Ruff, Light, and Parker, 1996).

Neurosensory Center Comprehensive Examination for Aphasia (NCCEA)[1] (Spreen and Benton, 1977; Spreen and Strauss, 1991)

This battery consists of 24 short subtests, 20 involving different aspects of language performance, and four "control" tests of visual and tactile functions. Most of the subtests normally take less than five minutes to administer. The control tests are given only when the patient performs poorly on a test involving visual or tactile stimuli. A variety of materials are used in the tests, including common objects, sound tapes, printed cards, a screened box for tactile recognition, and the Token Test "tokens." An interesting innovation enables patients whose writing hand is paralyzed to demonstrate "graphic" behavior by giving them "Scrabble" letters for forming words. All of the materials can be easily purchased, or they can be constructed by following instructions in the manual. Age and education corrected scores for each subtest are entered on three profile sheets, one providing norms for the performance of intact but poorly educated adults, a second with norms based on the performance of aphasic patients, and the third giving performance data on nonaphasic brain damaged patients. The first two profiles taken together enable the examiner to identify patients whose performance differs significantly from that of normal

[1]This battery may be obtained from the University of Victoria Neuropsychology Laboratory, P.O. Box 1700, Victoria, British Columbia, V8W 3P4, Canada.

adults, while providing for score discriminations within the aphasic score range so that small amounts of change can be registered.

This test has proven sensitivity, particularly for moderately and severely aphasic patients (Spreen and Risser, 2003) and also for distinguishing kinds and degrees of speech and language impairments after head injury (Sarno, 1980). It suffers from a low ceiling which diminishes its usefulness for examining well-educated patients with mild impairments (Spreen and Risser, 2003), and it omits assessment of spontaneous speech.

Porch Index of Communicative Ability (PICA) (Porch, 1983)

The PICA was developed as a highly standardized, statistically reliable instrument for measuring a limited sample of language functions. This battery contains 18 ten-item subtests, four of them verbal, eight gestural, and six graphic. The same ten common items (cigarette, comb, fork, key, knife, matches, pencil, pen, quarter, toothbrush) are used for each subtest with the exception of the simplest graphic subtest in which the patient is asked to copy geometric forms. Spontaneous conversation is not addressed. The examiner scores each of the patient's responses according to a 16-point multidimensional scoring system (Porch, 1971). Each point in the system describes performance. For example, a score of 1 indicates no response; a score of 15 indicates a response that was judged to be accurate, responsive, prompt, complete, and efficient. Qualified PICA testers undergo a 40-hour training period after which they administer ten practice tests. This training leads to high interscorer reliability correlation coefficients. Its validity as a measure of language and communication ability has been demonstrated (Spreen and Risser, 2003).

By virtue of its tight format and reliable scoring system, the PICA provides a sensitive measure of small changes in patient performance. This sensitivity can aid the speech pathologist in monitoring treatment effects so long as the patient's deficits are not so mild that they escape notice because of the test's low ceiling. Its statistically sophisticated construction and reliability make it a useful research instrument as well (McNeil, 1979). A.D. Martin (1977) called into question a number of aspects of the PICA, such as the assumption that the scaling intervals are equal, which can lead a score-minded examiner to misinterpret the examination, particularly with respect to the patient's capacity for functional communication. Auditory comprehension is not adequately examined by these procedures (Kertesz, 1989; Spreen and Risser, 2003). Thus, while some aphasia syndromes may be indicated by PICA findings, the data it generates are too limited for making diagnostic classifications or inferences about underlying structural damage (A.G. Davis, 1993).

Western Aphasia Battery (WAB) (Kertesz, 1979, 1982)

This battery grew out of efforts to develop an instrument from the Boston Diagnostic Aphasia Examination that would generate diagnostic classifications and be suitable for both treatment and research purposes. Thus, many of the items were taken from the Boston examination. The Western Aphasia Battery consists of four oral language subtests—*spontaneous speech, auditory comprehension, repetition,* and *naming*—that yield five scores based on either a rating scale (for Fluency and Information content of speech) or conversion of summed item-correct scores to a scale of 10. Each score thus can be charted on a ten-point scale; together, the five scores, when scaled, give *a profile of performance.* An *Aphasia Quotient* (*AQ*) can be calculated by multiplying each of the five scaled scores by 2 and summing them. Normal (i.e., perfect) performance is set at 100. The AQ gives a measure of discrepancy from normal language performance, but like any summed score in neuropsychology, it tells nothing of the nature of the problem. The profile of performance and the AQ can be used together to determine the patient's diagnostic subtype according to pattern descriptions for eight aphasia subtypes. In addition, tests of reading, writing, arithmetic, gestural praxis (i.e., examining for apraxia of gesture), construction, and Raven's Progressive Matrices are included to provide a comprehensive survey of communication abilities and related functions. Scores on the latter tests can be combined into a *Performance Quotient* (*PQ*); the AQ and PQ together give a summary *Cortical Quotient* (*CQ*) score for diagnostic and research purposes. The language portions of the test take about one and one-half hours, and less time with more impaired or particularly fluent patients. Reliability and validity evaluations meet reasonable criteria. Its statistical structure is satisfactory (Spreen and Risser, 2003).

Only the two scores obtained by ratings should present standardization problems. However, the other items leave little room for taking the qualitative aspects of performance into account and thus may provide a restricted picture of the patient's functioning which may account for some of the reported disparities between diagnostic decisions made by clinicians or generated by other aphasia tests and diagnostic classifications based on WAB data (e.g., see A.G. Davis, 1993). Another drawback is that the classification system does not address the many patients whose symptoms are of a "mixed" nature (i.e., have components of more than one of the eight types delineated in this classification system) (Spreen and Risser, 2003).

The WAB has been used to evaluate the language abilities of patients with a variety of neurological diseases. Patients with right hemisphere strokes performed as well as control subjects on all five scales while those with strokes on the left were significantly impaired (K.L. Bryan and Hale, 2001). Early language impairment in patients with primary progressive aphasia was detected on items involving fluency and naming, while comprehension and nonverbal cognition were retained (Karbe et al., 1993). Lower fluency, repetition, and naming scores distinguished left hemisphere stroke patients from patients with mild Alzheimer's disease; however, those with right hemisphere strokes could not be distinguished on any AQ measures (J. Horner, Dawson, et al., 1992). Patients with vascular dementia performed worse than Alzheimer patients on the writing scale while the latter scored lower on the repetition scale (Kertesz and Clydesdale, 1994).

Aphasia Screening

Aphasia screening tests do not replace the careful examination of language functions afforded by the test batteries. Rather, they are best used as supplements to a neuropsychological test battery. They signal the presence of an aphasic disorder and may even call attention to its specific characteristics, but they do not enable the examiner to make either a reliable diagnosis or the fine discriminations required for understanding the manifestations of an aphasic disorder. These tests do not require technical knowledge of speech pathology for satisfactory administration or determination of whether a significant aphasic disorder is present. However, excepting the Token Test which can elicit subtle deficits, conversations with the patient coupled with a mental status examination should, in most cases, make an aphasia screening test unnecessary. "All we need is a concept of what needs to be assessed, a few common objects, a pen, and some paper" (A.G. Davis, 1993, p. 215). Davis considers screening tests to be useful to the extent that "a standardized administration maximizes consistency in diagnosis, supports a diagnosis, and facilitates convenient measurement of progress" (p. 215).

Aphasia Screening Test (Halstead and Wepman, 1959)

This test was created by Wepman, not Halstead; Wepman was a member of Halstead's department at the time the article describing it was published. This is the most widely used of all aphasia tests since it or one of its variants has been incorporated into many formally organized neuropsychological test batteries. As originally devised, the Aphasia Screening Test has 51 items which cover all the elements of aphasic disabilities as well as the most common associated communication problems. It is a fairly brief test, rarely taking longer than 30 minutes to complete. There are no rigid scoring standards, but rather, the emphasis is on determining the nature of the linguistic problem once its presence has been established. Erroneous responses are coded into a diagnostic profile intended to provide a description of the pattern of the patient's language disabilities. Obviously, the more areas of involvement and the more a single area is involved, the more severe the disability. However, no provisions are made to grade test performance on the basis of severity, nor information provided for classifying patients, nor are guidelines given for clinical application.

Wepman (personal communication, 1975 [mdl]) rejected this test about 30 years after he had developed it, as he found that it contributed more confusion than clarity to both diagnosis and description of aphasic disorders. Aphasia and related conditions require more than an item or two to be identified and understood within the totality of the patient's communication abilities.

Reitan included it in the Halstead-Reitan Battery along with tests developed by Halstead and others. Reitan pared down the original test to 32 items but still handled the data descriptively, in much the same manner as originally intended (Jarvis and Barth, 1994; Reitan and Wolfson, 1993). A second revision of the Aphasia Screening Test appeared in E.W. Russell, Neuringer, and Goldstein's (1970) modification of Reitan's modification of Halstead's battery. This version, called the *Aphasia Test,* contains 37 items. It is essentially the same as Reitan's modification except that four easy arithmetic problems and the task of naming a key were added. E.W. Russell and his colleagues established a simple error-counting scoring system for use with their computerized diagnostic classification system, which converts to a 6-point rating scale. Other scoring systems have been developed typically based on a number correct (or error) score in which each item is evaluated on a "right" or "wrong" basis (W.G. Snow, 1987b).

In his item-by-item comparisons of responses made by 50 patients with lateralized lesions, W.G. Snow (1987a) found that only one item—copying the drawing of a key—discriminated the two groups: significantly more patients with right hemisphere disease made errors on this item than those with left-sided lesions. By and large, patients with left-sided damage did worse on verbal items; those with damage on the right had poorer performances on naming the triangle, on drawing, and on reading "7 SIX 2." More than half of a group of normal elderly (ages 65–75) subjects failed one or more of the repetition items, at least one drawing item was failed by a similar number, and more than one-third of this group failed at least one item classified as measuring language comprehension (Ernst, 1988). Additionally, significant correlations between this test and both mental ability and education have been recorded (Spreen and Risser, 1991). Thus, if one goes by score alone, this test cannot qualify for aphasia screening. Moreover, the manner in which it is presented to examiners allows naive ones to ascribe very serious neuropsychological deficits to a single error, such as reporting "acalculia" on the basis of a patient's inability to multiply 27 × 3 mentally (usually reflecting an attention disorder!) or interpreting the careless drawing of a key as "constructional apraxia." Ridiculous as it seems, I [mdl] have seen such crude and potentially harmful interpretations many times when reviewing examination protocols and reports. Probably the best way of handling this test is Wepman's: *junk it altogether*.

A very shortened version of Wepman's Aphasia Screening Test consists of four tasks (Heimburger and Reitan, 1961):

1. Copy a square, Greek cross, and triangle without lifting the pencil from the paper.
2. Name each copied figure.
3. Spell each name.
4. Repeat: "He shouted the warning"; then explain and write it.

This little test may aid in discriminating between patients with left and right hemisphere lesions, for many of the former can copy the designs but cannot write, while the latter have little trouble writing but many cannot reproduce the designs.

Token Test (Boller and Vignolo, 1966; De Renzi and Vignolo, 1962)

The Token Test is extremely simple to administer, to score and, for almost every nonaphasic person who has completed the fourth grade, to perform with few if any errors. Yet it is remarkably sensitive to the disrupted linguistic processes that are central to the aphasic disability, even when much of the patient's communication behavior has remained intact. Scores on the Token Test correlate highly both with scores on tests of auditory comprehension (Morley et al., 1979) and with language production test scores (Gutbrod et al., 1985). The Token Test performance also involves immediate memory span for verbal sequences and capacity to use syntax (Lesser, 1976). It can identify those brain damaged patients whose other disabilities may be masking a concomitant aphasic disorder, or whose symbolic processing problems are relatively subtle and not readily recognizable. However, it contributes little to the elucidation of severe aphasic conditions since these patients will fail most items quite indiscriminately (Wertz, 1979).

Twenty "tokens" cut from heavy construction paper or thin sheets of plastic or wood make up the test material. They come in two shapes (circles and squares[1]), two sizes (big and little), and five colors. The tokens are laid out horizontally in four parallel rows of large circles, large squares, small circles, and small squares, with colors in random order (e.g., see De Renzi and Faglioni, 1978). The only requirement this test makes of the patient is the ability to comprehend the token names and the verbs and prepositions in the instructions. The diagnosis of those few patients whose language disabilities are so severe as to prevent them from cooperating on this task is not likely to depend on formal testing; almost all other brain injured patients can respond to the simple instructions. The test consists of a series of oral commands, 62 altogether, given in five sections of increasing complexity (Table 13.1).

Examiners must guard against unwittingly slowing their rate of speech delivery as slowed presentation of instructions (*stretched speech* produced by slowing an instruction tape) significantly reduced the number of errors made by aphasic patients without affecting the performance of patients with right hemisphere lesions (Poeck and Pietron, 1981). However, even with slowed instructions, aphasic patients still make many more errors than do patients with right-sided lesions.

Items failed on a first command should be repeated and, if performed successfully the second time, scored separately from the first response. When the second, but not the first, administration of an item is passed, only the second performance is counted, under the assumption that many initial errors will result from such nonspecific variables as inattention and disinterest. Each correct response earns 1 point on the 62-point scale. The examiner should note whether the patient distinguishes between the Part 5 "touch" and "pick up" directions.

Boller and Vignolo (1966) developed a slightly modified version of De Renzi and Vignolo's (1962) original Token Test format. Their cut-off scores correctly classified 100% of the control patients, 90% of patients with right-hemisphere lesions, and 91% of apha-

[1]When originally published, instructions called for rectangles. Squares have been universally substituted to reduce the number of syllables the patient must process.

TABLE 13.1 The Token Test

PART I

(Large squares and large circles only are on the table)

1. Touch the red circle
2. Touch the green square
3. Touch the red square
4. Touch the yellow circle
5. Touch the blue circle (2)*
6. Touch the green circle (3)
7. Touch the yellow square (1)
8. Touch the white circle
9. Touch the blue square
10. Touch the white square (4)

PART II

(Large and small squares and circles are on the table)

1. Touch the small yellow circle (1)
2. Touch the large green circle
3. Touch the large yellow circle
4. Touch the large blue square (3)
5. Touch the small green circle (4)
6. Touch the large red circle
7. Touch the large white square (2)
8. Touch the small blue circle
9. Touch the small green square
10. Touch the large blue circle

PART III

(Large squares and large circles only)

1. Touch the yellow circle and the red square
2. Touch the green square and the blue circle (3)
3. Touch the blue square and the yellow square
4. Touch the white square and the red square
5. Touch the white circle and the blue circle (4)
6. Touch the blue square and the white square (2)
7. Touch the blue square and the white circle
8. Touch the green square and the blue circle
9. Touch the red circle and the yellow square (1)
10. Touch the red square and the white circle

PART IV

(Large and small squares and circles)

1. Touch the small yellow circle and the large green square (2)
2. Touch the small blue square and the small green circle
3. Touch the large white square and the large red circle (1)
4. Touch the large blue square and the large red square (3)
5. Touch the small blue square and the small yellow circle
6. Touch the small blue circle and the small red circle
7. Touch the large blue square and the large green square
8. Touch the large blue circle and the large green circle

TABLE 13.1 (continued)

9. Touch the small red square and the small yellow circle
10. Touch the small white square and the large red square (4)

PART V

(Large squares and large circles only)

1. Put the red circle on the green square (1)
2. Put the white square behind the yellow circle
3. Touch the blue circle with the red square (2)
4. Touch—with the blue circle—the red square
5. Touch the blue circle and the red square (3)
6. Pick up the blue circle or the red square (4)
7. Put the green square away from the yellow square (5)
8. Put the white circle before the blue square
9. If there is a black circle, pick up the red square (6)

N.B. There is no black circle.

10. Pick up the squares, except the yellow one
11. Touch the white circle without using your right hand
12. When I touch the green circle, you take the white square.

N.B. Wait a few seconds before touching the green circle.

13. Put the green square beside the red circle (7)
14. Touch the squares, slowly, and the circles, quickly (8)
15. Put the red circle between the yellow square and the green square (9)
16. Except for the green one, touch the circles (10)
17. Pick up the red circle—no!—the white square (11)
18. Instead of the white square, take the yellow circle (12)
19. Together with the yellow circle, take the blue circle (13)
20. After picking up the green square, touch the white circle
21. Put the blue circle under the white square
22. Before touching the yellow circle, pick up the red square

*A second number at the end of an item indicates that the item is identical or structurally similar to the item of the number in De Renzi and Faglioni's "short version" (see p. 510). To preserve the complexity of the items in Part 5 of the short version, item 3 of the original Part IV should read, "Touch the large white square and the *small* red circle."

From Boller and Vignolo (1966)

sic patients, for an overall 88% correctly classified (see Table 13.2, p. 509).

Part V alone, which consists of items involving relational concepts, identified only one fewer patient as "latent aphasic" than did the whole 62-item test of Boller and Vignolo. This finding suggests that Part V could be used without the other 40 questions to identify those patients with left hemisphere lesions misclassified as nonaphasic because their difficulties in symbol formulation are too subtle to impair communication for most ordinary purposes. Doubling the number of items increased the power of Part II to discriminate between patients with right hemisphere lesions and aphasics to 92.5% (R. Cohen, Gutbrod, et al., 1987).

TABLE 13.2 A Summary of Scores Obtained by the Four Experimental Groups on The Token Test

		BRAIN DAMAGED PATIENTS		
		RIGHT	LEFT	
Partial Scores	*Control Patients* (n = *31*)	(n = *30*)	*Nonaphasic* (n = *26*)	*Aphasic* (n = *34*)
Part I				
10	31	30	26	30
9 & lower				4
Part II				
10	31	29	25	23
9 & lower		1	1	11
Part III				
10	29	28	25	13
9	2	2	1	10
8 & lower				11
Part IV				
10	29	25	21	5
9	2	3	3	4
8 & lower		2	2	25
Part V				
20 and above	28	22	14	3
18 & 19	3	7	5	2
17 & lower		1	7	29
Total score				
60 & above	26	21	14	2
58–59	5	6	4	1
57 & lower		3	8	31

Adapted from Boller and Vignolo (1966)

Test characteristics. Age effects have been documented (De Renzi and Faglioni, 1978; Ivnik, Malec, Smith, et al., 1996; Spreen and Strauss, 1998). Although De Renzi and Faglioni (1978) reported education effects, Spreen and Strauss (1998) suggest that age corrections are unnecessary for persons with greater than eight years of education. Correlations with general mental ability (as measured by Raven's Matrices) become apparent only with brain impaired patients (Coupar, 1976). Men and women perform similarly (M.T. Sarno, Buonaguro, and Levita, 1985). Test-retest reliability was high with correlation coefficients between .92 and .96 when measured on aphasic patients (Spreen and Strauss, 1998); with intact elderly persons who make very few errors, the reliability coefficient was only .50 after a year's interval (W.G. Snow, Tierney, Zorzitto, et al., 1988). Practice effects measured on patients with no intervention and no degenerative disease are virtually nil (McCaffrey, Duff, and Westervelt, 2000b). Validation of its sensitivity to aphasia comes from a variety of sources (Spreen and Risser, 2003).

Neuropsychological findings. Despite the simplicity of the called-for response—or perhaps because of its simplicity—this direction-following task can give the observant examiner insight into the nature of the patient's comprehension or performance deficits. Patients whose failures on this test are mostly due to defective auditory comprehension tend to confuse colors or shapes and to carry out fewer than the required instructions. They may begin to perseverate as the instructions become more complex. A few nonaphasic patients may also perseverate on this task because of conceptual inflexibility or an impaired capacity to execute a series of commands.

For example, although he could repeat the instructions correctly, a 68-year-old retired laborer suffering multi-infarct dementia was unable to perform the two-command items because he persisted in placing his fingers on the designated tokens simultaneously despite numerous attempts to lead him into making a serial response.

This clinical observation was extended by a study of a group of dementia patients who performed considerably *below normal limits* on a 13-item form of this test (Swihart, Panisset, et al., 1989). These patients did best on the first simple command, "Put the red circle on the green square," with high failure levels (56% and 57%) on the two following items because of tendencies to perseverate the action "Put on" when these subsequent item instructions asked for "Touch." This study found the Token Test to be quite sensitive to dementia severity: it correlated more highly with the Mini-Mental State Examination ($r = .73$) than with an auditory comprehension measure ($r = .49$), indicating that failures were due more to general cognitive deficits than to specific auditory deficits.

When patients have difficulty on this task, the problem is usually so obvious that, for clinical purposes, the examiner may not find it necessary to begin at the beginning of the test and administer every item. To save time, the examiner can start at the highest level at which success seems likely and move to the next higher level if the patient easily succeeds on three or four items. When a score is needed, as for research purposes or when preparing a report that may enter into litigation proceedings, the examiner may wish to use one of the several short forms.

Token Test variants. Spreen and Benton developed a 39-item modification of De Renzi and Vignolo's long form, which is incorporated in the Neurosensory Center Comprehensive Examination for Aphasia (reproduced in Spreen and Strauss, 1998). From this, Spellacy and Spreen (1969) constructed a 16-item short form that uses the same 20 tokens as both the original and the modified long forms and includes many of the relational items of Part V. A 22-item Token Test is part of Benton, Hamsher, and Sivan's Multilingual Aphasia Examination battery. The first ten items contain representative samples from sections I to IV of the original test; the last 11 items involve the more complex relational concepts found in the original section V. A 16-item short form identified 85% of the aphasic and 76% of the nonaphasic brain damaged patients, screening as well as Part V of the 62-item long form but not quite as well as the entire long form. These data suggest that, for screening, either Part V or a short form of the Token Test will usually be adequate. Patients who achieve a borderline score on one of these shorter forms of the test should be given the entire test to clarify the equivocal findings. Age-corrected norms have been developed for the MAE version (Ivnik, Malec, Smith, et al., 1996).

TABLE 13.3 Adjusted Scores and Grading Scheme for the "Short Version" of the Token Test

CONVERSION OF RAW SCORES TO ADJUSTED SCORES		SEVERITY GRADES FOR ADJUSTED SCORES	
For Years of Education	*Change Raw Scores By*	*Score*	*Grade*
3–6	+1	25–28	Mild
10–12	−1	17–27	Moderate
13–16	−2	9–16	Severe
17+	−3	8 or less	Very severe

Adapted from De Renzi and Faglioni (1978)

A "Short Version" of the Token Test (De Renzi and Faglioni, 1978). This 36-item short version takes half the time of the original test and is therefore less likely to be fatiguing. It differs from others in the inclusion of a sixth section, Part 1, to lower the test's range of difficulty. The new Part 1 contains seven items requiring comprehension of only one element (aside from the command, "touch"); e.g., "1. Touch a circle"; "3. Touch a yellow token"; "7. Touch a white one." To keep the total number of items down, Part 6 has only 13 items (taken from the original Part 5), and each of the other parts, from 2 through 5, contains four items (see the double-numbered items of Table 13.1 and its footnote). On the first five parts, should the patient fail or not respond for five seconds, the examiner returns misplaced tokens to their original positions and repeats the command. Success on the second try earns half a credit. The authors recommend that the earned score be adjusted for education (see Table 13.3). The adjusted score that best differentiated their control subjects from aphasic patients was 29, with only 5% of the control subjects scoring lower and 7% of the patients scoring higher. A scheme for grading auditory comprehension based on the adjusted scores (see Table 13.3) is offered for making practical clinical discriminations. De Renzi and Faglioni reported that scores below 17 did distinguish patients with global aphasia from the higher-scoring ones with Broca's aphasia.

VERBAL EXPRESSION

> . . . sudden fits of inadvertency will surprize vigilance, slight avocations will seduce attention and casual eclipses will darken learning; and that the writer shall often in vain trace his memory at the moment of need, for that which yesterday he knew with intuitive readiness, and which will come uncalled into his thoughts tomorrow.
>
> *Samuel Johnson*

Tests of confrontation naming provide information about the ease and accuracy of word retrieval and may also give some indication of vocabulary level. Individually administered tests of word knowledge typically give the examiner more information about the patient's verbal abilities than just an estimate of vocabulary level. Responses to open-ended vocabulary questions, for example, can be evaluated for conceptual level and complexity of verbalization. Descriptions of activities and story telling can demonstrate how expressive deficits interfere with effective communication and may bring out subtle deficits that have not shown up on less demanding tasks.

Naming

Confrontation naming, the ability to pull out the correct word at will, is usually called *dysnomia* when impaired. The left temporal lobe is essential for this task in most right-handers (Hamberger et al., 2001). Lesions of the posterior superior temporal and inferior parietal regions are associated with semantic paraphasic errors, while lesions of the insula and putamen contribute to phonologic paraphasic errors (Knopman, Selnes, Niccum, and Rubens, 1984). Repetitive transcranial magnetic stimulation over the temporal lobe can facilitate picture naming (Mottaghy et al., 1999). The speech-dominant hippocampus also is a significant component of the overall neuroanatomical network of visual confrontation naming (Sawrie, Martin, et al., 2000). Dysnomia is usually a significant problem for aphasic patients. In its milder form, dysnomia can be a frustrating, often embarrassing problem that may accompany a number of conditions—after a concussion or with multiple sclerosis, for example.

Two months after being stunned with a momentary loss of consciousness when her car was struck from behind, a very bright doctoral candidate in medical sociology described her naming problem as "speech hesitant at times—I'm trying to explain something and I have a concept and can't attach a word to it. I know there's something I want to say but I can't find the words that go along with it."

In neurological examinations, confrontation naming is typically conducted with body parts and objects beginning with the most frequently used terms (e.g., *hand, pen*) and then asking for the name of the parts, thus going from the most frequently used name to names less often called upon in natural speech (e.g., *wrist* or *joint, cap* or *clip*) (e.g., Strub and Black, 2000). In formal aphasia and neuropsychological assessment, pictures are the most usual stimulus for testing naming facility. The examination of patients with known or suspected aphasia may also include tactile, gestural, and nonverbal sound stimuli to evaluate the naming process in response to the major receptive channels (Rothi, Raymer, et al., 1991).

Kremin (1988), noting that most confrontation naming tasks assess only nouns, recommended asking for verbs and prepositions to delineate the nature of the naming deficit for more accurate diagnosis. Identifying activities, shown in line drawings, with the appropriate verb appears to be a slightly easier task for intact adults than naming objects (M. Nicholas, Obler, Albert, and Goodglass, 1985). A little loss of retrieval efficiency for older adults in the 70s was documented on this task. The Boston Diagnostic Aphasia Examination has a number of activity pictures for just this purpose. Picture sets containing only very common objects are unlikely to prove discriminating when examining suspected or early dementia patients (Bayles and Tomoeda, 1983; Kaszniak, Wilson, et al., 1986).

For picture naming, Snodgrass and Vanderwart's 1980 set of 260 pictures has norms for "name agreement, image agreement, familiarity, and visual complexity." A.W. Ellis and his colleagues (1992) provided a list of 60 picture items taken from the Snodgrass and Vanderwart collection, arranged both according to frequency of occurrence in English and in sets of three. Each word in a set contains the same number of syllables but differs according to its frequency (high, medium, low), thus enabling the examiners to make up naming tasks suitable for particular patients or research questions. The vulnerability of object names to retrieval failure is related to the age of acquisition of the names, with later acquisition (usually less commonly used words) associated with more errors (B.D. Bell, Davies, Hermann, and Walters, 2000; Hodgson and Ellis, 1998).

Boston Naming Test (BNT) (E.F. Kaplan, Goodglass, and Weintraub, 1983; Goodglass and Kaplan, 2001)

This test consists of 60 large ink drawings of items ranging in familiarity from such common ones as "tree" and "pencil" at the beginning of the test to "sphinx" and "trellis" near its end. Adults begin with item 30 and proceed forward unless they make a mistake in the first eight items, at which point reverse testing is continued until eight consecutively correct responses are obtained. The test is discontinued after eight consecutive failures. When giving this test to patients with dementia or suspected dementia, K. Wild (personal communication, 1992 [mdl]) recommends the following instructions: "I'm going to show you some pictures and your job is to tell me the common name for them. If you can't think of the name and it's something you know you can tell me something you know about it." She advises that semantic cueing be conservative to assess for perceptual errors. When patients are unable to name a drawing, the examiner gives a semantic cue; if

(E. Kaplan, Fein, et al., 1991). Among each set of choices are one 2-point definition, a 1-point definition, and three 0-point definitions, including one that is phonetically similar to the test item word. This format is particularly helpful for patients with word retrieval problems who can recognize but not bring up spontaneously the correct definition.

Paper-and-pencil vocabulary tests

Single paper-and-pencil vocabulary tests are rarely used. Most of the time, the assessment of vocabulary takes place as part of an academic aptitude test battery, a reading test battery, or one of the multiple test guidance batteries. One single vocabulary test that has been used in numerous neuropsychological studies is the 80-word Mill Hill Vocabulary Scale (Raven, 1982). This multiple-choice test takes relatively little time to administer and is easily scored. Mill Hill raw scores convert to percentiles and a standard score (expressed as a "deviation IQ" score [Raven et al., 1976]) for age levels from 20 to 65. This well-standardized test has proven sensitivity to left hemisphere disease (L.D. Costa and Vaughan, 1962) and to dementia (R.G. Morris and Kopelman, 1992). Performance on the Mill Hill was only slightly (5 IQ score points) but significantly diminished in a group of TBI patients mostly tested within six months of injury (D.N. Brooks and Aughten, 1979). No Mill Hill score differences were found between groups of elderly patients with and without diffuse brain disease (Irving, 1971).

The *Gates-MacGinitie Reading Tests* (GMG) are well-suited for clinical evaluations of vocabulary level as they have both a vocabulary and a reading comprehension test presented in a four-choice format (see p. 523). The most recent edition expands the number of formats by including one for adults in addition to the senior high school norms which are applicable for many adult patients.

For research purposes, the *Verbal Comprehension* test of the *Employee Aptitude Survey* (EAS) (Ruch et al., 1963) provides a quick, simple means for assessing vocabulary. The four-choice format contains 30 words ranging in difficulty level from "keen" to "prolix," thus sampling a more mature vocabulary range than similar tests. L.M. Binder, Tanabe, et al., 1982 found that scores remained stable during treatment for cerebrovascular disorders, although some other measures showed improvement with middle cerebral artery bypass. I [mdl] included it in a study on mental efficiency with type II diabetes mellitus under the assumption that, as a vocabulary test, it would be relatively unaffected by disease severity: comparisons were made between one group whose members had newly identified diabetes, another of diagnosed and treated diabetics, and a healthy control group (U'Ren et al., 1990). However, contrary to expectations, scores on this test did reflect diabetes severity ($p < .001$), which suggested that selecting definitions for these mostly abstract words involves a significant amount of conceptual prowess, at least for persons within the 67–77 year age range.

Nonverbal response vocabulary tests

Vocabulary tests in which patients signal that they recognize a spoken or printed word by pointing to one of a set of pictures permit evaluation of the recognition vocabulary of many verbally handicapped patients. These tests are generally simple to administer. They are most often used for quick screening and for estimating the general ability level of intact persons when time or circumstances do not allow a more complete examination. Slight differences in the design and in standardization populations of the picture vocabulary tests in most common use affect their appropriateness for different patients to some extent.

Peabody Picture Vocabulary Test (PPVT-III) (L.M. Dunn and Dunn, 1997)

This easily administered vocabulary test has been standardized for ages $2^1/_2$ to 90+. It consists of 204 picture plates, each with four pictures, one plate for each word in the two reasonably equivalent test forms with the words arranged in order of difficulty. The subject points to or gives the number of the picture most like the stimulus word, which is spoken by the examiner or shown on a printed card. The simplest words are given only to young children and obviously retarded or impaired adults. The PPVT items span both very low levels of mental ability and levels considerably above average adult ability. Care should be taken to enter the word list at the level most suitable for the subject so that both basal (the highest six consecutive passes) and ceiling (six failures out of eight) scores can be obtained with minimal effort. Points for passed items are simply counted and entered into tables giving a standard score equivalent, percentile rank, stanine, and an age equivalent score. A Spanish version is available from the PPVT publisher.

The standardization for the current revision of the PPVT is based on a sample of 2,725 subjects drawn from different regions and occupational groups according to representation in the 1994 U.S. Census. Split-half and alternate form reliabilities were .94 (L.M. Dunn and Dunn, 1997). A study of adults found correlations of the PPVT-R (L.M. Dunn and Markwardt,

1981) with the WAIS-R VIQ score of .82 and .78 for Forms L and M respectively, with much lower correlations with the PIQ score (.46, .38) (Stevenson, 1986). Correlational studies between the original 1965 version of the PPVT and a number of other cognitive tests plus education found WAIS-R Vocabulary to be the only important contributor to PPVT variance (J.K. Maxwell and Wise, 1984). This research appears to reflect the essentially verbal nature of this test. Stevenson (1986) also found that PPVT-R mean scores ran consistently lower than did WAIS-R summation scores. The next edition, PPVT-III, underestimated the *superior* WIS-A scores in one study of college students (N.L. Bell et al., 2001).

Since administration begins at a level near that anticipated for a subject, this test goes quickly and, as such, may be a useful instrument for estimating mental ability levels generally. Although PPVT scores are often interpreted as representing premorbid intelligence, patients with lesions of the left hemisphere may have difficulty with this test (A. Smith, 1997). For severely impaired patients, particularly when their ability to communicate has been compromised, this test may give the examiner the best access to the patient's residual vocabulary and fund of information. In addition, the simplicity of the pictures makes it eminently suitable for those brain damaged patients who have so much difficulty sorting out the elements in a complex stimulus that they are unable to respond to the intended problem.

Quick Test (Ammons and Ammons, 1962)[1]

Although billed as an intelligence test from which IQ scores can be derived, this 50-item test primarily examines vocabulary (Swartz, 1985)—but vocabulary used in situational contexts. The subject is shown a card with four pictures: one, for example, depicting a traffic policeman with a whistle to his mouth guarding children on the way to school. As the examiner reads words from the list the subject points to the appropriate picture (e.g., "belt," "pedestrian," and "imperative" go with the policeman picture). Words are scaled in difficulty from "easy," ages six through 18+, to "hard." Its three forms are roughly equivalent. Based on data from ten studies, median correlations with the WAIS VIQ, Information, and Vocabulary tests were .82, .82 and .83 (Feingold, 1982). This test may underestimate the mental ability of the brightest subjects but is quite accurate for persons in the *average* ability ranges (Traub and Spruill, 1982). M.B. Acker and Davis (1989) found that scores on this test contributed significantly to predictions of outcome for TBI patients almost four years later as measured by both degree of independence and level of community activity. Taken together, these studies recommend the Quick Test for rapid screening of verbal ability.

[1]This test can be obtained from Psychological Test Specialists, Box 9229, Missoula, MT 59807.

Discourse

Story telling

Pictures are good stimuli for eliciting usual speech patterns. The *Cookie Theft* picture from the Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1983b) is excellent for sampling propositional speech since the simple line drawing depicts familiar characters (e.g., mother, mischievous boy) engaged in familiar activities (washing dishes) in a familiar setting (a kitchen). Patients' stories about this picture can help differentiate the types of language impairment of different aphasic groups (Ardila and Rosselli, 1993). Alzheimer patients have difficulty in describing the central meaning of stories and tend to focus on less important details (S.B. Chapman et al., 1995).

Describing activities

Open-ended questions about patients' activities or skills also elicit samples of their normal speech. I [mdl] have asked patients to describe their work (e.g., "Tell me how you operate a drill press"), a behavior day ("Beginning with when you get up, tell me what you do all day"), or their plans (see *Script Generation* [pp. 620–621] for a formalized procedure to elicit patients' descriptions of familiar activities). While these questions may enable the examiner to learn about the patient's abilities to plan and carry out activities, they do not allow for much comparison between patients (e.g., How do you compare a farmer's description of his work with that of a sawmill worker who pulls logs off of a conveyor belt all day?). Moreover, the patient's work may be so routine or work plans so ill-formulated that the question does not elicit many words. De Renzi and Ferrari (1978) solved the problem of comparability for their Italian population by asking men to describe how they shave and women how to cook spaghetti. "Tell me how to make scrambled eggs" is a counterpart of the spaghetti question that most Americans can answer. L.L. Hartley and Jensen (1991) instructed their patients to explain how to buy groceries in an American supermarket. I [mdl] ask patients what they like to cook and then have them tell me how to make it, or I may ask men to describe how to change a tire. Borod, Rorie, and their colleagues (2000) asked patients to recollect emotional and nonemotional experiences. Interestingly, emotional content enhanced discourse of left hemisphere lesioned patients and suppressed performance when the lesion was on the right.

Verbal Fluency

Following brain injury, many patients experience changes in the speed and ease of verbal production. Greatly reduced verbal productivity accompanies most aphasic disabilities, but it does not necessarily signify the presence of aphasia. Impaired verbal fluency is also associated with frontal lobe damage (R.W. Butler, Rorsman, et al., 1993; Janowsky, Shimamura, Kritchevsky, and Squire, 1989), particularly the left frontal lobe anterior to Broca's area (Baldo, Shimamura, Delis et al., 2001; B. Milner, 1975; Tucha et al., 1999). Stuss, Alexander, and their coworkers (1998) found that patients with left dorsolateral and/or striatal lesions were the most significantly impaired on a letter fluency task. Lesions restricted to the inferior medial area of the frontal lobes did not produce impairment. However, patients with superior medial frontal lesions of either left or right hemisphere had moderate impairment. Fluency also diminished with left parietal lesions. Patients with left frontal lesions perform poorly on verb generation tasks as well (Thompson-Schill et al., 1998). In keeping with these clinical data, the frontal lobes show increased activation on imaging studies during fluency tasks (Brannen et al., 2001; Warkentin and Passant, 1993).

Reductions in fluency may occur in patients with diffuse brain injury. In TBI patients, reduced verbal fluency is associated with both measures of severity (coma and PTA duration) and computed tomography suggesting that diffuse axonal injury is a major contributor to the cognitive inflexibility reflected in their poor fluency performances (Vilkki, Holst, Öhman, et al., 1992).

A fluency problem can show up in speech, reading, and writing; generally, it will affect all three activities (Perret, 1974; L.B. Taylor, 1979) and both free and responsive speech (Feyereisen et al., 1986). However, with aging, writing fluency tends to slow down much earlier than speech fluency, which healthy persons maintain well into the 70s (Benton and Sivan, 1984). Problems in word generation are prominent among the verbal dysfunctions of dementia.

Fluency of speech

Fluency of speech is typically measured by the quantity of words produced, usually within a restricted category or in response to a stimulus, and usually within a time limit. Almost any test format that provides the opportunity for unrestricted speech will test its fluency. Fluency has been measured by rate of speech production as well as word counts of spoken responses to pictures, to directed questions, or to questions stimulating free conversation (Feyereisen et al., 1986; L.L. Hartley and Jensen, 1991).

As Estes (1974) suggested, word fluency tests provide an excellent means of finding out whether and how well subjects organize their thinking. He pointed out that successful performance on these tests depends in part on the subject's ability to "organize output in terms of clusters of meaningfully related words." He also noted that word-naming tests indirectly involve short-term memory in keeping track of what words have already been said. Fluency tests requiring word generation according to an initial letter give the greatest scope to subjects seeking a strategy for guiding the search for words and are most difficult for subjects who cannot develop strategies of their own. Examples of effective strategies are use of the same initial consonant (e.g., content, contain, contend, etc.), variations on a word (shoe, shoelace, shoemaker), or variations on a theme (sew, stitch, seam). Fluency tests calling for items in a category (e.g., animals, what you find in a grocery store) provide the structure lacking in those asking for words by initial letter. However, even within categories, subjects to whom strategy-making comes naturally will often develop subcategories for organizing their recall. For example, the category "animals" can be addressed in terms of domestic animals, farm animals, wild animals, or birds, fish, mammals, etc.

Laine (1988) defined two kinds of conceptual clustering appearing as two or more successive words with similar features: *phonological clusters* share the same initial sound group for letter associates (*sal*ute, *sal*vage for *S*) or the same initial sound for animals (*b*aboon, *b*eaver); and *semantic clusters* in which meanings are either associated (soldier, salute) or shared (salt, sugar). When a cluster is exhausted, the subject must efficiently switch to a new one (Troyer, Moscovitch, and Winocur, 1997).

Age (particularly for persons over 70), sex, and education have been found to influence performance on these tests (Benton, Hamsher, and Sivan, et al., 1994), with women's performances holding up increasingly better than men's after age 55. Some studies have found no age differences on letter fluency tasks (D. Hughes and Bryan, 2002) but positive age effects appear on semantic fluency, e.g., "animals" (Troyer, 2000). In Troyer's study, advancing age was associated with slightly larger cluster sizes and fewer category switches. In evaluating fluency performances, premorbid ability levels need also be taken into account (Crawford, Moore, and Cameron, 1992), and especially educational and vocational accomplishments.

Controlled Oral Word Association (COWA) (Benton and Hamsher, 1989; Spreen and Strauss, 1998)

Benton and his group have systematically studied the oral production of spoken words beginning with a designated letter. The associative value of each letter of the alphabet, except *X* and *Z*, was determined in a normative study using normal control subjects (Borkowski et al., 1967; see Table 13.6). Control subjects of low ability tended to perform a little less well than brighter brain impaired patients. These findings highlight the necessity of taking the patient's premorbid verbal skill level into account when evaluating verbal fluency (Crawford, Moore, and Cameron, 1992).

The Controlled Oral Word Association test (first called the *Verbal Associative Fluency Test* and then the *Controlled Word Association Test*) consists of three word-naming trials. The set of letters that were first employed, F-A-S, has been used so extensively that this test is sometimes labelled "F-A-S." The version developed as part of Benton and Hamsher's (1989) Multilingual Aphasia Examination provides norms for two sets of letters, C-F-L and P-R-W. These letters were selected on the basis of the frequency of English words beginning with these letters. In each set, words beginning with the first letter of these two sets have a relatively high frequency, the second letter has a somewhat lower frequency, and the third letter has a still lower frequency. In keeping with the goal of developing a multilingual battery for the examination of aphasia, Benton and Hamsher also give the frequency rank for letters in French, German, Italian, and Spanish. For example, in French the letters P-F-L have values comparable to C-F-L. The COWA is one of three tests in the *Iowa Screening Battery for Mental Decline* (Eslinger, Damasio, and Benton, 1984; see p. 691). The FAS version is part of the Neurosensory Center Comprehensive Examination for Aphasia.

To give the test, the examiner asks subjects to say as many words as they can think of that begin with the given letter of the alphabet, excluding proper nouns, numbers, and the same word with a different suffix.

TABLE 13.6 Verbal Associative Frequencies for the 14 Easiest Letters

	WORDS/MINUTE		
	9–10	*11–12*	*>12*
Letters	A C D G	B F L M	P
	H W	R S T	

From Borkowski et al. (1967)

TABLE 13.7 Controlled Oral Word Association Test: Adjustment Formula for Males (M) and Females (F)

	AGE (YEARS)					
Education	*25–54*		*55–59*		*60–64*	
(Years Completed)	*M*	*F*	*M*	*F*	*M*	*F*
<9	9	8	11	10	14	12
9–11	6	5	7	7	9	9
12–15	4	3	5	4	7	6
≥16	—	—	1	1	3	3

Adapted from Benton, Hamsher, and Sivan (1994)

The Multilingual Aphasia Battery version also provides for a practice trial using the very high frequency letter "S." The practice trial ends when the subject has volunteered two appropriate "S" words. This method allows the examiner to determine whether the subject comprehends the task before attempting a scored trial. (The practice trial I [mdl] give lasts one minute to provide a genuine "warm-up"). The score, which is the sum of all acceptable words produced in the three one-minute trials, is adjusted for age, sex, and education (see Table 13.7). The adjusted scores can then be converted to percentiles (see Table 13.8). In addition, the examiner counts both errors (i.e., rule violations such as nonwords, proper nouns) and repetitions (noting whether they are repetitions, true perseverations, or variations on the just previously given word, e.g., "look," "looking," the latter word being a rule violation). Repeated words that count as repetitions do not occur successively but are evidence of an impaired ability to generate words and keep track of earlier responses simultaneously. A greater number of words is usually produced early compared to later in the trial. Fernaeus and Almkvist (1998) suggest scoring the first and second halves of each one-minute trial separately. Although this pattern holds for Parkinson patients, the COWA performance that best distinguished them from

TABLE 13.8 Controlled Oral Word Association Test: Summary Table

Adjusted Scores	*Percentile Range*	*Classification*
53+	96+	Superior
45–52	77–89	High normal
31–44	25–75	Normal
25–30	11–22	Low normal
23–24	5–8	Borderline
17–22	1–3	Defective
10–16	<1	Severe defect
0–9	<1	Nil–Trace

Adapted from Benton, Hamser, and Sivan (1976) control subjects was fewer words produced in the first 15 sec (Fama, Sullivan, Shear, et al., 1998).

Spreen and Strauss (1998) give means and standard deviations for different age and educational groups. Tombaugh, Kozak, and Rees (1999) report means and standard deviations for large normal samples ranging from 20 to 89 years with percentiles stratified by age and education. Norms are also available stratfied for education and sex (Ruff, Light, Parker, and Levin, 1996); for age, education, and ethnicity (African American and Caucasian: Gladsjo, Schuman, et al., 1999); and for age and ethnicity (African American, Hispanic, and Caucasian: Johnson-Selfridge et al., 1998). Sumerall and his colleagues (1997) provide data about qualitative errors in an elderly sample (70–95) without neurologic or psychiatric disease: perseverations (23% repeated the same word within 30 sec., 28% repeated the same word after 30 sec, and 40% repeated a word stem with a different ending), breaking set (4.3% gave words in which the first letter differed from the one required), and *proper noun* (13% gave one or more). Only subjects ages 81–95 and with fewer than 15 years' education broke set. Age did not affect productivity. (See Mitrushina, Boone, and D'Elia [1999] for a compilation of earlier norm sets.) Metanorms based on data from 32 studies with a total of 17,625 scores provide a "Summary of aggregate statistics for FAS Totals" giving means and standard deviations by sex, for four age groups (<40, 40–59, 60–79, 80–95) and for two education levels (0–12, >12) (Loonstra et al., 2001). Since variability at lower educational levels tends to be wide (e.g., Loonstra and her colleagues found the SD for 0–12 years = 13.09, for >12 = 12.37), the performances of persons with less education, particularly levels below high school, must be interpreted with caution.

Test characteristics. While mean scores for less educated older subjects slowly slide from a 50–54 year high (which at 41.52 does not differ from younger groups nor from their better educated age peers), means remain about the same for those with 13+ years of schooling until the 75+ years when the mean drops by an apparently nonsignificant amount (Spreen and Strauss, 1998). The performances of men and women do not differ (Ruff, Light, Parker, and Levin, 1996; Sarno, Buonaguro, and Levita, 1985; Zec, Andrise, et al., 1990).

On retesting elderly persons after one year, only the letter *A* (of the FAS set) had a reliability coefficient below .70 or .71, which were the reliability levels for the other letters and the total score, respectively (W.G. Snow, Tierney, Zorzitto, et al., 1988). COWA performance had a moderate correlation with WIS-A Digit Span (.45) and Vocabulary (.41), but practically inconsequential correlations with memory (.17 to .22) and figural fluency (.24) (Ruff, Light, Parker, and Levin, 1997). Comparing the first and second halves of the one-minute productions, initial responses related to Digit Span and memory free recall, while later responses were related to WIS-A Information, Similarities, and Vocabulary (Fernaeus and Almkvist, 1998). These researchers concluded that initial responses depended on rapid access of words from semantic memory with very little effort, while late productions depended on strategies for effortful searching of semantic memory.

Neuropsychological findings. Word fluency as measured by FAS, COWA, and similar techniques calling for generation of word lists has proven to be a sensitive indicator of brain dysfunction. Frontal lesions, regardless of side, tend to depress fluency scores, with left frontal lesions resulting in lower word production than right frontal ones (Miceli et al., 1981; Perret, 1974; Ramier et Hécaen, 1970). Benton (1968) found that not only did patients with left frontal lesions produce on the average almost one-third fewer FAS words than patients with right frontal lesions, but those with bilateral lesions tended to have even lower verbal productivity. Patients with left dorsolateral and superior medial frontal lobe lesions switched categories less frequently but produced normal cluster sizes (Troyer, Moscovitch, Winocur, et al., 1998a). Although both left and right temporal lobe partial resections for seizure control produced declines in COWA performance in the days following surgery, one year later performance exceeded preoperative levels for both groups (Loring, Meador, and Lee, 1994). Left temporal lobe epilepsy (N'Kaoua et al., 2001), multiple sclerosis (Matotek et al., 2001), and mild TBI (Raskin and Rearick, 1996) are often associated with deficits on letter fluency tests. Reduced capacity to generate words has been associated with every dementing process, although the underlying defect tends to vary (Tröster, Fields, et al., 1998; Troyer, Moscovitch, Winocur, et al., 1998b). In some conditions mental inflexibility seems to make an important contribution to the naming disorder (e.g., in some patients with Parkinson's disease); in others, semantic processing and recall abilities are impaired (e.g., Alzheimer's disease). Lexical–phonological functions are compromised in left hemisphere stroke patients. Performances on this test did not differentiate elderly depressed patients from those with diagnosed dementia (R.P. Hart, Kwentus, Taylor, and Hamer, 1988).

Category fluency

Category fluency is less difficult than letter fluency for adults. Whereas elderly control subjects generate about 12 to 16 words/min for letter fluency, animal fluency averages for control subjects range from 20.95 (50 to 59 age range) to 18.96 (70 to 79 age range) (e.g., see Mitrushina, Boone, and D'Elia, 1999). Even control subjects in their 80s produced more animals than FAS words (Koroza and Cullum, 1995). Category fluency declines with age (Fama, Sullivan, Shear, et al., 1998). Using the categories of animals, fruits, and vegetables, normative data stratified by language, age, sex, and education are available for well-educated elder (Lucas et al., 1998a) and for English and Spanish speakers living in the United States (Acevedo et al., 2000). In a study of four ethnic groups, Hispanics and African Americans named the fewest animals, Chinese and Vietnamese the most (Kempler et al., 1998). The authors suggested that variations in word lengths among languages contributed to these findings.

Patients with frontal lobe lesions have reduced letter and category fluency, which is consistent with the theory that they have deficent retrieval strategies (Baldo and Shimamura, 1998). Alzheimer patients have more difficulty with category fluency than letter fluency (Fama, Sullivan, Shear, et al., 1998), an impairment usually attributed to a breakdown in semantic knowledge about categories (Monsch, Bondi, Butters, et al., 1994). Using optimal cut-off scores, category fluency was superior (100% sensitivity, 90.9% specificity) to letter fluency (81.8% sensitivity, 84.1% specificity) in correctly differentiating Alzheimer patients from control subjects. Parkinson patients also have more difficulty with category than letter fluency compared to control subjects (Fama, Sullivan, Shear, et al., 1998). Monsch and her colleagues (1994), finding that Huntington patients were equally impaired on both types of tasks, suggested that their failures were due to reduced general initiation and/or retrieval capacities. However, another study reported that Huntington patients were relatively more impaired on categories (Baldo and Shimamura, 1998).

When both animal and letter naming tasks were used to compare dementia and depression effects on verbal fluency, depressed patients' better animal naming scores distinguished the two patient groups, although even on this easier task the depressed patients' output was inferior to that of the control subjects (R.P. Hart, Kwentus, Taylor, and Hamer, 1988). Compared to control subjects, category production of right brain damaged patients may be a little lower than their letter production and they tend to produce fewer clusters, perhaps due to reduced ability to develop semantic strategies (Joanette et al., 1990).

Other categories have been used to study verbal fluency. Examining the nature of the naming deficits of Parkinson, Huntington, and Alzheimer patients, Randolph, Braun, and their colleagues (1993) used "name things found in a supermarket" along with subcategory cues (e.g., "fruits and vegetables," "things people drink"); the subcategory cueing aided the Parkinson and Huntington patients substantially, but not those with Alzheimer's disease. In addition to the more usual fruit and vegetable naming task, Fuld (1980) asked her elderly subjects to name happy and sad events and found that, contrary to the usual pattern, depressed subjects named more sad than happy events. Other categories that have been used to examine fluency are "types of transportation" and "parts of a car" (Weingartner, Burns, et al., 1984). Studying the effect of set on the verbal productions of patients with Korsakoff's psychosis, Talland (1965a) asked his subjects to "name as many different things as you can that one is likely to see in the street." A 17 person control group (WAIS Vocabulary $M = 10$) averaged 15.7 street sights.

Action fluency

Subjects are instructed to "tell me as many different things as you can think of that people do. I don't want you to use the same word with different endings, like 'eat,' 'eating,' 'eaten.' Also, just give me single words, such as 'eat' or 'smell,' rather than a sentence. Can you give me an example of something that people do?" (Piatt et al., 1999). Parkinson patients were compared with elderly subjects on three fluency tasks: animal naming, FAS, and verb generation. Parkinson patients without dementia and control subjects generated more verbs than FAS words, but patients with dementia had disproportionate difficulty with action fluency. Yet others have reported impaired action fluency in Parkinson patients without dementia (Peran et al., 2003).

Writing fluency

Thurstone Word Fluency Test (TWFT) (L.L. Thurstone and Thurstone, 1962). A written test for word fluency first appeared in the *Thurstones' Primary Mental Abilities* tests (1938, 1962). Subjects must write as many words beginning with the letter S as they can in five minutes, and then write as many four-letter words beginning with C as they can in four minutes. The average 18-year-old can produce 65 words within the nine-minute total writing time. Adult norms are available (Heaton, Grant, and Matthews, 1991). B. Milner (1964, 1975) found that the performance of patients with left frontal lobectomies was significantly impaired on this test relative to that of patients with left temporal lobectomies whose frontal lobes remained intact, and to that of patients whose surgery was confined to the right hemisphere. She observed that this task is more discriminating than object naming fluency tests because the writing task, particularly for C words, is harder. This pattern of relative impairments (frontal output < nonfrontal, left < right hemisphere, left frontal < right frontal) showed up among patients with brain damage due to many different etiologies (Pendleton, Heaton, Lehman, and Hulihan, 1982). Those patients with diffuse damage (trauma and degenerative diseases) performed much like the frontal patients. In a validity study, patients with many kinds of brain injuries performed below control subjects' levels, but the test did not discriminate anterior from posterior lesions, left from right hemisphere lesions, or focal from diffuse lesions; test–retest reliability was high (M.J. Cohen and Stanczak, 2000).

Quantity of writing content

Clinical observations that many patients with right hemisphere damage tend to be verbose led to speculation that these patients may use more words when writing than do other persons (Lezak and Newman, 1979). The number of words used to answer personal and WAIS-type questions, complete the stems of a sentence completion test, and write interpretations to proverbs and a story to Thematic Apperception Test (TAT) card 13MF was counted for 29 patients who had predominantly right hemisphere damage, 15 whose damage was predominantly in the left hemisphere, 25 with bilateral or diffuse damage, and also for 41 control subjects hospitalized for medical or surgical care. On a number of these items, proportionately more patients with predominantly right hemisphere damage gave very wordy responses than patients with other types of brain damage or the control patients. This phenomenon appeared most clearly on the open-ended questions of the sentence-completion test and a personal history questionnaire, neither of which required much conceptual prowess or writing skill. On proverb interpretations and the TAT story, education level played the greatest role in determining response length except for the tendency of the left brain damaged group to give the shortest responses to proverbs.

Quality of writing

At the suggestion of David Spaulding, I [dbh] often ask dementia patients to write "Help keep America clean" on an unlined sheet of paper. This brief writing-to-dictation task gives an opportunity to observe spelling, use of capitalization, and orthographic skills as well as planning in the use of space on the page. More complex tasks offer an opportunity to examine grammar, syntax, and organization of thought processes. Croisile, Ska, and their associates (1996b) compared moderately demented Alzheimer patients' oral and written descriptives of the BDAE Cookie Theft picture, scoring for total number of words and their subtypes (nouns, adjectives, etc.), lexical errors, syntactic complexity, grammatical errors, amount of information, implausible details, and irrelevant comments. Oral descriptions were longer than written ones for both patients and control subjects. Oral descriptions proved to be more sensitive to word-finding difficulty in Alzheimer patients, while written descriptions showed a greater reduction in number of functor words and more implausible details. In addition, Alzheimer patients made more spelling errors.

Speed of writing

Talland (1965a) measured writing speed in two ways: speed of copying a 12-word sentence printed in one-inch type and speed of writing dictated sentences. On the copying task, his 16 control subjects averaged 33.9 seconds for completion, taking less time ($p < .05$) than patients with Korsakoff's psychosis. No significant score differences distinguished control subjects from patients in their speed of writing a single 12-word sentence. However, when writing a 97-word story, read to them at the rate of one to two seconds per word, the control subjects averaged 71.1 words within the three-minute time limit, whereas the patient group's average was 53.1 ($p < .02$). When writing speed has been slowed by a brain disorder, the slowing may become more evident as the length of the task increases. Moreover, the amount of time it took to write the word "television" with the nonpreferred hand differentiated neurologically normal and abnormal schizophrenic patients better than 30 other measures, mostly taken from the standard Halstead-Reitan Battery (G. Goldstein and Halperin, 1977). These investigators acknowledged being at a loss to explain this finding and wondered whether the task's sensitivity might be a function of its midrange level of complexity. Writing times in the range of 6.6 and 5.7 seconds were reported for the schizophrenic patients without neurological disease studied by Goldstein and Halperin and for medicated epileptics (R. Lewis and Kupke, 1992), respectively. Nondominant hand times tend to run just about twice as long as times for the dominant hand, suggesting that pronounced deviations from this pattern may reflect unilateral brain damage.

The Repeatable Cognitive-Perceptual-Motor Battery includes a test of writing speed, *Sentence Writing Time,* which requires subjects to write "The large dog runs fast" (Kelland and Lewis, 1994; R. Lewis, Kelland, and Kupke, 1990). They report a mean writing time of 7.4 ± 1.4 sec for 40 persons (20 of each sex) in the 18– to 30 year range. Writing time ranges for older age groups ran from 7.8 ± 1.3 for 33 subjects 45–59 years old to 11.0 ± 3.3 for 38 subjects age 70 and over. Administration of diazepam to healthy volunteers did not affect their writing time as measured by this test.

VERBAL ACADEMIC SKILLS

With the exception of aphasia tests, surprisingly few neuropsychological batteries contain tests of learned verbal skills such as reading, writing, spelling, and arithmetic. Yet impairment in these commonplace activities can have profound repercussions on a patient's vocational competence and ultimate adjustment. It can also provide clues to the nature of the underlying organic condition.

Reading

Reading may be examined for a number of reasons: to obtain a general appraisal of reading ability in patients without a distinctive impairment of reading skills; to evaluate comprehension of verbal material; for diagnostic purposes, particularly with patients who are aphasic or have significant left hemisphere involvement; or for fine-grained descriptions of very specific deficits for research or treatment purposes. Diagnosis and fine-grained descriptions require specialized knowledge that is usually available from speech pathologists or reading specialists who are also well acquainted with the appropriate test instruments. Cognitive neuropsychologists studying reading aberrations frequently devise their own examination techniques, designed for the specific problem or patient under study (e.g., see Baddeley, Logie, and Nimmo-Smith, 1985; Coslett, 2003; McCarthy and Warrington, 1990; Rapp et al., 2001).

Examiners are cautioned about evaluating reading ability on the basis of the multiple-choice questions for the reading passages in the Boston Diagnostic Aphasia Examination or the Western Aphasia Battery (L.E. Nicholas et al., 1986). Both control subjects and aphasic patients answered considerably more than half the items correctly (far beyond 25% correct by chance) without reading the passages, simply on the basis of inherent meaningfulness. TBI patients earned almost as high scores without reading the BDAE and WAB passages as after reading them (Rand et al., 1990). The paragraph in the Minnesota Test for Differential Diagnosis of Aphasia is so difficult that normal control subjects answered only 80% of the sentences correctly (L.E. Nicholas et al., 1986).

Gates-MacGinitie Reading Tests (GMRT), 4th ed. (MacGinitie, MacGinitie, Maria, and Dreyer, 2000)

These are academic skill tests that lend themselves to neuropsychological assessment. Although these paper-and-pencil multiple-choice tests come in separate forms for each year from Pre-Reading to sixth grade, three will be appropriate for most adults: grade 7/9, grade 10/12, and AR (Adult Reading).

The Gates-MacGinitie tests measure two different aspects of reading. The first subtest, *Vocabulary,* involves simple word recognition. The last subtest, *Comprehension,* measures ability to understand written passages. Both Vocabulary and Comprehension scores tend to be lower when verbal functioning is impaired. When verbal functions remain essentially intact but higher-level conceptual and organizing activities are impaired, a marked differential favoring Vocabulary over Comprehension may appear between the scores of these two subtests. The two tests have generous time limits. They can be administered as untimed tests without much loss of information since most very slow patients fail a large number of the more difficult items they complete outside the standard time limits.

SRA Reading Index (Science Research Associates, 1968)

This multiple-choice reading test provides brief assessments of five levels of reading skill: (1) *Picture–Word Association* (nine items) requires the subject to recognize the word that goes with a picture of a common object (*cow, car*); (2) *Word Decoding* (13 items) asks the subject to identify the one-word definition or description that completes short, incomplete sentences such as, "Apples grow on a . . . "; (3) in *Phrase Comprehension* (13 items) the subject must complete a sentence by choosing the correct phrase among similar phrases which differ in such aspects of grammar as prepositions or adverbs; (4) *Sentence Comprehension* (12 items) presents a sentence with four similar sentences, of which only one gives the target sentence's meaning correctly; (5) *Paragraph Comprehension* (13 items) consists of three sets of explanatory paragraphs (e.g., one gives the rules for a card game), each followed by a number of questions about the material it contains. This untimed test reportedly takes intact adults about 25 minutes to complete. With a vocabulary level that is quite basic, the breakdown into levels of reading skills may offer useful insights when reading impairment reflects neuropsychological dysfunction. Normative data are keyed to a variety of mostly blue-collar occupations, such as electrician or heavy equipment operator.

Reading Index-12 tests reading ability up to the 12th grade. Like the Reading Index, it is in a multiple-choice format. Its 72 items ask for comprehension of written materials ranging in length from phrases to paragraphs. Normative data are provided for workers in office/clerical and manufacturing positions.

Understanding Communication (T.G. Thurstone, 1992)

This reading comprehension test comprises 40 statements consisting of one to three sentences with the final wording incomplete. Four one-word or short phrase choices are offered to complete each statement, of which one makes good sense. As the test progresses, the statements become more difficult due to greater ideational complexity and more demanding vocabulary. Norms are provided for the 15-minute time limit, but examiners interested in how well patients slowed by brain dysfunction perform should allow them to complete as many items as they can. When performance on this test drops significantly below measured vocabulary level, the possibility of impaired reasoning and/or verbal comprehension may be considered.

National Adult Reading Test (NART) (H.E. Nelson and O'Connell, 1978); *National Adult Reading Test, 2nd ed. (NART-2)* (H.E. Nelson and Willison, 1991)

The NART list consists of 50 phonetically irregular words (see Table 13.9). Correct pronunciation of these words implies prior knowledge of them. This test is often used to estimate premorbid mental ability in adults because vocabulary correlates best with overall ability level and is relatively unaffected by most nonaphasic brain disorders (see pp. 92–94, 515). However, until recently, little direct evidence existed to support the assumption that current reading vocabulary is a good measure of prior intellectual ability. To assess whether NART scores correspond to premorbid mental ability, Crawford, Deary, and colleagues (2001) compared NART scores of a group of older adults (mean age 77 years) without dementia to their scores on an intelligence test taken at age 11 and found a high (.73) correlation. In contrast, NART scores had only a modest (.25) correlation with current MMSE scores in this group.

TABLE 13.9 The National Adult Reading Test

Ache	Subtle	Superfluous	Gouge	Beatify
Debt	Nausea	Radix	Placebo	Banal
Psalm	Equivocal	Assignate	Facade	Sidereal
Depot	Naive	Gist	Aver	Puerperal
Chord	Thyme	Hiatus	Leviathan	Topiary
Bouquet	Courteous	Simile	Chagrin	Demesne
Deny	Gaoled	Aeon	Detente	Labile
Capon	Procreate	Cellist	Gauche	Phlegm
Heir	Quadruped	Zealot	Drachm	Syncope
Aisle	Catacomb	Abstemious	Idyll	Prelate

Adapted from H.E. Nelson and O'Connell (1978)

Crawford (1992) and his colleagues conducted a series of studies in the United Kingdom on which they found that the NART IQ score correlates significantly with education ($r = .51$) and (not surprisingly) social class ($r = .36$); the $-.18$ correlation with age, while significant, accounted for practically none of the variance (Crawford, Stewart, Garthwaite, et al., 1988). There do not appear to be sex effects (Schlosser and Ivison, 1989). Scoring for errors, the Crawford group found a split-half reliability coefficient of .90 (Crawford, Stewart, Garthwaite, et al., 1988), interrater reliability coefficients between .96 and .98, and test–retest reliability coefficients of .98 (Crawford, Parker, Stewart, et al., 1989). In a factor analytic study combining the NART and the WAIS, they extracted a first factor, which they identified as "Verbal Intelligence", on which the NART error score had a high ($-.85$) loading (Crawford, Stewart, Cochrane, et al., 1989). In other studies comparing the NART and the WAIS IQ scores, they found that the NART predicted 72% of the VIQ variance but only 33% of the PIQ (Crawford, Parker, Stewart, et al., 1989). A correlation with demographic variables was .70 (Crawford, Allan, Cochrane, and Parker, 1990). These workers use the NART in conjunction with demographic variables for estimation of premorbid ability in deteriorating patients (Crawford, Cochrane, Besson, et al., 1990; Crawford, Nelson, et al., 1990; see also pp. 95–96).

When dementia patients have language disturbances, this procedure will underestimate premorbid ability (Stebbins, Gilley, et al., 1990; Stebbins, Wilson, et al., 1990). Alzheimer patients' reading problems were demonstrated by their decline in NART scores when examined annually over three years; the extent of decline was greatest for those with initially low Mini-Mental State Examination scores (Cockburn, Keene, et al., 2000). While NART scores do show a decrement with dementia severity, this decline is mild compared to measures of cognitive function showing marked declines (Maddrey et al., 1996). Although Spreen and Strauss (1998) recommend against using this kind of test with patients who are aphasic, dyslexic, or who have articulatory or visual acuity defects, Schlosser and Ivison (1989) pointed out that this test's sensitivity to the language deterioration in Alzheimer's disease may make it an effective early predictor of dementia.

NART variants. A short NART uses only the first half of the word list to avoid distressing patients with limited reading skills who can only puzzle through the more difficult half of the test (Crawford, Parker, Allan, et al., 1991). This format predicted WAIS IQ scores almost as well as the full word list (see p. 93).

North American Adult Reading Test (NAART, NART-R) (Blair and Spreen, 1989)[1]

This 61-word version of the NART has been modified for appropriateness for North American subjects, providing both U.S. and Canadian pronunciation guides as needed (see pp. 92–93). Twelve words from the NART generally unfamiliar to readers of North American English were replaced with 23 words more common to North Americans. Excellent interscorer reliability is reported and internal consistency is high. Like the NART, this instrument predicts WAIS-R VIQ well but not PIQ. In a large sample of healthy, well-educated adults ranging in age from 18 to 91 years, education was much more strongly related to performance than was age (Uttl, 2002). NAART scores increased with age up to 60 years and then leveled off. The correlation between NART scores and WAIS-R Vocabulary was .75. In this sample, 35 items were sufficient to predict WAIS-R Vocabulary reliably. This short version was recommended when time is limited.

American NART (AMNART) (Grober and Sliwinski, 1991)

A modification of the NART for American readers consists of 27 words from the British version and 23 new irregular American words of comparable frequency to the ones that were replaced. Grober and Sliwinski (1991) removed five words that had very low item-total correlations (see p. 93). Like the NART, this instrument predicts WAIS-R VIQ well but not PIQ.

Reading Subtest of the Wide Range Achievement Test–Revised (WRAT-R) (Jastak and Wilkinson, 1984), *Wide Range Achievement Test 3 (WRAT3)* (G.S. Wilkinson, 1993)

This test begins with letter reading and recognition at Level I (for children) and continues with a 75-word (WRAT-R) or 84-word (WRAT3) reading and pronunciation list. At Level II, Reading involves only the word list. The latest revision provides two forms (each a 42-item split-half) to facilitate retesting. The time limit for each response is 10 sec. The test is discontinued after ten failures. WRAT3 norms cover ages 5 to 75, but the highest WRAT-R age is "45 and over." For the WRAT3, normative data are available for the two split-half versions as well as the full 84-word list. African Americans matched for education with whites had scores about 5 points lower (Manly, Jacobs, Touradji, et al., 2002). For the WRAT3 normative sample, correlations with WAIS-R Vocabulary was .62. WRAT-R Reading and NART correlations are strong (.82) (Wiens, Bryan, and Crossen, 1993). No sex effects were found for a group of healthy participants ages 15 to 70 years (Klimczak et al., 2000).

The word pronunciation format of this test is identical to that of the NART, but it was developed to evaluate educational achievement rather than to assess premorbid ability. Both this test and the NART are based on the same assumptions: that familiar words will be pronounced correctly, and familiarity reflects vocabulary. It is further assumed in the WRAT that reading vocabulary provides a valid measure of reading ability. However, word recognition is not the same as reading comprehension; thus this test gives only a rough measure of academic achievement. Spreen and Strauss (1998) caution against using it for academic evaluations. It has not been used much in neuropsychological research protocols. One study did find a moderate association between right temporal lesions and poor performance, and a little weaker but significant association between right parietal lesions and poor performance on this test (Egelko, Gordon, et al., 1988).

A multiple-choice version, the Wide Range Achievement Test-Expanded Version (Robertson, 2002), adds a reading comprehension test in a multiple-choice format designed for children and adults up to age 24. Reading passages include selections from textbook, recreational, and other sources, designed to test word meaning in context as well as literal and inferential reading skills.

Reading Subtest of the *Kaufman Functional Academic Skills Test (K-FAST)* (A.S. Kaufman and Kaufman, 1994a)

This brief 34-item test assesses reading as it relates to everyday activities such as reading signs, understanding labels on medicines, and following directions in a recipe. The normative sample was a group of 1,434 people ages 15 to 85+. No sex effects were found for a 15 to 70 year-old group (Klimczak et al., 2000). Scores strongly correlated (.82) with WRAT3 Reading in this healthy sample. Whites performed slightly better than African Americans (T.H. Chen et al., 1994).

Writing

> Normal writing can be carried out only if a highly complex group of cortical zones remains intact. This complex comprises practically the whole brain and yet forms a highly differentiated system, each component of which performs a specific function . . . writing can

[1]The word list, pronunciation guide, and administration instructions are given in Spreen and Strauss (1998).

> be disordered by circumscribed lesions of widely different areas of the cerebral cortex, but in every case the disorder in writing will show qualitative peculiarities depending on which link is destroyed and which primary defects are responsible for the disorder of the whole functional system.
>
> *Luria, 1966, pp. 72–73*

Qualitative aspects of writing may distinguish the script of patients whose brain damage is lateralized (A. Brodal, 1973; Hécaen and Marcie, 1974). Patients with right hemisphere lesions tend to repeat elements of letters and words, particularly seen as extra loops on *m*, *n*, and *u*, and to leave a wider than normal margin on the left-hand side of the paper (A.W. Ellis, 1982; Roeltgen, 2003). Left visuospatial inattention may be elicited by copying tasks (see Fig. 10.8, p. 385). Difficulty in copying an address by patients with left visual inattention was significantly associated with right temporal lesions (Egelko, Gordon, et al., 1988). Generally, patients with left hemisphere lesions are more likely to have a wide right-sided margin, and they tend to leave separations between letters or syllables that disrupt the continuity of the writing line. Edith Kaplan has also noted that, frequently, aphasic patients will print when asked to write (personal communication, 1982 [mdl]). Different contributions of cortical regions to writing become apparent in the variety of writing disorders observed in patients with focal left hemisphere lesions (Coslett, Gonzalez, Rothi, et al., 1986; Roeltgen, 2003; Roeltgen and Heilman, 1985). Benson (1993) observed that "Almost every aphasic suffers some degree of agraphia." He therefore recommended that writing ability be examined by both writing to dictation and responsive writing (e.g., "What did you do this morning?").

Writing tests allow the examiner to evaluate other dysfunctions associated with brain damage, such as a breakdown in grammatical usage, apraxias involving hand and arm movements, and visuoperceptual and visuospatial abilities (Roeltgen, 2003). With brain disease, alterations in writing size (e.g., micrographia in Parkinson's disease) or writing output (diminished in dementia, increased in some conditions) may also occur. Figure 13.1 shows an attempt to write (*a*) "boat" and (*b*) "America" by a 72-year-old man with Alzheimer's disease of moderate severity and prominent apraxia. This difficulty in forming letters despite being able to spell the words orally is a form of apraxic agraphia.

In studying the writing disturbances of acutely confused patients, Chédru and Geschwind (1972) described a three-part writing test which shares some items with the Boston Diagnostic Aphasia Examination: (1) writing to command, in which patients were told to write a sentence about the weather and a sentence about their jobs; (2) writing to dictation of words (business, president, finishing, experience, physician, fight) and sentences ("The boy is stealing cookies." "If he is not careful the stool will fall."); and (3) Copying a printed sentence in script writing ("The quick brown fox jumped over the lazy dog."). They found that patients' writings were characterized by dysgraphia in the form of motor impairment (e.g., scribbling), spatial disorders (e.g., of alignment, overlapping, cramping), agrammatisms, and spelling and other linguistic errors. Moreover, dysgraphia tended to be the most prominent and consistent behavioral symptom displayed by them. The authors suggested that the fragility of writing stems from its dependence on so many different components of behavior and their integration. They also noted that for most people writing, unlike speaking, is far from being an overlearned or well-practiced skill. Signatures, however, are so overpracticed that they do not provide an adequate writing sample.

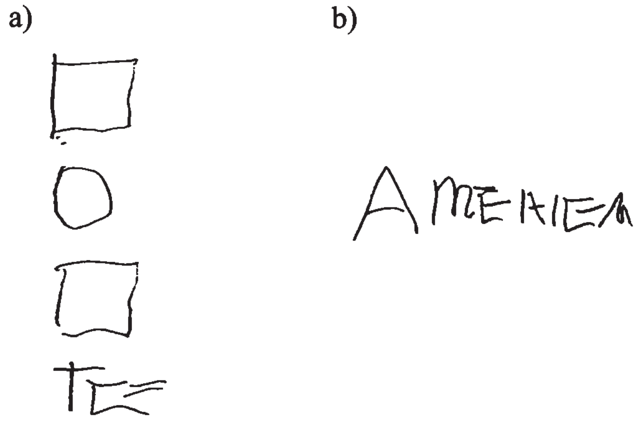

FIGURE 13.1 Alzheimer patient's attempt to write (*a*) "boat" and (*b*) "America."

When asked to write a description of "everything that is happening" in the Cookie Theft picture of the Boston Diagnostic Aphasia Examination, the responses of dementia patients were highly correlated (−.76) with ratings of dementia severity (J. Horner, Heyman, et al., 1988). Writing samples were scored according to (1) overall organization, relevance, and continuity of the writing; (2) vocabulary completeness and accuracy of word usage; (3) grammatical completeness and accuracy; (4) spelling accuracy; and (5) mechanics and legibility of writing, e.g., form, accuracy, and placement of letters and words. Evaluations were based on the sum of these scores.

Spelling

Poor spelling in adults can represent the residuals of slowed language development or childhood dyslexia, of poor schooling or lack of academic motivation, or of bad habits that were never corrected. Additionally, it may be symptomatic of adult-onset brain dysfunction. Thus, in evaluating spelling for neuropsychological purposes, the subject's background must be taken into account along with the nature of the errors. Both written and oral spelling should be examined because they can be differentially affected (McCarthy and Warrington, 1990).

Spelling Subtest of the Wide Range Achievement Test-Revised (WRAT-R) (Jastak and Wilkinson, 1984), *Wide Range Achievement Test 3 (WRAT3)* (Wilkinson, 1993)

This subtest calls for written responses. Young children begin with name and letter writing. The WRAT-R list consists of 46 words; the WRAT3 has 80. Normative data are available for two split-half versions of the WRAT3 containing 40 words each. The test is discontinued depending upon the subject's spelling skills. Following each word reading the examiner also reads a sentence containing the word. Fifteen seconds is allowed for each word. Ten failures is the criterion for discontinuing. No means for analyzing the nature of spelling errors is provided.

Johns Hopkins University Dysgraphia Battery (R.A. Goodman and Caramazza, 1985)

This test was developed to clarify the nature of spelling errors within the context of an information processing model (Margolin and Goodman-Schulman, 1992). It consists of three sections: I. *Primary Tasks* includes (A) Writing to dictation of material varied along such dimensions as grammatical class, word length, word frequency, and nonwords; and (B) Oral spelling. In II. *Associated Tasks,* the subject (C) writes the word depicted in a picture, (D) gives a written description of a picture, and (E, F) copies printed material either directly or as soon as it is withdrawn from sight. The subject's errors are evaluated in section III, *Error Coding,* according to one of 11 different kinds of error along with scoring categories for "Don't know" and "Miscellaneous errors." Margolin and Goodman-Schulman give examples of how these procedures can help to explicate different kinds of dysgraphic disorder.

Knowledge Acquisition and Retention

Information (Wechsler, 1955, 1981, 1997a)

Although many tests of academic achievement examine general knowledge, Information is the only one that has been incorporated into neuropsychological assessment batteries and research programs almost universally. The Information items test general knowledge normally available to persons growing up in the United States. WIS-A battery forms for other countries contain suitable substitutions for items asking for peculiarly American information. The items are arranged in order of difficulty from the four simplest, which all but severely retarded or organically impaired persons answer correctly, to the most difficult, which only few adults pass. Some Information items were dropped over the years because they became outdated. The relative difficulty of others can change with world events; e.g., the increased popular interest in Islamic culture will necessarily be reflected in a proportionately greater number of subjects in 2004 who know what the Koran is than in 1981 when this item was first used. In addition, increases in the level of education in the United States, particularly in the older age groups, probably contribute to higher mean scores on the WAIS-R version of Information and to lower mean scores on the more recently standardized WAIS-III Information (Lezak, 1988c; Quereshi and Ostrowski, 1985; see K.C.H. Parker, 1986, for a more general discussion of this phenomenon).

Administration suggestions. I [mdl] make some additions to Wechsler's instructions. I spell "Koran" after saying it since it is a word people are more likely to have read than heard, and if heard, it may have been pronounced differently. When patients who have not gone to college are given one or more of the last four items, I usually make some comment such as, "You have done so well that I have to ask you some questions that only a very few, usually college-educated, people can answer," thus protecting them as much as possible from unwarranted feelings of failure or stupidity if they are unfamiliar with the items' topics. When a patient gives more than one answer to a question and one of them is correct, the examiner must insist on the patient telling which answer is preferred, as it is not possible to score a response containing both right and wrong answers. I usually ask patients to "vote for one or another of the answers."

Although the standard instructions call for discontinuation of the test after five failures, the examiner may use discretion in following this rule, particularly with brain injured patients. On the one hand, some neurologically impaired patients with prior average or higher intellectual achievements are unable to recall once-learned information on demand and therefore fail several simple items in succession. When such patients give no indication of being able to do better on the increasingly difficult items and are also distressed by their failures, little is lost by discontinuing this task early. If there are any doubts about the patient's inability to answer the remaining questions, the next one or two questions can be given later in the session after the patient has had some success on other tests. On the other hand, bright but poorly educated subjects will often be ignorant of general knowledge but have acquired expertise in their own field, which will not become evident if the test is discontinued according to rule. Some mechanics, for example, or nursing personnel, may be ignorant about literature, geography, and religion but know the boiling point of water. When testing alert persons with specialized work experience and limited education who fail five sequential items not bearing on their personal experience, I usually give all higher-level items that might be work-related.

I have found it a waste of time to give the first few items where the usual administration begins (items 5 to 7, 8, or 9) to well-spoken, alert, and oriented persons with even as little as a tenth grade education. Thus, I begin at different difficulty levels for different subjects. Should a subject fail an item or be unable to retrieve it without the cueing that a multiple-choice format provides (see below), I drop back two items, and if one of them is failed I drop back even further; but having to drop back more than once occurs only rarely.

When giving the Information test to a patient with known or suspected brain dysfunction, it is very important to differentiate between failures due to ignorance, loss of once-stored information, and inability to retrieve old learning or say it on command. Patients who cannot answer questions at levels higher than warranted by their educational background, social and work experiences, and vocabulary and current interests have probably never known the answer. Pressing them to respond may at best waste time, at worst make them feel stupid or antagonize them. However, when patients with a high school education cannot name the capital of Italy or recognize "Hamlet," I generally ask them if they once knew the answer. Many patients who have lost information that had been in long-term storage or have lost the ability to retrieve it, usually can be fairly certain about what they once knew but have forgotten or can no longer recall readily. When this is the case, the kind of information they report having lost is usually in line with their social history. The examiner will find this useful both in evaluating the extent and nature of their impairments and in appreciating their emotional reactions to their condition.

When patients acknowledge that they could have answered the item at one time, appear to have a retrieval problem or difficulty verbalizing the answer, or have a social history that would make it likely they once knew the answer, information storage can be tested by giving several possible answers to see whether they can recognize the correct one. I always write out the multiple-choice answers so the patient can see all of them simultaneously and need not rely on a possibly failing auditory memory. For example, when patients who have completed high school are unable to recall Hamlet's author, I write out, "Longfellow, Tennyson, Shakespeare, Wordsworth." Often patients identify Shakespeare correctly, thus providing information both about their fund of knowledge (which they have just demonstrated is bigger than the Information score will indicate) and a retrieval problem. Nonaphasic patients who can read but still cannot identify the correct answer on a multiple-choice presentation probably do not know, cannot retrieve, or have truly forgotten the answer. (The WAIS-R NI provides a prepared set of multiple-choice answers.)

The additional information that the informal multiple-choice technique may communicate about the patient's fund of knowledge raises scoring problems. Since the test norms were not standardized on this kind of administration, additional score points for correct answers to the multiple-choice presentation cannot be evaluated within the same standardization framework as scores obtained according to the standardization rules. Nevertheless, this valuable information should not be lost or misplaced. To solve this problem, I use *double scoring;* that is, I post both the age-graded standard score the patient achieves according to the standardization rules and, usually following it in parentheses, another age-graded standard score based on the "official" raw score plus raw score points for the items on which the patient demonstrated knowledge but could not give a spontaneous answer. This method allows the examiner to make an estimate of the patient's fund of background information based on a more representative sample of behavior, given the patient's impairments. The disparity between the two scores can be used in making an estimate of the amount of deficit the patient has sustained, while the lower score alone indicates the patient's present level of functioning when verbal information is retrieved without assistance.

On this and other WIS-A tests, an administration adapted to the patient's deficits with double-scoring to document performance under both standard and adapted conditions enables the examiner to discover the full extent of the neurologically impaired patient's capacity to perform the task under consideration. Effective use of this method involves both testing the limits of the patient's capacity and, of equal importance, standardized testing to ascertain a baseline against which performance under adapted conditions can be compared.

In every instance, the examiner should test the limits only after giving the test item in the standard manner with sufficient encouragement and a long enough wait to satisfy any doubts about whether the patient can perform correctly under the standard instructions.

Test characteristics.[1] Information scores hold up well with aging. When education effects are controlled (by covariance), Information scores stay steady into the 70s (A.S. Kaufman, Kaufman-Packer, et al., 1991; A.S. Kaufman, Reynolds, and McLean, 1989), and for an educationally relatively privileged group, they decline only slightly into the 90s (Ivnik, Malec, Smith, et al., 1992b). Significant sex differences of around 1 scaled score point on all forms of the WIS favor males (A.S. Kaufman, Kaufman-Packer, et al., 1991; A.S. Kaufman, McLean, and Reynolds, 1988; Snow and Weinstock, 1990). Of course, education weighs heavily in performances on this test, accounting for as much as 37 to 38% of the variance in the over-35 age ranges. After controlling for the effects of age, education, and sex, African Americans with traditional African American practices, beliefs, and experiences had significantly lower WAIS-R Information scores than African Americans who were more acculturated (Manly, Miller, et al., 1998). These authors propose that due to their educational and cultural experiences, some African Americans are not routinely exposed to item content on Information. In another study, African Americans obtained mean scores that were $1^1/_2$ to 2 scaled score points below those of whites, but education differences between these two groups were not reported (A.S. Kaufman, McLean, and Reynolds, 1988). Urban subjects over age 55 performed significantly better than their rural age peers, but this difference did not hold for younger people: "Perhaps the key variable is the impact of mass media, television . . . on the accessibility of knowledge to people who are growing up in rural areas" (A.S. Kaufman, McLean, and Reynolds, 1988, p. 238).

Test–retest reliability coefficients mostly in the .76 to .84 range have been reported, varying a little with age and neuropsychological status (Rawlings and Crewe, 1992; J.J. Ryan, Paolo, and Brungardt, 1992; Snow, Tierney, Zorzitto, et al., 1989; see also McCaffrey, Duff, and Westervelt, 2000a), with only a schizophrenic group providing an exceptional correlation coefficient of .38 (G. Goldstein and Watson, 1989). The highest reliabilities (.86–.94) are reported for samples of the normative populations (Wechsler, 1955, 1981, 1997a). Split-half reliability coefficients are high (.85 to .96) in clinical groups although somewhat lower (.74) in mentally retarded adults (Zhu, Tulsky, et al., 2001). TBI patients who took this test four times within a year did not gain a significantly greater number of score points than did patients who only took the first and last of the test series (Rawlings and Crewe, 1992). Older subjects retested within a half year made a significant but small gain (about 1/2 of a scaled score point) on this test (J.J. Ryan, Paolo, and Brungardt, 1992). In factor analytic studies, Information invariably loads on a Verbal Comprehension factor (see p. 650). Information's high correlations with other mental ability tests led Feingold (1982) to conclude that it can be used alone as a measure of general ability. As could be expected, correlations with measures of executive functioning are minimal (Isingrini and Vazou, 1997).

Information and Vocabulary are the best WIS-A measures of general ability, that ubiquitous test factor that appears to be the statistical counterpart of learning capacity plus mental alertness, speed, and efficiency. Information also tests verbal skills, breadth of knowledge, and—particularly in older populations—remote memory. Information tends to reflect formal education and motivation for academic achievement. It is one of the few tests in the WIS-A batteries that can give spuriously high ability estimates for overachievers or fall below the subject's general ability level because of early lack of academic opportunity or interest.

Neuropsychological findings. Glucose metabolism increases in the left temporal lobe and surrounding areas during this test, with much smaller increases also noted in the right temporal lobe (Chase et al., 1984). In brain injured populations, Information tends to appear among the least affected of the WIS-A tests (Donders, Tulsky, and Zhu, 2001; O'Brien and Lezak 1981; E.W. Russell, 1987). Although a slight depression of the Information score can be expected with brain injury of any kind, because performance on this test shows such resiliency, particularly with focal lesions or trauma, it often can serve as the best estimate of the original ability. In individual cases, a markedly low Information score suggests left hemisphere involvement, particularly if verbal tests generally tend to be relatively depressed and the patient's history provides no other kind of explanation for the low score. Thus, the Information performance can be a fairly good predictor of the hemispheric side of a suspected focal brain lesion (Hom and Reitan, 1984; A. Smith, 1966; Spreen and Benton, 1965). However, contrary to folklore that Information holds up well with dementia, it is actually one of the more sensitive of the WIS verbal tests and

[1]Most of the following data come from WAIS-R studies.

appears to be a good measure of dementia severity (Larrabee, Largen, and Levin, 1985; Storandt, Botwinick, and Danziger, 1986).

Information (WAIS-R NI) (E. Kaplan, Fein, et al., 1991). In the initial administration of Information, WAIS-R NI instructions recommend that all items be given unless the subject becomes too discouraged or frustrated. The multiple-choice test is given after the standardized tests. An analysis of item content (into "number facts," "directions and geography," academically related information, and responses requiring names) relates error patterns to possible interpretations of them. Using the multiple-choice technique, subjects in the 50- to 89-year age range averaged one-and-one-half raw score points more than on the standard administration (Edith Kaplan, personal communication, February, 1993 [mdl]). The WAIS-R NI benefit increased with age and was greatest (gain of 2.41 raw score points) for subjects 80–89 years old.