Another difference relates to recovery. In the case shown in Figure 3.13, the language-related cortices are not damaged to the same extent as in the case shown in Figure 3.12, and recovery is far superior. These patients may not recover to normal speech and language, but neither do they remain severe global aphasics.

One other anatomical pattern in global aphasia is that of a patient with a lesion in the left frontal operculum, underlying white matter, basal ganglia, insula, and even part of the parietal operculum, but it spares the tem-
FIGURE 3.13. Magnetic resonance template of a patient with global aphasia but without hemiparesis (JMCC0656). In this patient there were two lesions in the left hemisphere, one in the superior sector of the frontal operculum and premotor cortex immediately above, and another in the angular gyrus.

FIGURE 3.14. Magnetic resonance template of a patient with global aphasia and hemiparesis (BD0638) in the acute stage, who later became a severe Broca's aphasic. Note that the lesion involved the left frontal operculum (areas 44 and 45), the premotor and motor cortices immediately behind and above Broca's area, as well as the insula and basal ganglia, but spared completely the temporal and parietal lesions.
poral lobe. Such patients also tend to recover and, in the chronic stable state, come to resemble a Broca' aphasic (Figure 3.14).

Alexia without Agraphia (Pure Alexia)

Alexia without agraphia is not an aphasic disorder as such, inasmuch as speech output and aural comprehension are intact. The condition is associated with a remarkably consistent anatomical localization (A. Damasio & Damasio, 1983). Figure 3.15 shows a case with alexia without agraphia. The lesion extends from the occipital cortex deep into the white matter, reaching the left lateral ventricle at the level of the trigone and occipital horn (the paraventricular area). It involves the primary visual cortex (area 17) and part of the visual association cortices (areas 18 and 19), and extends into the mesial occipitotemporal junction involving mesial area 37 and the posterior sector of the parahippocampal gyrus. The corpus callosum is intact but the lesion disrupts interhemispheric connectional systems that course through the splenium of the corpus callosum, in the forceps major,

FIGURE 3.15. Computerized tomographic template of a patient with alexia without agraphia (PA0321). The lesion involved the mesial sector of the left occipital lobe, the mesial occipitotemporal junction, and the white matter in the paraventricular region (where the forceps major courses), but spared the corpus callosum proper.
and interlock the visual cortices. Thus, even if the corpus callosum proper
does not show any damage, its outflow is compromised.

Atypical Aphasias

The advent of modern neuroimaging techniques led to the identification
of the left basal ganglia as the lesion correlate for a group of aphasias
known as "atypical," for lack of a better term. These aphasias are generally
of the fluent type, in some way resembling Wernicke's aphasia. Yet, unlike
typical fluent aphasias, there are also disturbances of articulation and,
even more deviantly, a right hemiparesis is present (A. R. Damasio, Dama-
sio, Rizzo, Varney, & Gersh, 1982; H. Damasio & Damasio, 1989; H. Dama-
sio, Eslinger, & Adams, 1984; Naeser et al., 1982).

The lesions are located deep in the left hemisphere and invariably in-
clude portions of the caudate nucleus and putamen and the anterior limb
of the internal capsule. They often occur in younger patients and are
caused by embolic events, where an embolus becomes lodged in the prox-
imal segment of the middle cerebral artery at the level of the lenticulostrici-
ate arteries, which supply the head of the caudate, the anterior limb of the
internal capsule, and the lenticular nucleus. Another atypical aphasic syn-
drome, with a strong resemblance to transcortical sensory aphasia, can oc-
cur with infarcts in the left thalamus when the anterior nuclei are involved
(Graff-Radford, Damasio, Yamada, Eslinger, & Damasio, 1985). In neither
of these cases is there cortical damage in acute or chronic stages.

Conclusion

A large variety of acquired aphasic syndromes and of closely associat-
ed disturbances (mutism and pure alexia) can be correlated to relatively
specific brain lesions located at varied sites in the left cerebral hemisphere.
More than 100 years of study of anatomoclinical correlations, with autops-
sy material as well as CT and MR scans, has proven that in spite of the in-
evitable individual variability, the correlation between aphasia types and
locus of cerebral damage is surprisingly consistent. Not surprisingly, there
are exceptions, which can be found, in particular, in left-handed subjects,
whose cerebral dominance for language is nonstandard and variable from
subject to subject, and even in a minority of right-handed subjects who
have right cerebral dominance for language (lesions within right hemi-
spheres in these subjects produce "crossed" aphasias).

The value of these consistent correlations for clinical management is un-
questionable. It should be clear, however, that the correlations per se pro-
FIGURE 3.16. Left hemisphere, lateral and mesial aspects, with identification of major fissures and gyri. Note that the insula cannot be seen on a lateral view because it is buried in the depth of the sylvian fissure, in its anterior portion, and is covered by the frontal operculum (the more posterior and inferior portion of the inferior frontal gyrus and the inferior portion of the precentral gyrus) and that Heschl gyri also cannot be seen because they occupy the superior surface of the superior temporal gyrus, buried inside the sylvian fissure. H.G., Heschl gyri (seen only in the superior surface of the superior temporal gyrus); L.G., lingual gyrus; F.G., fusiform gyrus; R.S., retrosplenial area; S.M.A., supplementary motor area.
FIGURE 3.17. Left hemisphere, lateral and mesial view with Brodmann’s cytoarchitectonic nomenclatures. Note that areas 41 and 42, corresponding to Heschl gyri, are not seen in this view because they occupy the superior surface of the superior temporal gyrus, inside the sylvian fissure.

Figures 3.16 through 3.21 present different views of the brain with its areas, fissures, and gyri delineated.

vide only limited information about the neurobiological mechanisms of language, in health and in disease.
FIGURE 3.18. Inferior and superior views of both hemispheres. On the left-hand side of each view, the major gyri and fissures are marked, and on the right-hand side Brodmann’s cytoarchitectonic fields are shown.
FIGURE 3.19. Left hemisphere, lateral and mesial aspects, with identification of the major functional areas. Note that the auditory cortex occupies both the lateral aspect of the superior temporal gyrus and the superior aspect, inside the sylvian fissure (not seen in this lateral view) where the transverse temporal gyri (the primary auditory areas) and the planum temporale are located.
FIGURE 3.20. Left hemisphere with the ventricular system. $fh + blv + tr + oh + th =$ left lateral ventricle; $fh$, frontal horn; $blv$, body; $tr$, trigone; $oh$, occipital horn; $th$, temporal horn; $3v$, third ventricle which connects with both lateral ventricles through the foramina of Monro (the left one is marked with an arrow) and continues caudally into the aqueduct (double arrow).

FIGURE 3.21. Left hemisphere with the basal ganglia seen in lateral view. The insert shows the basal ganglia seen from the occipital pole.
Acknowledgment

This work was supported by NINCDS Grant PO1 NS19632.

References


Assessment of Aphasia

OFTFRIED SPREEN and ANTHONY H. RISSE

Even the earliest records of medical knowledge refer to language disorders after brain damage (Benton, 1964). Accounts of simple clinical examinations were often included in such reports, but it was not until the second half of the nineteenth century, specifically since the publications of Broca (Joynt, 1964), that aphasia was explored more systematically. Case reports by Wernicke (1874/1908) and contemporaries contained detailed descriptions of examination procedures for individual patients. Whereas some of these examinations were probably standard procedure in certain hospitals, others were invented on the spot to explore individual features of a specific syndrome of aphasia. Understandably, these reports focused on the patient's specific disorder rather than on the examination procedure.

The clinical examination as developed in the late nineteenth century has been modified and augmented, but it has remained the essential tool of the clinical neurologist. Such examinations are exemplified in the writings of Jackson (1915) and Pick (1913). The standard repertoire of many clinical examinations includes such routine procedures as the Paper Test of Pierre Marie (1883), the Hand–Eye–Ear Test of Henry Head (1926), and Geschwind's (1971) "no ifs, ands, or buts" repetition as a simple task with high multiple demands on the patient's understanding, processing, and repetition ability.

The clinical examination has a number of disadvantages, which gradually led to the development of more generally applicable and standardized assessment instruments. Clinical examinations tend to vary from one place to another, both in content and in the way in which they are administered. What is considered abnormal remains up to the subjective judgment of the
A clinician. Early attempts to produce a more standardized examination were published by Head (1926), who insisted on a detailed "clinical protocol." Another examination procedure was published by Froeschels, Ditttrich, and Wilhelm (1932).

The first comprehensive battery of tests for aphasic patients was used by Weisenburg and McBride (1935) in a 5-year study of 60 aphasic patients. Schuell (Schuell, Jenkins, & jiminez-Pabon, 1964) called this study a landmark because it was the first to use control subjects, to compare aphasic with nonaphasic brain-damaged subjects, and to use standardized methodology. Several other batteries were developed in the 1950s by Wepman (1951), Eisenson (1954, 1993), and Schuell (1955), partly as a result of intensive treatment efforts with World War II veterans. Benton (1964) reviewed the development of assessment procedures and noted the work done in various centers, criticizing that none of these procedures had been published in "usable form." The descriptions of procedures were insufficient, no standardization information was presented, and neither exact criteria for scoring nor detailed guides for interpretation were included. He compared the state of the art of aphasia testing with the "pre-Binet stage" in intelligence testing: "We are today where intelligence testing was in 1900" (p. 263).

The situation has changed in the 35 years since Benton's review. Instruments have been published that present detailed administration and scoring criteria and, at least in part, provide information on standardization and interpretation procedures. Benton (1994), although accepting limitations of the status quo, concluded that standardized tests have proven to be effective in monitoring the status of aphasia and in guiding rehabilitation strategies. Our review deals primarily with contemporary assessment techniques. Other reviews have been presented by Davis (1993), Kertesz (1979), Lezak (1995), and Spreen and Strauss (1998).

Purposes of Assessment and Testing

Assessment procedures vary greatly, depending on the examiner's goal. It is important to consider the purpose in evaluating and choosing specific instruments. Matching the assessment procedure to the patient requires a flexible and knowledgeable approach to assessment and testing. Four general types of evaluation purposes may be distinguished: (a) screening, (b) diagnostic assessment, (c) descriptive testing for rehabilitation and counseling, and (d) progress evaluation. A balanced approach using any of the four types when appropriate is empirically and clinically sound, and requires thoughtful decision-making by the practitioner.
Screening

Screening refers to a brief and cursory examination to detect the presence of a disorder. Three types of aphasia screening can be identified: (a) the bedside clinical examination; (b) screening tests per se; and (c) tests of specific aspects of language functioning that are sensitive to the presence of aphasia.

The bedside clinical examination is a clinical evaluation in the tradition of classical neurology (see Benson, 1979b; Strub & Black, 1985) and historically has been the primary method of assessing aphasia. It permits a brief and practical evaluation of language disorders, and remains a standard tool for many attending physicians, neurologists, and speech clinicians. The skilled clinician makes maximal use of his or her communicative interactions with the patient to rule out aphasia, establish a diagnosis, or reach the decision whether or not a more comprehensive assessment is warranted.

Screening tests reflect the once-popular application arising out of the clinical psychology of the 1950s of screening for the presence or absence of "organicity" (a loose term referring to any form of damage to the nervous system affecting psychological functions), particularly in high-risk groups and in conjunction with psychiatric evaluations. In relation to aphasia, some relatively brief and highly sensitive screening tests are available, but testing for the purpose of screening has lost its attractiveness and usefulness since the 1950s. One reason is that the accuracy of screening devices is limited, usually around 80% (Spreen & Benton, 1965). Another reason is that in clinical practice such cursory "detective work" is rarely necessary, as most patients are referred with an established clinical impression of aphasia, a known organic etiology, and neuroradiological localization. Finally, the information obtained from such instruments offers poor specificity (i.e., there are many reasons why a nonaphasic patient may fail a language screening test) and reveals little to indicate the severity of the detected aphasia. Thus, screening tests have been all but abandoned.

Specific-function tests that explore a highly sensitive aspect of language functioning are easily incorporated into comprehensive evaluations, such as the neuropsychological assessment, to "screen" (explore) the nature of language functioning and to determine the need for additional and/or ancillary testing to better describe an observed deficit. The Token Test (TT) has proven itself to be the most durable and broadly used of such specific-function tests. The Boston Naming Test also has developed appeal as a stand-alone "screen" of language. Subtests of established aphasia batteries are also used, such as the Neurosensory Center Comprehensive Examination for Aphasia (NCCEA) Word Fluency ("F-A-S") task.
**Diagnostic Assessment**

Diagnostic assessment refers to an overall examination of a patient's language performance to arrive at both a diagnostic impression and a detailed description of areas of associated strengths and weaknesses. Because of the comprehensive nature of this examination, it is suitable for patients who are medically stable in the acute or postacute period of their recovery and for initial and/or follow-up evaluations of patients with subjective complaints of language problems. Diagnostic assessments tend to elicit brief samplings of performance in many different areas and may not necessarily be of use to the speech clinician interested in a detailed exploration of a particular problem. When the evaluation is confined to performances on language and aphasia-related tasks, the diagnostic impression may either refer to the type of aphasia present (i.e., classification) or go beyond the description of the functional deficit and arrive at speculative conclusions about the nature and location of the underlying brain disorder. Impressions from a broader cognitive (neuropsychological) evaluation include the type of aphasia as only one of a number of differentially diagnosable neurobehavioral syndromes, such as dementia, confusional states, amnestic syndrome, attentional disorder, and so forth, in order to determine the full spectrum of the patient's deficits (i.e., differential diagnosis).

**Descriptive Assessment**

For the purpose of rehabilitation and counseling, a descriptive assessment would seem to be the most sensible approach to take, choosing, when warranted, a variety of assessment procedures—a diagnostic battery, a functional communication scale, and a number of ancillary tests. Rehabilitation and counseling pose different questions than can be answered from a strictly diagnostic assessment. In particular, it is important to gain as much information as possible about areas of functional strength, as well as about the presence of deficits, as this allows better-reasoned advice on what treatment activities to pursue, what vocational options remain open to the patient, and the actual communicative level the patient can attain. The assessment shifts from a strict testing situation to the observation of communicative behavior. A bridging of test contexts is necessary to compare the relation between specific language behaviors and deficits with the general ability to communicate. Descriptive assessment in the rehabilitation setting also involves (a) making predictions about recovery and response to treatment, and (b) measuring the ability of the patient to process, learn, and remember new material and, hence, to participate actively in and benefit from individualized treatment programs.
Progress Evaluation

Closely related to the descriptive assessment, progress evaluation permits an examination of spontaneous recovery when initial tests are repeated or when daily progress is charted in treatment settings. The clinician caring for the patient would like to be able to chart changes accurately over time rather than rely on subjective judgments or enthusiastic endorsements of the usefulness of therapy made by the patient or relatives. No formal tests have been developed specifically for this purpose, mainly because such progress assessments have to be tailor-made for each individual and her or his current level and range of deficit. Additionally, criteria need to be established for what will qualify as significant progress, as this likely would vary on a case-by-case basis depending on the patient’s premorbid characteristics. Ad hoc assessments may be formulated for entire domains of language functioning or for the modification of specific language behaviors.

The ability of the patient to relearn or compensate for what he or she has lost should be part of the progress evaluation. This is a neglected aspect of aphasia assessment. Most tests merely measure the status quo but deliberately exclude any practice or learning during the testing procedure. Providing cues to the patient during diagnostic evaluations usually is seen as contributing to measurement error, and hence should be avoided. If a test were designed to provide information on the relearning capacity of the patient, the approach would systematically include a variety of short learning trials with different kinds of cues to determine whether the patient’s language performance improves, at least within the immediate testing situation. It should be obvious that the inclusion of such procedures in the assessment of aphasia would dramatically change the usual form of testing, affect test reliability, and presumably add to the length of the test. Yet it is our impression that the benefit of such tests will outweigh the additional problems of test construction and validation, and that such tests will be a major focus of test development in the future.

Psycholinguistic Evaluation of Aphasic Language

Some studies have attempted to record and analyze conversational speech of aphasic patients in a setting that makes no specific demands on the patient. The main goal of such studies is not, however, an assessment of communicative abilities but a more detailed study from a psycholinguistic point of view. Studies of use and abuse of syntax, grammar, word
choice, frequency of word usage, pauses, and hesitations, speed of utter-
ance, and so forth, can be conducted with such "free-speech" samples. The
alternative approach is to focus on each aspect of psycholinguistic analy-
sis individually and construct an experimental setting that allows an
analysis of the types of errors produced by an aphasic patient.

Both the open-ended free speech and the experimental approach have
been used extensively in aphasia research (Goodglass & Blumstein, 1973;
Spreen, 1968). Insofar as these studies represent experiments rather than
tries to assess the aphasic patient's deficit, such studies are not re-
viewed here. However, some of these studies have led to conclusions about
the nature of the deficit in specific types of aphasia and can be translated
into suitable methods of assessment.

The first comprehensive psycholinguistic studies were conducted by
Wepman and collaborators (Wepman & Jones, 1964). They calculated var-
ious linguistic parameters from stories told in response to looking at pic-
tures, including grammatical form class usage, grammatical correctness,
intelligibility, and word-finding problems. The studies involved complex
calculations as well as judgments by linguistically trained researchers; they
cannot be readily translated into more directly accessible forms of assess-
ment.

The second major project was conducted by Howes and collaborators
(Howes, 1964, 1966, 1967; Howes & Geschwind, 1964) and involved a de-
tailed analysis of conversational speech of 5000 words each of more than
80 aphasic and nonaphasic speakers. The analyses concentrated on lexical
diversity (i.e., the frequency of word usage in aphasic vs. nonaphasic
speakers) and also distinguished between "fluent" and "nonfluent" speak-
ers, who were viewed as similar to Wernicke's and anomic aphasics and to
Broca's aphasics, respectively. Benson (1967) developed a simplified and
clinically useful rating scale system based on the information from Howe's
study with some additional rating dimensions.

A third comprehensive psycholinguistic analysis of conversational
speech by Spreen and Wachal (1973) used a computer-scored interaction
analysis of aspects of spoken language. Crockett (1972, 1976) designed 5-
point rating scales for 17 characteristics of speech—including rate of
speech, prosody, pronunciation, hesitation, and so on—to translate psy-
cholinguistic speech characteristics into basic rating scale dimensions. In-
terrater agreement among five judges was satisfactory after some training,
and after a carefully worded description of each characteristic was given.

Although Benson's (1967) and Crockett's (1972, 1976) ratings of psy-
cholinguistic aspects of aphasic speech appeared to be quite successful
within the limited scope of the research problems under investigation, fur-
ther use of this approach has been limited. Several of the ratings have been
incorporated into the Boston Diagnostic Aphasia Examination (BDAE). Shewan (1988) developed a system to describe and quantify aphasic subjects’ connected language in describing pictures, using phonologic, semantic, and syntactic components, as well as in general parameters of output (e.g., number of utterances, paraphasias). The Shewan Spontaneous Language Analysis (SSLA) also received some psychometric support, indicating adequate intrajudge, interjudge, and test–retest reliability, and validity based on clinical judgment of severity of connected language impairment. More recently, L. E. Nicholas and Brookshire (1995) have developed what they termed a Correct Information Unit (CIU) analysis of connected language. The CIU is a standardized, rule-based scoring system for which psychometric properties have been documented. Another detailed analysis of free conversational speech samples (Crockford & Lesser, 1994) stressed the analysis of editing elements (Schlenk, Huber, & Willmes, 1987) produced by the patient, amount and type of collaborative “repair” (Barnsley, 1987), and the proportion of conversational “load” (P. Hawkins, 1989; Miller, 1989) carried by the patient and the conversational partner, as well as unfilled and filled pauses, unsolicited repetitions, phonemic approximations, circumlocutions, and neologisms. In addition, Gerber and Gurland (1989) developed an Assessment Protocol of Pragmatic Linguistic Skills (APPLS), and P. M. Lesser and Coltheart (1992) published the Psycholinguistic Assessment of Language Performance in Aphasia (PALPA).

These studies demonstrate clearly that the somewhat elusive aspects of speaking style can be translated into scales that are readily understood and usable. Perhaps one reason for the infrequent use of such ratings has been that most have been developed in the context of relatively complex research rather than as part of the development of assessment tools. Another reason may be that psycholinguistic aspects of aphasic speech are rather complex in themselves and are not readily understood without prior linguistic training; hence, the clinically oriented examiner tends to shy away from psycholinguistic evaluations and to use relatively more concrete standard testing methods.

Construction Principles of Aphasia Tests

Both clinical examinations and tests may investigate the same areas of difficulty; the distinction lies in the quantification of the test examination and in the opportunity to compare quantitative scores with reference norms. Hence, a test could be defined as a clinical examination that meets a number of psychometric requirements.

The following section describes the psychometric requirements for a
well-constructed test to establish the information that should be critically evaluated before a test is put to use in daily practice. Few tests in the area of aphasia assessment fully meet the stringent psychometric requirements often demanded by the psychometric specialist and by groups concerned with the standards of testing (American Psychological Association, 1985). The reason for this is that most tests in the field of aphasia have been developed in individual laboratories, in the context of clinical work, and are not generally adopted by a large number of services and institutions. The collection of norms and the validity and reliability studies proceed slowly and are almost entirely dependent on the resources of test authors and their collaborators. The demand for such tests remains small compared, for example, with tests of general intelligence. In other words, test development is demanding in terms of both time and money. Aphasia tests are not best-sellers; as a result, development has been less than optimal.

**General Requirements for Tests**

The most frequently stated requirements for tests of any kind are demonstrated reliability, validity, and standardization (Anastasi, 1988; Cronbach, 1990).

**STANDARDIZATION**

STANDARDIZATION refers first to the test administration itself, which should be constant from patient to patient and from one examiner to another. If test administration and the conditions under which the test is administered are kept as controlled as possible, measurement error can be kept to a minimum. Any deviation from a standard administration procedure (e.g., prompting if the patient cannot respond readily, extending the time limits for answering) will inevitably produce more variability in test scores, and hence undesirable variance when the scores of patients or groups of patients are compared. The clinician may be tempted to use the test material to explore how much a patient may improve as a result of simple aids given during the testing. However justified, it should be understood that test results achieved under such modified conditions are no longer comparable to the published norms; that is, they contain an undesirable degree of measurement error. Exploration of the impact of aids and cues in a nonstandard, ad hoc basis has often been referred to as “testing the limits,” and is frequently used for descriptive rather than diagnostic purposes.

Standardization also refers to the establishment of norms against which the performance of an individual patient can be compared. Norms are essentially a set of scores obtained from a reference group. For example, a
score at the 90th percentile indicates a performance better than that of 90% of the reference population. If the test is constructed for a variety of populations, separate norms have to be established (e.g., if scores tend to vary greatly with age, sex, or educational level). Often, additional norms can be avoided by using correction scores for these factors, but this may be impractical if more than two of these factors interact with each other. Norms are usually produced for a group of healthy men and women without neurological impairment or aphasia, which allows the examiner to see where a given person’s performance lies in the distribution of scores. Clinically, the examiner can determine whether a patient’s performance is within normal limits, is a borderline performance, or is defective (e.g., a performance lower than the 1st or 5th percentile).

RELIABILITY

RELIABILITY is the demonstration that on repeat administration after a reasonable time interval and under the same conditions similar results will be obtained for the same subject. Reliability is often demonstrated by giving an alternate form of the test during the same or at a subsequent session, by comparing alternate items of the test, or by subdividing the test. Generally, reliability is best demonstrated with normal, healthy subjects, as the measurement error in patient populations and the likelihood of change in performance due to changes in the patient’s condition are both high. In practice, interscorer differences can be reduced to a minimum if the test manual contains sufficient scoring instructions and samples of how a given response can be scored. One form of expressing scoring reliability is to give the test records to two or more independent scorers and compute a correlation between scorers. Such interscorer reliability is highly desirable, as poor reliability of this type will obviously not only affect the general reliability of the test, but also introduce measurement error into studies of validity and other psychometric properties.

VALIDITY

Validity is probably the most crucial requirement of any test. It refers to the demonstration that a test measures what it claims. Validity can be measured in a variety of ways; usually, a distinction is made between PREDICTIVE (or criterion-related), CONTENT, and CONSTRUCT VALIDITY. Of the three forms, the demonstration that a test is a valid “predictor” of whether the patient is aphasic is the most popular, but of limited value in several ways. The demonstration of validity relies entirely on the fact that the aphasic patient’s performance can be discriminated from that of normal subjects’ on the basis of test results; in other words, the demonstration of validity comes close to the screening problem described earlier. Such a demonstration re-
lies on the clinical judgment made for the aphasic group but neglects the fact that the discrimination between aphasics and normals could result from entirely irrelevant (for aphasia) or trivial test items. In the ideal case, other contrast groups in addition to healthy, normal subjects (i.e., brain-damaged patients without aphasia) should be used.

Construct validity is often demonstrated by investigating the correlation of a new test with another test of known validity. Since few tests in the field of aphasia have known validity, another form of validity examination—the demonstration of factorial construct validity—is frequently used. In this case, factor analytic statistical techniques are used to show whether the tests in a given battery all contribute to a major factor of common variance that represents language functions.

Content validity refers to the adequacy of sampling from the domain of behaviors to be measured. In the case of testing for aphasia, for example, measuring verbal fluency alone would not be sufficient, because it does not appear to sample adequately the whole range of language behavior. In other words, test items should be based on sound reasoning, and should not be trivial or selectively biased. The content should also agree with the content areas as defined by other researchers. The range and diversity of the content of a test can also be explored by factor analysis.

**Specific Requirements for Tests with Brain-Damaged and Aphasic Patients**

In addition to the general requirements, several specific problems frequently occur in tests that are designed for use with brain-damaged patients and specifically with aphasics. These problems include the range of item difficulty, the need to clarify the nature of specific deficits revealed by the tests, the overlap of examinations for aphasia with measures of intelligence, the usefulness of a test in conjunction with recovery and therapy, and the overall conceptualization of the nature of aphasia.

**RANGE OF ITEM DIFFICULTY**

Range of item difficulty is usually determined by selecting from a range of "very easy" to "very difficult" items. In a well-constructed test, items should be homogeneously distributed; that is, the difficulty range (expressed in percentage of subjects passing each item) should rise in linear fashion from the first to the last item. The principle of homogeneity of item distributions is relatively easy to follow if we are dealing with a test for a normal population. However, if aphasics were given such a test, most items would be far too difficult for a majority of the patients. As a result, most aphasics would have scores in the bottom range of the distribution.
Consequently, aphasia tests must shift the difficulty of item distribution toward the lower or "easy" end to make it possible to discriminate mild, moderate, and severe levels of aphasia and to determine aphasic subtypes. This shift inevitably produces a "ceiling" effect if the test is then applied to normal subjects, as they would likely score at or near the 100% correct range. It is, of course, possible to include items that are easy enough to discriminate different degrees of aphasia as well as items difficult enough for a normal population. Such a test, however, could be lengthy and impractical.

OVERLAP WITH INTELLIGENCE TESTS

The overlap of assessment of aphasia with measures of intelligence, such as the Wechsler Adult Intelligence Scale (WAIS and its WAIS-R and WAIS-III revisions), has often gone unnoticed. It deserves special consideration in the context of item selection and in the context of other defects found in association with aphasia. It should be stressed that in the examination of aphasia the demands on the general intellectual abilities of the patient should be kept to a minimum. In addition, previously acquired knowledge of specific concepts and terms should influence the assessment of aphasia as little as possible.

The problem does not usually arise with the "easy" items used in aphasia tests. But when items for the "difficult" level are constructed, the separation of what is strictly language and what is intelligence becomes blurred. For example, naming tasks can be advanced to any level of difficulty by adding rare words and concepts that are likely to be found only in the vocabulary of the college-educated person of above-average intelligence. Tasks requiring definitions invariably tend to place higher value on abstract, elegant wording and penalize the uneducated, less intelligent subject. Tasks requiring oral arithmetic reasoning or the finding of superordinate concepts or similarities are, in fact, part of standard intelligence tests presently in use. For this reason, tests must be carefully scrutinized for content that exceeds the basic examination of language abilities. If such content cannot be avoided because of the range of item difficulty, the test must contain separate norms for patients of different ranges of intellectual and educational background or must apply adequate corrections for such factors.

USE IN MEASURING RECOVERY

The use of tests in the context of recovery and therapy poses two problems. The first is essentially an additional validity problem, that is, whether the test is suitable for the measurement of recovery. Tests adequate for the
measurement of recovery may be slightly different in content from tests that indicate the presence or type of aphasia, and may require more items in certain difficulty ranges to allow measurement of small steps in recovery. Related questions are the ability of the test to predict recovery or predict response to therapy, which must be established independently or in addition to other validation procedures. Interpretation may be complicated if changes over time merely reflect the patient's learning of the test ("practice effect") or remembrance of it, rather than actually indicating any underlying recovery.

CONCEPTUALIZATION OF THE NATURE OF APHASIA

The conceptualization of the nature of aphasia underlies many of the considerations outlined in this section. Test construction is directly influenced by whether we see aphasia as a specific disorder of selected abilities or as a pervasive disturbance of communication, and by whether we conceive of aphasia as unitary in nature or as consisting of many "subtypes." Benton (1967) pointed out that the choice of a model of language functioning determines what kind of test we construct or use. He indicated that the problem is similar to the one posed by the conceptualization of intelligence; it is similar also in the sense that no common agreement exists. Two approaches to test construction should be recognized as equally reasonable at this time:

1. To construct tests on the basis of one of the currently accepted conceptions of aphasia. This "taxonomic" or diagnostic approach ensures that the test measures all aspects viewed as important in a specific theoretical approach, but makes it probable that it will not be widely used as long as different conceptualizations of aphasia are held by other workers in the field.

2. To approach the problem pragmatically, avoid specific conceptualizations, and construct a test that contains a variety of probes of all abilities described by researchers of widely differing theoretical viewpoints. This pragmatic approach, quite commonly used in the field of intelligence testing, will not be fully satisfactory to any of the prevailing schools but may gain wider acceptance if the test instrument is otherwise well constructed and of demonstrated use in clinical practice.

Both approaches have been applied in the construction of currently used tests. Whether evaluation proceeds from an objective psychometric or from a theoretically based framework is a topic that has also been discussed by Kertesz (1994). In the following description of individual tests, we make specific reference to the conceptual framework used in each for the information of the reader unfamiliar with a given instrument.
Current Methods for the Assessment of Aphasia

The following selective review provides introductory information to readers unfamiliar with some of the assessment procedures to help them choose those methods most likely to meet their needs. Table 4.1 lists, in alphabetical order, the rating scales, and Table 4.2 the tests, that will be reviewed along with each test’s primary reference and the acronym used in this chapter. Information is also given on the test procedure itself, its psychometric properties, the theoretical position of the test authors, and the most likely areas of use.

Clinical Examination

The advantage of the clinical examination lies in its flexibility, brevity, and suitability for even severely physically impaired patients. The examiner can conduct a cursory examination at the bedside during acute recovery and follow up on any errors made by the patient by administering additional tasks and quickly skipping across areas of strength where there is no obvious impairment.

Numerous versions of the clinical examination have been recorded, in formal descriptions within the contexts of the mental status examination in neurology (Strub & Black, 1985), a general neurological examination (Poeck, 1974), and specifically designed clinical examinations (Benson, 1979b), as well as in individual case descriptions (e.g., Geschwind & Kaplan, 1962). Luria (1966) provided a detailed description of a clinical examination. The clinical examination usually includes examining spontaneous and conversational speech, repetition, comprehension of spoken language, word-finding (naming common objects and object parts), reading (headlines or text from newspapers or magazines), and writing (including dictation).

<table>
<thead>
<tr>
<th>Scale</th>
<th>Abbreviation used in text</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communicative Abilities of Daily Living</td>
<td>CADL</td>
<td>Holland (1980)</td>
</tr>
<tr>
<td>Communicative Effectiveness Index</td>
<td>CETI</td>
<td>Lomas et al. (1989)</td>
</tr>
<tr>
<td>Functional Abilities of Communication Profile</td>
<td>FACS</td>
<td>Frattali et al. (1995)</td>
</tr>
<tr>
<td>Functional Communication Profile</td>
<td>FCP</td>
<td>M. T. Sarno (1969)</td>
</tr>
<tr>
<td>Pediatric Evaluation of Disability Inventory</td>
<td>PEDI</td>
<td>Haley et al. (1992)</td>
</tr>
</tbody>
</table>
TABLE 4.2
Aphasia Assessment Instruments

<table>
<thead>
<tr>
<th>Full title</th>
<th>Abbreviation used in text</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aphasia Screening Test</td>
<td>AST</td>
<td>Reitan (1991)</td>
</tr>
<tr>
<td>Appraisal of Language Disturbances</td>
<td>ALD</td>
<td>Emerick (1971)</td>
</tr>
<tr>
<td>Arizona Battery for Communication Disorders of Dementia</td>
<td>ABCD</td>
<td>Bayles and Tomoeda (1990)</td>
</tr>
<tr>
<td>Auditory Comprehension Test for Sentences</td>
<td>ACTS</td>
<td>Shewan (1988)</td>
</tr>
<tr>
<td>Boston Assessment of Severe Aphasia</td>
<td>BASA</td>
<td>Helm-Estabrooks et al. (1989)</td>
</tr>
<tr>
<td>Boston Diagnostic Aphasia Examination</td>
<td>BDAE</td>
<td>Goodglass and Kaplan (1983)</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>BNT</td>
<td>Kaplan and Goodglass (1983)</td>
</tr>
<tr>
<td>Controlled Oral Word Association</td>
<td>COWA</td>
<td>Spreen and Benton (1977)</td>
</tr>
<tr>
<td>Discourse Comprehension Test</td>
<td>DCT</td>
<td>Brookshire and Nicholas (1997)</td>
</tr>
<tr>
<td>Minnesota Test for Differential Diagnosis of Aphasia</td>
<td>MTDDA</td>
<td>Schuell (1965, 1973)</td>
</tr>
<tr>
<td>Multilingual Aphasia Examination</td>
<td>MAE</td>
<td>Benton et al. (1994)</td>
</tr>
<tr>
<td>Neurosensory Center Comprehensive Examination for Aphasia</td>
<td>NCCEA</td>
<td>Spreen and Benton (1977)</td>
</tr>
<tr>
<td>Pantomime Recognition Test</td>
<td>—</td>
<td>Benton et al. (1983)</td>
</tr>
<tr>
<td>Phoneme Discrimination Test</td>
<td>—</td>
<td>Benton et al. (1983)</td>
</tr>
<tr>
<td>Porch Index of Communicative Ability</td>
<td>PICA</td>
<td>Porch (1981)</td>
</tr>
<tr>
<td>Reporter’s Test</td>
<td>—</td>
<td>De Renzi (1980)</td>
</tr>
<tr>
<td>Sklar Aphasia Scale</td>
<td>SAS</td>
<td>Sklar (1973)</td>
</tr>
<tr>
<td>Sound Recognition Test</td>
<td>SRT</td>
<td>Spreen and Benton (1974)</td>
</tr>
<tr>
<td>Token Test</td>
<td>TT</td>
<td>many versions, see text</td>
</tr>
<tr>
<td>Western Aphasia Battery</td>
<td>WAB</td>
<td>Kertesz (1982)</td>
</tr>
</tbody>
</table>

Because the clinical examination varies greatly in form as well as detail from one setting to another, we do not attempt a comprehensive evaluation of different examination methods. Clinical examination skills must be acquired under close supervision in a clinical setting.

In her reevaluation of the Short Examination for Aphasia, Schuell (1966) carefully debated the merits of the clinical examination in comparison to the comprehensive test. She stressed that only a comprehensive test can assess all aspects of “aphasia, [which] deals with one of the most complex and perhaps the only unique function of the human brain” (p. 138).

Screening Tests

Screening tests are designed to screen quickly for the presence or absence of aphasia. They are described here as screening tests because they
do not claim to provide a detailed description of the aphasic disorder but rather to check for and focus the direction on the problem if aphasia is present. Screening tests have been developed by Fitch-West and Sands (1987), Keenan and Brassell (1975), Lecours, Mehler, Parente, and Beltrami (1988), and Orzeck (1964; Inglis & Lawson, 1981), but have found only limited use. The Frenchay Aphasia Screening Test is a British measure that has led to some additional investigations (Al-Khawaja, Wade, & Collins, 1996; Enderby & Crow, 1996).

**APHASIA SCREENING TEST**

The Reitan (1991; Reitan & Wolfson, 1985; Wheeler & Reitan, 1962) version of the Halstead-Wepman screening test is designed to determine whether the patient can perform such simple tasks as spelling a word or naming an object. The Aphasia Screening Test (AST) procedures are such that the clinician should elicit the patient’s best performance. A broad array of language functions is briefly assessed by one or two items each. For example, the patient is required to draw a shape, name it, and spell it; to read (e.g., “See the black dog”); to do a single pencil-and-paper and a single “in-head” arithmetic problem; and to demonstrate object use and picture drawing.

The test takes approximately 20 min to complete. The test manual provides many illustrative examples of performances. The test is usually given within the context of a complete neuropsychological test battery intended to assess the full range of psychological deficits after brain damage. Interrater reliability has been reported as high (Barth, 1984). The screening efficiency of the test as a single measure (discrimination between aphasic and nonaphasic brain-damaged patients) has been reported as 80% correct (Krug, 1971). Ernst (1988) reported normative data for elderly subjects.

It should be noted that Barth (1984), Goldstein and Shelley (1984), and Werner, Ernst, Townes, Peel, and Preston (1987) reported significant correlations between the AST and IQ as well as educational level in large neuropsychiatric populations. In a study by Snow (1987), only 1 of 33 items of the AST differentiated between right- and left-hemisphere lateralized tumor and stroke patients.

**SKLAR APHASIA SCALE**

The Sklar Aphasia Scale (SAS) (Sklar, 1973) provides a brief assessment of the aphasic patient’s abilities along four dimensions: auditory decoding, visual decoding, oral encoding, and graphic encoding. Each of the four subtests is represented by five areas that each comprise five items. The SAS is constructed within a framework of a decode (input), transcode (process), and encode (output) model of language and its disabilities.
Each response on the SAS is scored on a 5-point scale, from a "correct" response (0) to no response (4). An impairment score for each test is obtained by finding the mean value of the four subtest impairment scores ($0 = \text{no impairment}, 100 = \text{full impairment}$). The total impairment index may be used prognostically in terms of potential benefit of therapy if modified by both the recency of the impairment and the patient's overall state of health. However, such prognostication is based only on a simple index score (e.g., an aphasic performing with a score of 70 has a better prognosis than an aphasic scoring 15), and psychometric evaluations of the prognostic significance of the SAS have not yet been presented.

The SAS test items were standardized on a sample of only 20 adults ranging in age from 29 to 78 years. The test author reported high correlations between SAS performance and performance on Eisenson's aphasia examination, Schuell's short version of the Minnesota Test for Differential Diagnosis of Aphasia (MTDDA), and the Halstead-Wepman AST in a sample of 12 patients. Although reliability data are not presented, the author describes five studies on the validity of the SAS as an instrument to assess language ability in aphasics.

**Tests of Specific Aspects of Language Behavior**

Several tests have been constructed for the detailed assessment of a specific language function. Such tests make no claim to cover all aspects of aphasia but do provide a relatively thorough assessment of the function in question. Because such functions are usually central to the aphasic disorder, however, these tests may also provide a reasonable discrimination between aphasic and nonaphasic patients in general. Some of these tests have been used as screening devices because of their good discrimination, although this was not necessarily the intent of the test authors.

**AUDITORY COMPREHENSION TEST FOR SENTENCES**

The Auditory Comprehension Test for Sentences (ACTS) (Shewan, 1988) provides an examination of auditory comprehension. The sentences are spoken by the examiner, and the patient must point to one of four visual stimuli to indicate which correctly represents the spoken sentence. Four training trials are permitted, which also serve as a screening device to determine whether the patient is too impaired to perform the entire task. These trials are followed by the 21 test items, which vary along parameters of sentence length, vocabulary difficulty, and syntactic complexity. Pass-fail scoring and qualitative error analysis are possible; the use of each system is made easy by the clear and concise ACTS response sheet. An average of 10 to 15 min is required to complete the ACTS.
A score of 18 of 21 (i.e., approximately 2 SD below the mean) is considered to be the lower bound of normal limits for adults with at least an elementary school education. Shewan (1988) reported an internal consistency correlation of .82, as well as a test–retest reliability of .87. Flanagan and Jackson (1997) recently reconfirmed the test–retest stability of the measure in a sample of neurologically intact individuals aged 50 to 76 years. Two measures of validity are provided in the manual. First, a correlation of .80 was obtained between the ACTS and an 8-point clinical rating of functional auditory–verbal comprehension. Second, when compared with established tests, there were correlations of .52 between the ACTS and the BDAE auditory comprehension section, and .89 with the WAB comprehension section. The lower ACTS–BDAE correlation was attributed to the wider range of abilities assessed by the BDAE compared with the ACTS.

Information about the ACTS standardization sample of 150 aphasics and 30 normal controls is provided in the manual. Means and standard deviations of the performance of aphasic subtypes (diagnosed clinically) are included: Wernicke’s aphasics performed the poorest, followed in turn by Broca’s aphasics, aphasics with anomia, and normal controls. The ACTS (along with the WAB) has been used to examine recovery and the differential impact of treatment in a subtyped sample of 100 aphasics (Shewan & Kertesz, 1984); improvements over time were similar for treated and untreated patients.

The ACTS is brief, easy to administer and score, and requires only simple nonverbal responses by the patient. The test has demonstrated reliability and validity, and appears to be a promising test of comprehension for sentence-length material with systematic variations of sentence length, difficulty, and complexity.

BOSTON NAMING TEST

The Boston Naming Test (BNT) (Kaplan & Goodglass, 1983; Kaplan, Goodglass, & Weintraub, 1978) has emerged as a popular test of visual confrontation naming not only for aphasia but also in dementia and other geriatric work. The current 60-item version has several variants: the original 85-item experimental form (Kaplan et al., 1978), the short 15-item version which is part of the CERAD screening battery for dementia (Morris, Mohs, Hughes, Van Belle, & Fillenbaum, 1989), two 42-item equivalent versions (Huff, Collins, Corkin, & Rosen, 1986), and four 15-item short versions, drawn from the full-length, 60-item test (Mack, Freed, Williams, & Henderson, 1992). A Spanish adaptation is available (Taussig, Henderson, & Mack, 1988; Ponton et al., 1992). Morrison, Smith, and Sarazin (1996) used the test with normal French-speaking subjects in Quebec, Canada. The test may be suitable for children from age 4 years.
The BNT stimuli are line drawings of objects with increasing naming difficulty, ranging from simple, high-frequency vocabulary (tree) to rare words (abacus). Administration requires a spontaneous response within a 20-sec period; if such a response is not made, two kinds of prompting cues (one phonemic, one semantic) may be given. Rules allow for discontinuation and for starting the test at an advanced level, thus saving considerable time for subjects without obvious impairment. Scoring counts the number of spontaneously produced correct responses, the number of cues given, and the number of responses after phonemic cuing and after semantic cuing. M. Nicholas, Obler, Au, and Albert (1996) developed a rating of BNT errors based on relatedness to the correct stimuli.

Reliability has been assessed in a number of independent settings. Test-retest reliability after 8 months in 51 adult intractable epileptics was reported as .94 (Sawrie, Chelune, Naugle, & Zuders, 1996). Henderson, Mack, Freed, Kemperer, and Andersen (1990) reported an 80% response consistency for both uncued and cued responses in Alzheimer’s disease (AD) patients after 6 months. Huff et al. (1986) divided the original BNT into two equivalent forms and obtained between-forms correlations of .81 in healthy control subjects and .97 in patients with AD. Thompson and Heaton (1989) compared the 85-item form with the standard 60-item form, and with the two nonoverlapping 42-item versions in 49 clinical patients. They found correlations of .96 to .84. The authors recommend the use of the short forms, as they may be more suitable if repeat testing is required. Another study constructed an “odd-item” and “even-item” version, as well as an experimental version of the BNT and found that all three short versions successfully discriminated AD, other dementing diseases, and normal older (mean age 73.7) subjects (Williams, Mack, & Henderson, 1989). Another 30-item version, developed for a Chinese population, showed a sensitivity between 56 and 80%, and a specificity between 54 and 70% in separating demented and nondemented subjects of low and high educational background (Salmon, Jin, Zhang, Grant, & Yu, 1995).

Age stability is a common finding in healthy elderly subjects (Ganguli, Seaburg, Ratcliff, Belle, & DeKosky, 1996; Mitrushina & Satz, 1995). However, Lansing, Randolph, Ivnick, and Cullum (1996) examined various short forms with a population of 717 controls and 237 AD subjects in the age range of 50 to 98 years and found significant correlations with age and education as well as gender effects for all forms, including the original full-length version. Correct classification rates varied from 58 to 69% for AD patients, and from 77 to 87% for normal controls. The authors used a discriminant function analysis to develop an empirical 15-item version balanced for gender. K. A. Hawkins et al. (1993) also found correlations between .74 and .87 between the Gates–McGinitе Reading Vocabulary Test
and the BNT across normal and clinical adult populations; they demonstrated that norms for the test may lead to many false-positive rates for naming deficit, and that corrections should be applied, especially for subjects with a lower-than-average reading level. Concurrent validity with the Visual Naming Test of the Multilingual Aphasia Examination (MAE) (Benton, Hamsher, Rey, & Sivan, 1994) was described by Axelrod, Ricker, and Cherry (1994).

The manual provides means for aphasics with a BDAE severity level of 0 to 5, which are well below the level for normal adults. However, the range for aphasics with severity levels 2 to 5 extends well into the range for normals. This is not surprising, because naming is not necessarily impaired in all types of aphasia. Sandson and Albert (1987) found that aphasic patients made more perseverative errors than patients with right-hemisphere lesions; furthermore, perseverations were more frequent in patients with posterior rather than frontal lesions.

Knopman, Selnes, Niccum, and Rubens (1984) reported good measurement of recovery of naming after strokes of small volume in the left posterior superior temporal—inferior parietal and the insula—putamen areas. Welsh et al. (1995) found that semantic errors and circumlocutions in AD patients were associated with left mesial and lateral temporal lobe metabolism, as measured by positron-emission tomography (PET) and fluorodeoxyglucose (FDG) emission techniques. The left anterior temporal area has also been implicated (Tranel, 1992). However, Trenerry et al. (1995) reported that carefully limited anterior right or left temporal lobectomy in 31 left and 24 right lobectomy patients with left-hemisphere language lateralization did not impact positively or negatively on BNT performance. The BNT was also not sensitive as to side of epileptic focus in a study of patients with idiopathic epilepsy (Haynes & Bennett, 1990) and in patients with anterior temporal lobectomy (Cherlow & Serafetinides, 1976). The test is sensitive to subcortical disease (multiple sclerosis and Parkinson’s disease), even when global mental status is only mildly affected; in addition, responses were slower than in normal controls (Beatty & Monson, 1989; Lezak, Whitham, & Bourdette, 1990).

As with other tests, visual—perceptual integrity should be checked if errors occur. Kaplan and Goodglass (1983) noted that, particularly in patients with right frontal damage, “fragmentation responses” may be made (e.g., the mouthpiece on a harmonica is interpreted as a line of windows in a bus; Lezak, 1995).

**CONTROLLED ORAL WORD ASSOCIATION**

The Controlled Oral Word Association (COWA) provides a fast, efficient assessment of verbal fluency. The task requires the patient to produce in a
limited time period (usually 1 min) as many oral word associations as possible to a specific letter of the alphabet or to some category, such as animals or foods. The test does not measure verbal fluency of conversational speech. Deficits in aphasic patients are common, but defective COWA has also been studied in patients with dementia, in nonaphasic patients with anterior left-hemisphere lesions, and in patients with frontal lobe pathology manifesting disorders of executive functions. Several standardized tests are available. The first, developed by Spreen and Benton (1969, 1977), was introduced as a subtest of the NCCEA, but has been widely used as a single test. Other versions were created by Benton et al. (1994) as part of the MAE, by Goodglass and Kaplan (1983) as part of the BDAE, and by R. T. Wertz (1979).

Spreen and Benton's version requires the patient to say as many words as possible that begin with the letters F, A, and S within 1-min time periods for each letter. Proper names and words that differ only in suffix are not acceptable; performance is gauged in terms of the sum of admissible words in all three trials. Normative data, as well as corrections for age and level of education, are available (Spreen & Benton, 1977; Spreen & Strauss, 1998).

The letters used in the NCCEA version are of the "easy" level. Borkowski, Benton, and Spreen (1967) examined the number of associations by normal adult females for the letters of the alphabet. The number of associations was related to the difficulty, as defined both by the Thorndike-Lorge (1944) word count (.80) and the number of words per letter in Webster's New Collegiate Dictionary (.74). The authors also reported that a heterogeneous sample of brain-damaged patients performed less well than normal adults at all levels of severity, lending validity to the testing method. Patients with low IQ scores were better differentiated with easy-level letters, whereas patients with high IQ scores were better distinguished with more difficult letters.

Interscorer reliability is very high, and 1-year test–retest reliability has been reported as .70 (Snow & Tierney, 1988). Concurrent validity has been established in several studies, generally indicating better validity for letters than categories (Coelho, 1984), although in dementia and the subcortical damage category (semantic) fluency tends to be more affected than letter association (phonetic fluency) (Butters, Granholm, Salmon, Grant, & Wolfe, 1987; Monsch et al., 1994; Monsch, Bondi, Butters, & Salmon, 1992). Correlation with age, education, WAIS-R Verbal IQ, and Performance IQ is quite low (Yeudall, Fromm, Reddon, & Stefanyk, 1986). FAS-COWA was part of the computerized Alzheimer's Disease Assessment Battery (Bracco, 1986) and discriminated well between AD and normal control subjects. Consistent with clinical–anatomic expectations FAS-COWA corre-
lates with glucose metabolic rates in left prefrontal regions in poststroke aphasic patients (Karbe, Kessler, Herholz, Fink, & Heiss, 1995). Benton et al.'s MAE version (1978; Benton, 1994) differs from the NCCEA in using three letters of progressively increasing associative difficulty, otherwise the testing is quite similar. Two equivalent versions (CFL and PRW) are available for repeated or follow-up administrations.

The BDAE uses category (animals) COWA, a format traditionally used in child evaluation settings. Patients are instructed to say as many different animal names as possible. A 90-sec recording period is allowed, with responses recorded in 15-sec blocks. The best 60-sec period (usually the first 60 sec) is scored. COWA tests are quick and simple. The versions described here have proven discriminative validity. COWA tests may not be very sensitive in distinguishing at lower levels of ability, but they are capable of screening for the presence of less severe disability.

DISCOURSE COMPREHENSION TEST

The Discourse Comprehension Test (DCT) (Brookshire & Nicholas, 1997) arose out of psycholinguistic research (L. E. Nicholas & Brookshire, 1995), which suggested that word and sentence comprehension tests are unlikely to predict comprehension of multiple-sentence messages and are unable to examine contextual aspects of such messages, that is, the ideas that are implicit across sentences. The authors also cite work by Brownell (1988) that suggests that word and sentence comprehension measures may underestimate comprehension problems in patients with right-hemisphere lesions.

The DCT has a narrative format. Paragraph-length stories are played for the patient on audiotape. The patient is queried after each story. Eight yes/no questions are divided as follows: two deal with the story's main ideas, two with implied main ideas, two with stated details, and two with implied details. The authors found that regardless of group, normals and aphasics score higher on main ideas than on details, and higher on stated than implied material. Ten stories are divided into two equivalent 5-story sets. After one or two practice stories, one set is administered; if performance is variable, the second set is given.

Scores on the DCT include an overall score and scores for each type of question. The test uses cutoff scores to denote defective performance, set at the 5th percentile of the normal sample. Standardization was based on 40 normal adults, aged 55 to 75 years, with a range of education from 8 through 18 years. Performances by 20 aphasics, 20 patients with right-hemisphere lesions, and 20 patients with traumatic head injury were reported. Overall scores correlated minimally with age and mildly with education. Test–retest reliability after 2 to 3 weeks was .87 for aphasics and
.95 for patients with right-hemisphere lesions. Studies of content validity are reported in the manual. Criterion-related validity ranged from .54 to .76 for the DCT with the BDAE Auditory Comprehension subtests, and with the four-subtest short version of the PICA (DiSimoni, Keith, & Darley, 1980).

The manual includes instructions for a silent-reading comprehension administration of the DCT with the same stimuli. The authors' Sentence Comprehension Test is also included, which is used to determine whether consistent yes/no responses can be elicited.

The DCT fills a niche in the assessment of comprehension at the level of narrative discourse which is only approximated in other assessment batteries. The authors report that results may generalize to the level of highly structured communicative dialogue, but it remains unclear whether it can be generalized to less structured dialogue. In addition to helping describe the communicative skills of brain-damaged patients, the DCT offers a potentially useful research instrument at a level between concise measures of single word and sentence comprehension (e.g., the ACTS) on the one hand, and subjective functional communication rating scales on the other. A developmental study (Newton, 1994) reported that with young children the use of a picture that depicts the initial circumstances of a narrative story helps them maintain a functional mental model to supplement comprehension as the story progresses. The use of such pictures for the DCT might provide a useful treatment-oriented adjunct, as well as an opportunity to compare patient performance with and without such visual aids.

The question remains whether the length of the stimuli and the number of questions per story introduce an undesirable attentional-memory component into the test which may influence performance. For example, Tompkins, Bloise, Timko, and Baumgaertner (1994) commented on the relationship between memory capacity and discourse passages that require a revision of the plot as the story unfolds.

The weakness of the DCT is the relatively small size of the standardization sample. It remains open whether the cutoff scores would remain stable with a larger sample of normal controls. Larger samples of both aphasic and control subjects would also permit a more refined percentile scoring system.

"IOWA-GROUP" TESTS

In addition to developing two comprehensive aphasia batteries, Benton and his colleagues have developed a number of specific-function tests for the evaluation of aphasic patients. Some of these tests are described briefly.

The SOUND RECOGNITION TEST (SRT; Spreen & Benton, 1974; Spreen &
Assessment of Aphasia

Strauss, (1998), presented on audiotape, examines auditory object recognition (e.g., train whistle) by requiring the identification of familiar sounds. The original form of the SRT presented two equivalent forms of 13 items each. Three modes of administration were offered: verbal response, pointing to one of four multiple-choice names, and pointing to one of four multiple-choice line drawings. Varney (1980, 1984b) modified the format such that all 26 items are presented and multiple-choice line drawings are used as the response format. Scoring standards are detailed in Spreen and Strauss (1998). Spreen and Benton (1974) described norms for normal adults and children, and the performance of brain-damaged adults. Impairments of sound recognition are typically limited to aphasic patients and are frequently associated with aural comprehension deficits. Aphasics with aural comprehension deficits during the acute stage of recovery from cerebrovascular accident but with normal sound recognition show rapid and near complete recovery of aural comprehension, whereas patients with acute aural comprehension and sound recognition deficits show a much poorer outcome (Varney, 1984b). Defects in sound recognition are associated with a number of lesion sites, none of which appear specific for the manifestations of the defect (Varney & Damasio, 1986).

The PANTOMIME RECOGNITION TEST (Benton, Hamsher, Varney, & Spreen, 1983), presented on videotape, shows a male pantomiming different activities (e.g., using a telephone). The patient is presented with four multiple-choice responses: one correct, one semantic error, one neutral error, and one odd error. Defective pantomime recognition is an infrequent finding, observed most frequently in aphasic patients, and appears to be related to associated defects in reading comprehension level. Demented patients may perform defectively on the task. Taking another approach, Records (1994) designed an experimental test evaluating the impact of gestures on comprehension in aphasic patients.

The PHONEME DISCRIMINATION TEST (Benton et al., 1983) is a brief test of the discrimination of phonemic sounds. The test consists of 30 pairs of one- or two-syllable nonsense words on audiotape. The patient has to indicate whether the pair members are the same or different. Practice or pretraining is encouraged to determine whether the patient can make same–different responses reliably. Normative data and samples of 100 aphasic left-hemisphere damaged and 16 nonaphasic right-hemisphere damaged patients are reported in the test manual. Comparisons of the performance of aphasic patients on the Phoneme Discrimination Test with the MAE Aural Comprehension Test are also reported (see also Varney & Benton, 1979). Varney (1984a) presented longitudinal data on the Phoneme Discrimination Test in the context of evaluating additional measures of comprehension.
PEABODY PICTURE VOCABULARY TEST-III

The Peabody Picture Vocabulary Test-III (PPVT-III) (Forms A and B, 3rd ed., Dunn & Dunn, 1997) assesses auditory comprehension of picture names, and is also part of the Florida Kindergarten Screening Battery (Satz & Fletcher, 1982). It was initially constructed as a test of hearing vocabulary in children, but has since been standardized for adults and used with a variety of clinical populations. The test requires the subject to choose one of four items displayed on a card as depicting the word spoken by the examiner. After 5 training items, 204 items of increasing difficulty can be given, but usually only 35 to 45 items need to be administered if a suitable entry point (six consecutive correct) is chosen; the test is discontinued after consecutive failures on 6 out of 8 items. It has been restandardized for the age range of 2½ to 90+ years of age. The time required is about 10 to 20 min. In the new edition, a number of items have been revised or added to correspond well to the negatively accelerating growth curve of vocabulary with age. A Spanish version is available.

The new edition has been standardized on a sample of 2725 subjects representative of the 1994 U.S. census estimate, ranging in age from 2½ to 90 years. Canadian norms are available from the distributor (Psycan). Kamphaus and Lozano (1984) note that in 6- to 11-year-old children with Spanish surnames (about one half of whom spoke Spanish at home), standard scores were about 12 to 13 points below the national norms, although the scores showed regular, expected increases with age. Sattler (1988) recommends special care in the interpretation of scores of ethnic minority children because their scores tend to be lower and reflect their verbal and experiential differences rather than ability.

The score on this test is simply the number of items passed including the items before the entry point. The manual allows translation of these scores into "age equivalents," "standard score equivalents," stanines, and percentiles. The authors have added a "true confidence band," indicating the range of scores in which the subject's true score can fall 68 times out of 100.

Split-half reliability has been reported as ranging from .61 to .88 in children and adolescents, and .82 in adults. The reliability of alternate forms ranged from .73 to .91 (Stoner, 1981; Tillinghast, Morrow, & Uhlig, 1983). Retest reliability with the alternate form after a minimum of 9 days showed a median correlation of .78. In children, retest stability over 11 months has been reported as .84 for the revised PPVT (Bracken & Murray, 1984), .81 in retarded children over a period of 7 months (Naglieri & Pfeiffer, 1983), and .71 in a mixed clinical neuropsychiatric population after 2½ years (Brown, Rourke, & Cicchetti, 1989). Internal consistency ranged from .96 to .98 in
children 6 to 11 years old (Kamphaus & Lozano, 1984). Construct validity of the test as a measure of scholastic aptitude is good (Hinton & Knights, 1971). Bracken and Murray (1984) report a predictive validity of .30 with spelling, .54 with reading recognition, .58 with reading comprehension, and .59 with the total Peabody Individual Achievement Test (PIAT) for the revised PPVT; similar values were reported for the first edition (Naglieri, 1981) and for mentally handicapped children with the revised edition (Naglieri & Pfeiffer, 1983). Concurrent validity with similar tests—the Bracken Basic Concept Scale, the preschool version of the Boehm Test of Basic Concepts and its revised version—were .68, .65, and .62, respectively (Zucker & Riordan, 1988). Because vocabulary is the single most important subtest of most intelligence tests, the test also correlates with the WISC-R, although the results of several studies are somewhat contradictory. In children, correlations with measures of verbal (.87), performance (.80), and full-scale IQ (.88) have been reported (Crofoot & Bennett, 1980). The PPVT also correlates highly with the McCarthy Scales in children (Naglieri, 1981), with the 1986 Stanford-Binet Intelligence Scale (Carvajal, Gerber, & Smith, 1987), and with the achievement scale of the Kaufman Assessment Battery for Children in a learning-disabled population (.78; D’Amato, Gray, & Dean, 1987). However, Faust and Hollingsworth (1991) found a correlation of only .34 with the WPPSI-R Full-Scale IQ, and of .30 and .31 with the PIQ and VIQ, respectively, in normal preschoolers. Altepeter (1989; Altepeter & Johnson, 1989) found only modest correlations in healthy adults with the WAIS-R (.47 with full-scale IQ) and warns that in this age range the test tends to overestimate IQ in the lower ability ranges and to underestimate IQ in the higher ability ranges; in a cross-tabulation of IQs in 10-point steps, less than one half of the clients were classified correctly. Price, Herbert, Walsh, and Law (1990) reported similar discrepancies in adult psychiatric inpatients.

Hollinger and Sarvis (1984) also stress the role of perceptual–organizational ability in PPVT performance of school-age children, and L. J. Taylor (1975) reached the same conclusion for preschool children based on a factor analysis of the WPPSI and the ITPA as well as the PPVT. Children with impaired oral language production (Rizzo & Stephens, 1981) and nonpsychotic, emotionally disturbed adolescents (Dean, 1980) tend to produce variable results. Elliott et al. (1990) found also that the PPVT results in children between 6 and 11 years of age with normal pure-tone hearing were strongly influenced by the ability to make fine-grained auditory discriminations, with consonant–vowel stimuli varying in timing and place of articulation. The PPVT showed significant differences between 28 young adults with specific language impairment and 28 controls (means of 82.36 and 92.79, respectively; Records, Tomblin, & Buckwalter, 1995). Das,
Mishra, Davison, and Naglieri (1995) demonstrated that the PPVT is sensitive to mental decline (dementia) in older (age greater than 50 years) Down's syndrome patients, with results parallel to those in a dementia rating scale.

The test is relatively nonthreatening and requires little verbal interaction; because it also allows gestural or pointing responses, the test is suitable for language-impaired as well as autistic or withdrawn patients. The auditory and visual-perceptual integrity of the patient should be carefully considered in interpreting the results of this test. Considering the extensive research with the PPVT and its continuing revision, it is clearly the preferred measure of vocabulary for children and is now valid for adult and geriatric populations.

REPORTER'S TEST

De Renzi used the stimuli and most of the commands from the Token Test to construct the Reporter's Test, a test for expressive deficits in aphasics (De Renzi, 1980; De Renzi & Ferrari, 1979). The Reporter's Test was designed to meet two specific goals: (a) to elicit organized speech and (b) to limit the range of what the patient is expected to say. Although picture description tasks (e.g., the BDAE "Cookie Theft" card) adequately fulfill the first goal, they fail on the second. Other task-description tests have been used in nonaphasic clinical settings (e.g., the "dice game"; McDonald & Pearce, 1995), but lack the range limitations of the Reporter's Test. The patient is required to act as a "reporter"; that is, she or he must report the actions of the tester to an imaginary third person. For example, if the examiner were to touch the large red circle, the patient must verbalize the relevant information necessary for a third person to reproduce the tester's actions ("Touch the large red circle"). The Reporter's Test begins with several sample items to acquaint the patient with the task. The test comprises five sections; the first four are taken from the Token Test. De Renzi recommended that it be used after the Token Test, so that the patient is acquainted with the stimuli and the required commands.

De Renzi (1980) described initial findings for the Reporter's Test in discriminating 24 aphasic patients from 40 hospitalized, nonaphasic, non-brain-damaged controls. In this study, an actual third person sat next to the patient and performed the instructions given by him or her. Years of education, but not age, were significantly related to performance; scores were corrected to account for education. Using a cutting score expected to produce 5% false positives, a 97% hit rate was obtained. Classification accuracy was higher for the Reporter's Test than for four other tests of verbal expression: visual naming, oral fluency, sentence repetition, and story telling.

De Renzi and Ferrari (1979) described aphasic performance using the
original pass-fail scoring, partial credit for correct performance after repetition, and a weighted scoring (1 point for each bit of information on a trial but without credit for repetition). Aphasic patients, nonaphasic left brain-damaged patients, and nonaphasic right brain-damaged patients were discriminated with an 82% hit rate. The authors recommended the use of both scoring systems to offset the weaknesses of each: lower classification for the weighted system and overly severe evaluation with the pass-fail scoring. Unfortunately, the test has not been used as widely in clinical settings and psychometric research as the Token Test.

TOKEN TEST

The Token Test (TT) was introduced as a brief test by De Renzi and Vignolo in 1962 to examine subtle auditory comprehension deficits in aphasic patients, by having patients respond gesturally to the tester’s verbal command. Since its inception, the original TT has been used, modified (De Renzi, 1980; De Renzi & Faglioni, 1978), and included in some batteries (Benton et al., 1994; Spreen & Benton, 1977). The original test has spawned many variants: short forms (Boller & Vignolo, 1966; Spellacy & Spreen, 1969; Van Harskamp & Van Dongen, 1977), a concrete-objects version (Martino, Pizzamiglio, & Razzano, 1976), a format with both auditory and visual presentation of commands (Kiernan, 1986), a TT “battery” (Brookshire, 1978), and a version with expanded linguistic examination (McNeil & Prescott, 1978). Equivalent versions in several languages are available (e.g., Italian, German, Portuguese; Fontanari, 1989; Kannada; Vena, 1982).

The TT is a portable test that, in most versions, contains 20 plastic token stimuli of two sizes (large and small), two shapes (square/rectangular and round), and five colors. The tokens are laid out in front of the patient, typically in a standard 4 × 5 matrix. The test has a variable number of sections that increase in sentence length and linguistic complexity (e.g., from “Point to a square” to “Put the small red square on the large blue circle”). The McNeil and Prescott (1978) version provides the most complex commands.

Some authors have reported age (Emery, 1986) and level of education (De Renzi, 1980; De Renzi & Faglioni, 1978) differences on certain versions of the TT. Gallaher (1979) reported day-to-day retest reliabilities for one version of the TT and its subsections to be greater than .90. Validation studies have shown the TT to be a strong and accurate discriminator between the performance of aphasic patients and that of normal hospitalized adults (De Renzi, 1980), nonaphasic right-hemisphere-damaged adults (Boller & Vignolo, 1966; Swisher & Sarno, 1969), and nonaphasic diffuse and focal brain-damaged adults (Orgass & Poeck, 1966). Morley, Lundgren, and Haxby (1979) found the TT to discriminate particularly well between normals and aphasics with high levels of ability compared with discrimina-
tions on the BDAE comprehension section and the Porch Index of Commu-
nicative Ability (PICA). Poeck, Kerschensteiner, and Hartje (1972) demon-
strated independence of TT performance and the fluency–nonfluency di-
mension in aphasic patients.

Cohen, Kelter, and Shaefer (1977), and Cohen, Lutzweiler, and Woll
(1980) studied construct validity and other aspects of TT validity. The
memory component of TT performance was examined by R. Lesser (1976),
Cohen, Gutbrod, Meier, and Romer (1987), and Gutbród, Mager, Meier,
and Cohen (1985), who concluded that the test measures deficits in the
short-term storage of highly specific information in aphasics. In contrast,
Riedel and Studdert-Kennedy (1985) claimed that a general cognitive
deficit is responsible for poor TT performance. In AD patients, Swihart and
Panisset (1989) found that a short version of the test correlated only weak-
ly with other simple auditory–verbal comprehension tasks, but correlated
highly with the Mini-Mental-State Examination because of strong perse-
verative tendencies found in that patient population.

The TT has maintained consistent popularity as both a clinical and an
investigative instrument, and has been examined for use as a therapeutic
tool (Holland & Sonderman, 1974; West, 1973). Two major compilations of
work with the TT are available (Boller & Dennis, 1979; McNeil & Prescott,
1978), and at least three English-language versions are commercially avail-
able (Benton et al., 1994; McNeil & Prescott, 1978; Spreen & Strauss, 1998).

The TT's advantages lie in sound discriminative validity, portability,
and brief administration time. Brookshire's (1973) early advice remains
valid: The clinician should keep in mind that although it is a sensitive in-
dicator of comprehension deficits, the TT relies on a limited stimulus ar-
ray. Other tests of auditory comprehension (e.g., the ACTS) may be used
to supplement the TT. Rao (1990) also points out that the test introduces a
somewhat artificial test situation and therefore has less "eco-validity" than
other, functional communication measures. Other comprehension tests in-
clude Lexical Understanding with Visual and Semantic Distractors (LUVS;
Bishop & Byng, 1984), which focuses on semantic comprehension, and an
object-manipulation test designed to measure syntactic comprehension
(Caplan, 1987).

**Rating Scales**

Rating scales assume a position somewhere between clinical assessment
and psychometric tests. The clinician who assigns a label of mild, moder-
ate, or severe to the symptoms of a patient is actually providing a basic rat-
ing of severity. Ratings are frequently used (a) as a summary judgment of
severity of any symptom or syndrome, (b) as a specific judgment of aspects
of a patient’s behavior that cannot be readily measured, (c) to rate the “quality of life” (QOL), or (d) to represent an entire class of functioning. A specific example of (a) is a professional’s summary judgment in disability or worker’s compensation cases. The second application has its origin in ratings of “activities of daily living” (ADL), frequently used by occupational therapists to rate numerous aspects of the patient’s behavior and to measure the progress of the patient’s general behavioral impairment during therapy. Some aspect of the patient’s speech behavior—for example, ability to communicate in the home setting—is rated on a scale of levels or points. QOL ratings, application (c), have also become an important aspect of the success of rehabilitation (M. T. Sarno, 1997). In this context, QOL addresses the question of “how a person experiences aphasia, how its meaning influences his/her behavior and interaction . . . A total of human response to an unexpected and unwanted life event” (M. T. Sarno, 1997, p. 675), and includes psychosocial factors such as loneliness, difficulty making friends, lowered self-esteem, and depression. Specific instruments have been designed to measure QOL (Caregiver Burden Interview, Zarit, Reever, & Bach-Peterson, 1980; Functional Life Scale, J. E. Sarno, Sarno, & Levita, 1973; Geriatric Evaluation of Relative’s Rating Instrument, Schwartz, 1983).

The last application is best demonstrated by the increasing use of functional independent measures (FIMs) (Frattali, 1993; Hamilton, Granger, Sherwin, Zielzny, & Tashman, 1987) in rehabilitation units, which include items ranging from sphincter control and feeding to communication. Such functional communication ratings have gained further importance in recent years because they are often used for public policy and reimbursement issues (Frattali, 1992, 1993). Progress in aphasia rehabilitation is now frequently defined as progress in functional communication, although Crockford and Lesser (1994) state that the actual use of such scales by practicing clinicians seems to be limited. Such measures sometimes attempt to reduce entire cognitive domains, such as communication, to a single scale of independence and various levels of dependence. It should be noted that single FIMs representing entire cognitive domains such as language or memory would seem to be of questionable value and validity.

Several comprehensive batteries, such as the BDAE, MAE, and Western Aphasia Battery (WAB), include qualitative rating scales as part of their set of subtests. Usually, a rating scale should not exceed seven points (from “normal” to “very severe”), since it has been demonstrated that the use of more than seven points does not enhance the accuracy of the ratings but merely provides a false impression of greater accuracy.

Rating scales should be subjected to careful interjudge reliability studies. Reliability can be improved if careful descriptions of each rating point
are provided. For example, instead of marking the lowest point as "normal," and the highest as "very severe," each point should be illustrated in as much detail as possible with examples. Rating scales are no substitute for psychometric testing, but they are extremely valuable if the information being rated cannot be readily tested or is too complex to be documented in test scores. Ratings of communicative ability in the home or in a conversational setting are often made by an informant (e.g., a relative or a member of the nursing staff) rather than by the clinician who sees the patient only in a highly structured, isolated, and somewhat artificial examination or therapy situation. The Functional Communication Profile (M. T. Sarno, 1969) is an example of a standardized rating scale of such communicative features.

The trend toward using rating scales for functional communication, particularly for program evaluation, has been criticized by Sacchett and Marshall (1992). They argue that material that has been "squeezed" out of a patient by role-playing or verbal prompts does not reflect natural conversational abilities, and that ratings of functional communication or QOL, for the sake of justifying speech therapy costs, may lead to forms of intervention that merely promote abilities in the areas of speech subject to rating. Instead, they advocate a case study approach along psycholinguistic lines, arguing that this will "repair" or improve language processing, generalizing to overall language use.

Communication Profiles

A distinction between selective language elicited in the typically structured test setting and the patient's ability to communicate in everyday environments has intuitive value to the speech clinician. The Functional Communication Profile (FCP) was the first standardized attempt to assess the functional usefulness of language ability in the everyday life of the aphasic patient. The Communicative Abilities of Daily Living (CADL; Holland, 1980) was the second psychometric measure to index the degree of disability faced by the patient in attempting to communicate in daily life, but using observations of actual interactions. The Communicative Effectiveness Index (CETI; Lomas et al., 1989) relies mainly on the patient's communicative interaction with spouses or significant others in 16 different situations, and correlates with the improvements shown on the WAB. The Functional Assessment of Communication Skills (FACS; Frattali, Thompson, Holland, Wohl, & Ferketic, 1995) returns to the FCP method by using the observations of both the informant and the speech clinician. Other, infrequently used instruments (e.g., Communicative Competence Evaluation Instrument; Houghton, Pettit, & Towey, 1982) are summarized and re-
viewed by Manochiopinig, Sheard, and Reed (1992). A new British instrument, the Assessment of Communicative Effectiveness in Severe Aphasia (Cunningham, Farrow, Davies, & Lincoln, 1995) uses standard open-ended conversational questions, objects, and pictures. It is fully described in the first publication, but has so far only been examined for interrater and test-retest reliability with 10 aphasic and 10 normal speakers. The rating of functional language and verbal processing skills in daily living has also become a common feature in the many geriatric rating scales (see the review by DeBettignies & Mahurin, 1989), and is available as separate language ratings (Patient Functional Communication Screening Instrument, FCS; Toner, Gurland, & Gasquoine, 1984; Toner, Gurland, & Leung, 1990) or in mental competence batteries (e.g., the Cognitive Competency Battery; Wang, Ennis, & Copland, 1986). However, these omnibus measures for use with the elderly often lack suitable standardization and item range for use with aphasic patients. Le Dorze, Julien, Brassard, Durocher, and Boivin (1994) reported on the development of a scale for use with long-term care patients, including those with aphasia and/or dementia, based on 196 statements about communicative acts in daily living.

The CADL, FACS, and FCP defer obtaining pure, isolated samples of specific language behaviors (as obtained by diagnostic tests) in favor of sampling complex communicative behaviors, such as the ability to communicate on the telephone, handle money, read newspapers and product labels, and ask for, correct, and impart significant information to and from others. As such, the information gauged on these profiles is a unique contribution to the overall assessment of the aphasic patient, providing the clinician with descriptive information about the communicative status of the individual, which can be treated as a second dimension of information not directly obtained from diagnostic testing procedures (M. T. Sarno, 1984a).

COMMUNICATIVE ABILITIES OF DAILY LIVING

The Communicative Abilities of Daily Living (CADL) (Holland, 1980) was designed to measure the functional communicative ability of aphasic patients. Much of the test involves patient performance during simulated, cue-context daily activities, such as dealing with a receptionist, communicating with a doctor, driving, shopping, and making telephone calls. Given its inherent focus on communicative rather than language ability per se, accurate communication whether oral, written, gestural, or by any other modality is acknowledged as significant.

The CADL has 68 items, which are scored as either "correct" (2 points), "adequate" (1 point), or "wrong" (0 points). For example, at one point the examiner asks the patient, "Your first name is _____, right?" (filling in a
fictitious name). If the person’s response includes both a negative response (“No,” headshake, written response, etc.) and his or her correct name, 2 points are given. If the patient simply replies with a negative response without further elaboration, the response is considered adequate but not fully appropriate, and is scored 1 point. If the patient responds affirmatively, perseverates, echoes the question, responds incoherently, or simply does not respond, no points are allotted. Requests for repetition are considered legitimate communicative statements and are not penalized. However, if the patient fails to respond within 5 sec, only partial credit (1 point) is allowed for a correct response. Given the generally slowed psychomotor and information-processing speed that frequently accompanies acquired brain damage, one wonders whether this time restriction is overly strict and does not indeed violate one of the goals of the CADL, which is to score communication success regardless of transmission method. Testing time ranges from 30 to 90 min.

The test situation is deliberately informal, and the examiner is instructed to act in various roles as much as possible (e.g., by changing his or her voice, introducing humor, etc.) to increase the “contextual richness” of the situations created during the test. The total score is the sum of all points scored on the 68 items (maximum score is 136). Holland reported reliability for 20 subjects retested after 1 to 3 weeks by a different examiner as .99, and internal consistency as .97. Concurrent validity was .84 with the BDEA, .93 with the PICA, and .87 with the FCP. Correlations for 23 ratings by staff and family were .67. Criterion validity was established by comparing CADL performance with behavior during a 4-hour observation period. Furthermore, the distribution of scores for different types of aphasia followed the clinical impression of their severity. Global aphasics showed the poorest CADL performance. Wernicke’s aphasics performed more adequately than did the global aphasics, but less well than Broca’s aphasics. The anomic group had near-normal scores. Aphasics living at home consistently had higher scores than those living in an institution.

Normative data are based on a sample of 130 normal adults (fluent English speakers without history of mental disorder or brain damage, or vision or hearing impairment). The differences between occupational levels, institutionalized versus noninstitutionalized, and between males and females were not significant. However, a slight decline with aging (over 65 years) was significant. The performance of 130 aphasic patients is also described in the manual.

CADL performances have been examined in patients with Wernicke’s aphasia, with normal controls, with AD, and in depressive patients (Fromm & Holland, 1989). Aphasics had markedly different profiles compared with AD patients, who showed performances corresponding to AD
severity.Depressed patients also showed lower scores than controls, but tended to show incomplete responses rather than the irrelevant, vague, or rambling responses seen in AD. A group of adult mentally retarded subjects with IQs between 50 and 80 obtained scores in the aphasic range; IQ and CADL score correlated .72 (Holland, 1980). A group of hearing impaired subjects (with hearing aids) showed near-normal scores (Holland, 1980).

The CADL is an excellent supplement to other aphasia examinations because it allows an estimate of the patient's communication ability rather than the accuracy of language. The "staged" quality of some sets of items requires a certain amount of acting ability on the part of the examiner, and may not always be successful with patients who refuse or cannot enter into such simulated interactions. It is not clear from the manual how this affects scores, but it is probably wise to take note of a patient's inability to follow the play-acting, and to make allowance for this in the interpretation of the total score. These interactions, created by specific instructions in the CADL, would also seem to require the patient to show a certain amount of new learning during the test situation which is not required for the FACS or FCP ratings. Apraxic patients may also be problematic (R. Wertz, LaPointe, & Rosenbek, 1984), although the test may serve as a supplementary instrument in such a population. Davis (1993) warns that the CADL is "still a test and does not provide for observing natural interactions" (p. 65).

Italian (Pizzamiglio et al., 1984), Spanish (P. Martin, Manning, Munoz, & Montero, 1990), and Japanese (Watamori et al., 1987) versions are available. A new edition, the CADL-2, will soon be published (Holland, Frattali, & Fromm, 1998). This edition eliminates redundant items and avoids role-playing problems by substitution of hypothetical situations (A. Holland, personal communication).

COMMUNICATIVE EFFECTIVENESS INDEX

The Communicative Effectiveness Index (CETI) (Lomas et al., 1989) uses a 16-item questionnaire. Each question is concerned with an everyday situation and paired with a visual analogue scale (a line marked "Not at all" at one end, and "As able as before the stroke" on the other), which the relative or other caregiver has to mark. For example, one question asks: "Does he intentionally let you know how he is feeling rather than you 'reading' his emotion?" A template is used to convert each mark into a 10-point scale. This measure is designed to use little of the therapist's time. The authors applied the scale to 11 recovering (6 to 10 weeks postonset) and 11 stable (more than 15 months postonset) aphasics. The recovering group showed, with the exception of two patients, the same or more recovery as the WAB; and with the exception of three patients, the CETI showed the
same or less change compared with the WAB for the stable group. Crockford and Lesser (1994) found, however, that the scale has limited potential for planning appropriate intervention. In a follow-up of eight aphasics, retested after 3 months, the CETI was less effective as a measure of stability or change of communicative effectiveness.

The interrater reliability has been reported as .73 for the combined group, and as .94 for the stable group only. Test–retest reliability was .94, and the difference between first and second testing for recovering aphasics was significant. Internal consistency was .90. Concurrent validity was .61 with the WAB Aphasia Quotient on first testing, and .52 on second testing. Correlation with a global rating of aphasia severity by caregivers was .79 on first testing and .62 on second testing. However, correlation with a speech questionnaire was only between .46 and .43 for speaking, and between .47 and .56 for understanding. The authors (Lomas et al., 1989) ascribe this finding to the emphasis on language in the speech questionnaire, whereas the CETI stresses functional communication.

FUNCTIONAL ASSESSMENT OF COMMUNICATION SKILLS FOR ADULTS

The Functional Assessment of Communication Skills for Adults (FACS) (Frattali et al., 1995) uses 43 items rated on a 7-point scale. The ratings are based on the observation of the therapist, but may include observations reported by others. Ratings for communication independence cover Social Communication (e.g., refers to familiar people by name, understands facial expressions); Communication of Basic Needs (recognizes familiar faces, makes needs to eat or to rest known); Reading, Writing, and Number Concepts (makes basic money transactions, fills out short forms); and Daily Planning (tells time, follows a map, keeps scheduled appointments). In addition, each domain is rated on a 5-point scale for qualitative dimensions of communication: Adequacy (understanding of message), Appropriateness (communication is relevant and done under the right circumstances), Promptness (responds without delay and in an efficient manner), and Communication Sharing (burden on the communication partner). A combined communication independence score can be calculated. The ratings for each domain of communication independence and for the qualitative dimensions may be entered into profiles which facilitate plotting of the patient’s progress during therapy. Each rating point is well-defined, as reflected in interrater reliability between .88 and .92 for communication independence scores (total .95) and quality of communication (.72 to .84, total .88) measured during a field test involving 31 examiners and 185 adult patients (Frattali et al., 1995). Internal consistency averaged .82 within domains, and .78 between items and overall score.
The external validity of the overall score for the aphasic group was .73, measured against the WAB Aphasia Quotient, and .78 for the qualitative dimension, against the Functional Independence Measure (State University of New York, 1993). Construct validity, measured with a principal component factor analysis, resulted in one major factor, with minor factors representing the four domain scores. Overall ratings of functional communication by clinicians and family members correlated with FACS scores at a level between .58 and .63.

Normative data are not appropriate for the FACS since a client with normal language would obtain perfect scores on all ratings. However, means for the aphasic and the traumatic brain injury (TBI) group are given and show the sensitivity of the FACS. This is further documented by an account of the discrimination of the scores between groups of patients with different levels of severity of impairment.

In comparison with the CADL, the FACS uses ratings based on general observations rather than on observations under specific conditions of testing.

FUNCTIONAL COMMUNICATION PROFILE

The Functional Communication Profile (FCP) (M. T. Sarno, 1969; M. L. Taylor, 1965) is designed to measure natural language use in everyday communication. The FCP attempts to index the aphasic patient’s ability to use language in common situations, relative to the patient’s estimated premorbid level of ability. “Normal” performance on the profile is defined by the clinician’s skilled estimation of the patient’s previous language ability based on available evidence (e.g., education level, occupation, interviews with family members). The effectiveness of a clinician-based rating scale of this type is directly related to the experience and skill of the user; therefore, the FCP is not recommended for use by testing technicians or clinicians with limited experience, or in settings where few adult aphasics are likely to be seen. Its usefulness may also be limited in situations where little premorbid information is available.

The clinician’s primary role is to create an informal rapport with the patient that allows the clinician to observe the patient’s natural communicative behavior without resorting to formal testing. Forty-five behaviors are rated on a 9-point scale of current ability as a proportion of estimated former ability. The scale ranges from “normal” (100%) to “absent” (0%) ability. Examples of functional abilities include the abilities to indicate yes or no, to read newspaper headlines, and to make change. The 45 behaviors are clustered into five categories: Movement, Speaking, Understanding, Reading, and Miscellaneous (e.g., calculation and writing) abilities. Over-
all cluster scores are obtained by determining the mean rating of the items in a cluster.

Despite the subjective clinical nature of the scoring system, M. T. Sarno (1969) reported interrater reliability coefficients larger than .87 for each of the five FCP categories.

Gains on psychometric testing do not automatically imply improved functional abilities, and, conversely, functional gains may not alter diagnostic classification. The distinction between functional ratings and psychometrically measured language functioning was examined by J. E. Sarno et al. (1971). Measurements of improvement were determined by comparing original and follow-up performances on the NCCEA Visual Naming and Identification by Sentence (i.e., TT) subtests and the FCP Speaking and Understanding subscales. Only a modest relationship was found between the original and follow-up scores on each of the two speech measurements (i.e., NCCEA Visual Naming and FCP Speaking), and no correlation was found between score changes on the two comprehension measures (NCCEA Identification by Sentence and FCP Understanding). On the other hand, M. T. Sarno and Levita (1981) reported some concordance between NCCEA TT performance and FCP Understanding in the examination of global aphasics at 1-year follow-up examinations.

The information obtained from the FCP is not designed to replace a comprehensive examination of the aphasic patient's language abilities and disabilities. Rather, its goal is to provide information about natural communication capacity (M. T. Sarno, 1984a). The information yielded by a properly administered FCP may well translate more easily into a description of the patient's everyday capabilities than the information provided by a standard comprehensive examination. It is not a diagnostic test. For example, M. T. Sarno, Buonaguro, and Levita (1987) found that scores for fluent (Wernicke's) and nonfluent (Broca's) aphasics on the FCP were highly similar after age, education, and time of onset were controlled. When properly used, the FCP may provide information on the functional consequences of the patient's aphasic condition that is not otherwise (except anecdotally) available. Studies examining the accuracy of medical personnel and family members in estimating or predicting the aphasic patient's performance levels (e.g., McClenahan, Johnston, & Densham, 1992) indicate the potential educational value of the results of the FCP and similar measures in counseling the family and in fine-tuning therapist and caregiver communicative interactions with the patient. Repeated FCP administration may provide information on the recovery process of functionally relevant communicative ability (Sands, Sarno, & Shankweiler, 1969; M. T. Sarno, Buonaguro, & Levita, 1985; M. T. Sarno & Levita, 1979).

None of the diagnostic tests discussed later in this chapter provides di-
rect measurements of functional communication. Hence, any of the described communication profiles would make a significant addition to aphasia assessment in association with a comprehensive diagnostic battery.

**Comprehensive Examinations**

Comprehensive examinations of the aphasic's language ability seek a diverse sampling of performances at different levels of task difficulty along all dimensions or functions that the test author deems relevant to language disability. Examples of dimensions common to most of these tests include naming, oral expression, auditory comprehension, repetition, reading ability, and writing ability. Other dimensions vary according to the theoretical orientation of the authors.

For practical purposes, we limit the current review to instruments designed for and/or primarily used in English. Tests developed and available mainly in another language (e.g., Aachen Aphasia Battery; Willmes, Poeck, Weniger, & Huber, 1983) are excluded. We also omit an older, infrequently used instrument (Appraisal of Language Disturbance; Emerick, 1971).

**ARIZONA BATTERY FOR COMMUNICATION DISORDERS OF DEMENTIA**

The Arizona Battery for Communication Disorders of Dementia (ABCD) (Bayles & Tomoeda, 1990) is designed as a comprehensive assessment of language and other communicative functions in patients with dementia, and includes four screening tasks: Speech Discrimination, Visual Perception and Literacy (reading sentences), Visual Field (circling all A's on a page of randomly scattered letters), and Visual Agnosia (naming or describing pictured objects). These tasks are designed to alert the examiner to other disorders that may interfere with communicative functioning. The main test consists of 14 subtests. A number of these subtests reflect traditional items found in most batteries, such as confrontation naming, category (animal) fluency, yes/no comprehension, and repetition. The battery also includes assessment of mental status (i.e., orientation), free-speech description of three pictured objects, a subtest similar to the Wechsler Vocabulary task, figure copying, and object drawing. Importantly, the ABCD differs from more typical aphasia batteries by measuring immediate and delayed recall of a story containing 17 pieces of information.

The response record provides relatively clear scoring guidelines, allowing 1 point for each item or part of an item. The authors suggest converting raw scores into "summary scores" between 1 and 5 to make subtest
scores comparable to each other. The summary scores are based on the performance of 50 healthy subjects of age (mean = 71 years), premorbid IQ, and education similar to that found in dementing patients. A summary score of 5 was achieved by most normal subjects, a score of 3 to 5 by patients with mild dementia, and a score of 2 to 5 by patients with moderate dementia. A construct summary score can be obtained by averaging summary scores across subtests contributing to each of five major areas (Mental Status, Episodic Memory, Linguistic Expression, Linguistic Comprehension, Verbal Visuospatial Construction). An overall performance score can be calculated by averaging across all subtests.

The ABCD has been used in only a few published studies. The manual reports 1-week retest reliability for 20 patients with AD as ranging from .01 (Reading Comprehension, Words) to .86 (Figure Copying), and concordance rates from .65 to .87. Internal consistency is high for all subtests (.63 to .98). The test has been successfully used in the United Kingdom without a change of items (Armstrong, Borthwick, Boyles, & Tomoeda, 1996).

Criterion validity has been demonstrated by highly significant correct classification rates between 50 normal and 50 AD subjects. Correlational validity of individual subtests with several measures of the severity of dementia in 50 AD patients ranged from .59 to .85. The test has also been used in a study of patients with multiple sclerosis (Wallace & Holmes, 1993), and in a comparison of semantic and phonetic (letter) word fluency by Bayles, Salmon, Tomoeda, Jacobs, & Caffrey, 1989).

The ABCD was standardized on 50 healthy elderly subjects with a mean age of 71 and slightly higher than average education and premorbid IQ. These subjects also form the basis for transformations into summary scores. It is not clear whether the test is subject to age, education, or gender effects, or whether a profile of subtest scores may contribute to differential diagnoses.

The current version of the ABCD is obviously still in the research stage and requires further work and independent confirmation. The test provides a wide range of measures similar to the WAB or the NCCEA, but also includes a general measure of mental status, several measures of memory, and diverse screening measures for visual perception, visual fields and/or neglect, auditory discrimination, and visual agnosia. These additional features are suitable for a quick, superficial assessment of the status of mental functions of dementing patients in the context of speech therapy. However, neuropsychologists may wish to use more detailed and better validated tests for these areas. In fact, one could use the MMSE, the BNT, FAS-COWA, and a list-learning task in half the time and obtain information similar to that tested in the ABCD. Although designed for elderly patients with dementia, the ABCD may be suitable for a larger range of pa-
tients, or for differential diagnosis with other dementing conditions or specific aphasia syndromes.

**BOSTON ASSESSMENT OF SEVERE APHASIA**

The purposes of the Boston Assessment of Severe Aphasia (BASA) (Helm-Éstabrooks, Ramsberger, Moyan, & Nicholas, 1989) are to provide a full assessment of language and other communicative functions in severely aphasic patients, and to identify those abilities that might form the initial steps of direct rehabilitation. It consists of 61 items in 15 subtests: Social Greetings and Simple Conversation; Personally Relevant Yes/No Questions; Orientation to Time and Place; Bucco-Facial Praxis; Sustained “Ah” and Singing; Repetition; Limb Praxis; Comprehension of Number Symbols; Object Naming; Action Picture Items; Comprehension of Coin Names; Famous Faces; Emotional Words, Phrases, and Symbols; Visuo-Spatial Items; and Signing One’s Name.

Several items can be administered in a relatively informal manner or as part of a conversation. The test takes approximately 30 to 40 min. Items are scored at several levels of performance, depending on the task demands, from no verbal or gestural response and partially communicative gestural response, to fully communicative verbal response and verbal response with affective quality. For each subtest, exact scoring criteria are provided. Correct responses are summed across five item clusters (auditory comprehension, praxis, oral-gestural expression, reading comprehension, gesture recognition + writing + visuo-spatial tasks). The total of correct responses forms the BASA total score. The total of other responses can also be summed across clusters and for the total test.

Internal consistency is high (between .72 and .89 for the five areas and .94 for the total score, slightly lower in global aphasics). Test-retest reliability after 2 months ranged from .52 to .73 for the area scores, and was .74 for the total score. Interrater reliability ranged from 80 to 100% in two patients. Concurrent validity with the BDAE is modest to adequate. This is probably to be expected because of the low difficulty range of the BASA subtests and the wide range of the BDAE. The correlation for 43 patients between BASA total score and BDAE aphasia severity rating was .67. Cluster total score correlations ranged from .44 to .76, suggesting some independence of each area score. A factor analysis showed expressive, visuo-spatial, and comprehensive language factors.

The BASA samples a wide range of communicative functions at a relatively low level, and lends itself to bedside examination of severely language-impaired patients. The scoring of affect and perseveration, and of partial verbal and gestural responses can provide useful information for the examiner and others, but so far these have not been further investigat-
Townes (1995) states that the inclusion of gestural communication is a unique feature of the BASA, but this also makes extra demands on the examiner's attention during testing.

Norms are presented as preliminary and are based on the performance of 111 patients with severe aphasia, including 47 with global aphasia. To use the norms, raw scores for the five areas are converted into standard scores, which in turn can be used to assign a percentile rank within this aphasic population. Norms for normal healthy speakers are not provided because they would presumably achieve a perfect score. Effects of age, gender, or education are not reported.

The BASA is a useful addition to the other comprehensive batteries in that it provides the opportunity for a broad communication assessment in severely aphasic patients by extending the range of items in the lowest range of performance. So far, research with this relatively new instrument has been limited.

**BOSTON DIAGNOSTIC APHASIA EXAMINATION**

The original Boston Diagnostic Aphasia Examination (BDAE) (Goodglass & Kaplan, 1972) has been revised (Goodglass, Kaplan, & Weintraub, 1983). The primary focus of the BDAE, which remains one of today's most popular aphasia examinations (e.g., Beele, Davies, & Muller, 1984), is the diagnosis of classic anatomically based aphasic syndromes. This diagnostic goal is attained by comprehensive sampling of language components that have previously proven themselves valuable in the identification of aphasic syndromes.

Goodglass and Kaplan stated that the design of their instrument is based on the observation that various components of language function may be selectively damaged by central nervous system (CNS) lesions; this selectivity is an indication of the anatomical neural organization of language and the functional interactions of various parts of the language system. A number of studies have validated this stated purpose (e.g., Meffer & Jeffrey, 1984; Naeser & Hayward, 1978).

The BDAE is divided into five language-related sections: (a) conversational and expository speech, (b) auditory comprehension, (c) oral expression, (d) understanding written language, and (e) writing. Each section contains a variety of subtests. Each subtest attempts to measure the specific function in as purely isolated a manner as possible.

The detailed manner of examining conversational and expository speech remains an important and relatively unique aspect of the BDAE. A "speech characteristics profile" is derived from samples of both free-conversational speech and narrative speech in the description of a line drawing (the "cookie theft" card). The profile indexes verbal prosody (melodic
Assessment of Aphasia

Fluency, articulation, grammatical level, paraphasias, and word-finding difficulties. Repetition and auditory comprehension are also rated, but ratings are derived from subtest performances. Finally, an overall "severity rating scale" can be determined from conversational speech samples, ranging from "no usable speech or auditory comprehension" (a score of 0) to "minimal discernable speech handicaps" (a score of 5). The reliability of the speech characteristic profile was examined for the original BDAE by three judges who rated the tape-recorded speech samples of 99 patients. The lowest correlations were .78 and .79 for word-finding difficulties and paraphasias, respectively; the other dimensions had coefficients of at least .85. Other interrater agreement studies have also shown satisfactory results (Davis, 1993).

Auditory comprehension is tested by word discrimination (pointing to objects, etc., on cards), body-part identification, commands (e.g., "make a fist"), and complex ideational material ("will a stone sink in water?").

Oral expression is gauged by oral agility (mouth movements); verbal agility (rapid repetition of words); automated sequences (e.g., reciting days of the week); recitation, singing, and rhythm (e.g., reciting nursery rhymes); repetition of words of increasing length and difficulty; and repeating phrases and sentences of high and low probability. Oral expression also includes word reading, responsive naming ("what do we tell time with?"); visual confrontational naming, animal naming, and oral sentence reading.

Understanding written language is assessed by symbol and word discrimination, word recognition, comprehension of oral spelling, word-picture matching, and reading sentences and paragraphs. Writing includes a 4-point rating of the mechanics of writing, serial writing (alphabet and numbers), primer-level dictation (letters, numbers, and words), spelling to dictation, written confrontation naming, narrative writing ("cookie-theft" card), and sentences to dictation.

In addition, several optional supplementary language tests explore psycholinguistic aspects of language, for example, comprehension of prepositions of location, passive subject–object discrimination, possessive relationships, expression of indicative, interrogative, and conditional verb and tense usage; specific repetition tasks for conduction aphasia; naming by touch; instructions for exploring minor hand agraphia; and the BNT as an extended naming test. Supplementary nonlanguage tests (spatial–quantitative tests, formerly called the Boston Parietal Lobe Battery) are included as a separate chapter as well as on the profile summary. These include constructional apraxia (drawing to command), finger agnosia, right–left orientation, acalculia, and ideational apraxia, as well as commands to test for bucco-facial apraxia.
Internal consistency ranges from .98 for visual confrontation naming to .68 for body-part identification. Test–retest reliability is not reported.

Knowledge of the "Boston school" approach to aphasia classification is necessary to interpret the BDAE (e.g., Benson, 1979a). The speech characteristics profile and the severity ratings are central to diagnostic decision making with the BDAE; particularly important is the fluency–nonfluency dimension. More detailed diagnoses may incorporate corroborative information from the profile sheet delineating subtest performances.

The original BDAE was standardized on a sample of 207 aphasic patients with relatively distinct cerebrovascular accidents (CVAs) and isolated, well-defined symptoms. Although standardization on such a large sample of patients is psychometrically useful, the clinician whose referrals do not reflect this select sample (i.e., referrals with a different bias in symptomatology and severity) may not be able to reference her or his patients directly to the BDAE sample. The revised BDAE was standardized on a new sample of 242 patients. As in the first sample, selective aphasias produced by CVAs predominate. However, the authors report that this sample included more patients with larger lesions and more severe aphasic symptoms.

The BDAE was also standardized on 147 healthy normal adult subjects (Borod, Goodglass, & Kaplan, 1980) to provide cutoff scores at 2 SD below the mean and include age and education corrections. These means, ranges, and suggested cutoff scores are included in the manual.

Emery (1986) found only a minimal, insignificant decline of scores on all subtests when comparing 20 healthy adults aged 30 to 42 years with a similar group aged 75 to 93 years. Whitworth and Larson (1989) found no significant effects for gender and education in their sample. Heaton, Grant, and Matthews (1991) presented norms for the comprehension of complex material of the BDAE in scaled scores, corrected for gender, education, and age, that are based on 553 normal subjects. However, these norms should be used with caution, because the number of subjects in some of the cells is quite small. Rosselli, Florez, and Castro (1990) presented norms based on 180 normal Spanish speakers from Colombia, broken down by educational level and three age ranges (16 to 30, 31 to 50, 51 to 65 years). These data show a significant effect of educational level for most tasks, and an age effect for some of the tasks. Norms for the parietal lobe battery in older subjects (ages 40 to 89) were presented by Farver and Farver (1982).

Borod, Carper, Goodglass, and Naeser (1984) applied the battery to 163 right-handed aphasics and found four factors: construction, visual schemata, verbal components of the Gerstmann syndrome, and visual finger recognition. Impairment was strongest in patients with lesions in both left parietal and frontal areas. The spatial–quantitative tests, together with
the WAIS, were applied to right- and left-handed aphasics: left-handed aphasics were significantly poorer on both, especially on tasks involving visuospatial construction, suggesting that in left-handers, the left hemisphere is typically dominant for tasks usually thought of as right-hemisphere specific (Borod, Carper, Naeser, & Goodglass, 1985).

Construct validity has been examined by reviewing the intercorrelation matrix of the 43 language and 23 nonlanguage measures of the BDAE and by factor analysis. In an earlier analysis (Goodglass & Kaplan, 1972), a strong general language factor emerged, as expected, with other factors covering the spatial–quantitative–somatognostic domain, as well as articulation/grammatical fluency, auditory comprehension, and paraphasia domains. A second factor analysis (Goodglass & Kaplan, 1983), omitting ratings and nonlanguage tests, resulted in five factors (comprehension/reading/naming, recitation/repetition, writing, oral agility/singing/rhythm, auditory comprehension). When rating scales were included, three additional factors emerged (fluency, reading, paraphasia). The addition of the spatial–quantitative tests resulted in a 10-factor solution, including a strong spatial–quantitative factor, a finger identification factor, and a factor labeled “freedom from paraphasia.” Discriminant validity between cases of Broca’s, Wernicke’s, conduction, and anomic aphasia was optimal when the following tests were entered into the equation: body-part identification, repetition of high-probability sentences, paraphasia rating, word-finding rating, phrase-length rating, and verbal paraphasias.

Divenyi and Robinson (1989) reported correlations of .86 and .93 of the auditory comprehension measured in the BDAE with the TT and with the respective part of the PICA. However, the BDAE auditory comprehension subtest was not an adequate predictor of auditory paragraph comprehension in independent standardized material (Brookshire & Nicholas, 1984); a second study (L. E. Nicholas, MacLennan, & Brookshire, 1986) showed that both aphasic and healthy subjects were able to answer a similar number of questions about a paragraph without having actually read the passage, suggesting a high passage dependency of this test. This dependency applied not only to the BDAE, but also to similar tasks in the MTDDA and the WAB. Dyadic interaction measures also did not correlate well with the BDAE (Behrmann & Penn, 1984). Decision rules for the “diagnosis” of the individual subtypes are not always clearly defined, although Reinvang and Graves (1975) attempted such clarification. Crary, Wertz, and Deal (1992) tried to isolate subtypes of aphasia empirically by means of a Q-type factor analysis for the BDAE and the closely related WAB; the resulting seven patient clusters (labeled Broca, anomic, global, Wernicke, conduction, and two unclassified clusters) agreed only poorly (in 38% of 47 patients) with the classification obtained using the classification rules of the test it-
self; the results were even worse for the WAB. The study, aside from its limited subject population for factor analytic studies and the use of a somewhat dated cluster-analysis technique, suggests that BDAE classification rules are based on clinical rather than construct validity. Similarly, Naeser and Hayward (1978) and Reinvang (1985) pointed out that scale profiles can aid in the classification but do not firmly classify patients into subtypes of aphasia. The test authors acknowledge that 30 to 80% of aphasic patients are not classifiable; this is also consistent with clinical experience that a majority of aphasic patients show mixed rather than pure symptomatology.

The BDAE predicted progress in therapy (Helm-Estabrooks & Ramsberger, 1986; Davidoff & Katz, 1985). In aphasics, word reading (Selnes, Niccum, Knopman, & Rubens, 1984) and confrontation naming (Knopman et al., 1984) showed striking improvement 6 months postinsult. Specifically, Marshall and Neuberger (1994) found that a measured pretreatment effort in self-correction (but not success of self-correction) and good auditory comprehension were related to improvement during treatment, as measured by the BDAE and the PICA. Narrative response to the Cookie-Theft card has been examined as a method of gleaning additional information about patients' self-monitoring and self-correction of errors. Nonaphasic and nondemented healthy elderly self-correct 72 to 92% of their errors in the description of this card (McNamara, Obler, Au, Durso, & Albert, 1992).

The test is lengthy (90 to 120 min) and probably more useful for assessments in detailed studies of aphasia and aphasia rehabilitation than as a routine language test included in a general assessment. The somewhat unwieldy number of tasks is clearly the result of trying to fit classical clinical-neurological testing into a psychometric format. The test includes useful directions for observing and recording many specific types of errors (e.g., paraphasias) found in aphasia, demonstrating the Boston process approach. However, even if the full-length test is not used, a number of sub-tests can be useful additions to clinical assessment depending on the presenting symptoms of the patient.

A Spanish (Goodglass & Kaplan, 1986), French (Mazaux & Orgogozo, 1985), and Hindu version (Kacker, Pandit, & Dua, 1991) are available. Computerized scoring and interpretation software is also available (Code, Heer, & Schofield, 1990).

MINNESOTA TEST FOR DIFFERENTIAL DIAGNOSIS OF APHASIA

The Minnesota Test for Differential Diagnosis of Aphasia (MTDDA) (Schuell, 1995, 1973) was ranked among the more popular batteries in 1984 (Beele et al., 1984), although few studies with the test have been published in recent years. It is a comprehensive examination designed to observe the
level at which language performance is impaired in each of the principal language modalities at different levels of task difficulty. To Schuell, the goal of a careful and comprehensive description of impairment in the aphasic patient is to provide a guide for effective therapeutic intervention.

The current version of the MTDDA is the result of numerous systematic revisions of the original experimental version of the late 1940s. The author employed empirical factor analytic techniques (Schuell, Jenkins, & Carroll, 1962) as well as clinical experience to construct and revise the test. The construction of the MTDDA reflects Schuell's theoretical view of aphasia as a unitary reduction of language that crosses all language modalities, which may or may not be complicated by perceptual or sensorimotor involvement, by various forms of dysarthria, or by other sequels of brain damage (Schuell, 1974b; Schuell & Jenkins, 1959; Schuell et al., 1964). The MTDDA is composed of five sections: auditory disturbances (represented by 9 subtests), visual and reading disturbances (9 subtests), speech and language disturbances (15 subtests), visuomotor and writing disturbances (10 subtests), and numerical relations and arithmetic processes (4 subtests). Within each section, the subtest order is generally arranged from the least difficult to the most difficult. Each section may be started at an estimated level of difficulty corresponding to the patient's ability, and then continued to the point where the patient fails 90% or more of the items. Both the test manual (Schuell, 1965) and the companion monograph (Schuell, 1973) describe supplementary tests that should be considered, as well as the factor and intercorrelation structure for the tests.

Differential diagnosis with the MTDDA identified five aphasia syndromes: simple aphasia, aphasia with visual involvement, aphasia with sensorimotor involvement, aphasia with scattered findings compatible with generalized brain damage, and an irreversible aphasia syndrome (Schuell, 1974a). Schuell (1966, 1973) also added two additional "minor syndromes": mild aphasia with persistent dysfluency (dysarthria), and aphasia with intermittent auditory imperception. However, as Zubrick and Smith (1979) pointed out, the MTDDA was not designed to deal with broader issues of aphasic differential diagnosis (e.g., distinguishing aphasia from memory loss, dementia, severe hearing loss, and confusional state). The test has been successfully used to measure language recovery after stroke and head trauma, and to show that language recovery is relatively independent from intelligence (Bayley & Powell, 1981; David & Skilbeck, 1984). Armstrong and Walker (1994) tested older adults with an MTDDA short version to examine any gender differences, but no differences were found.

The length of the MTDDA presents a problem for the user of the test. Short forms (Schuell, 1957) and "very short" forms (Powell, Bayley, &
Clark, 1980) have been created, reducing the 43 subtests to as few as four. Schuell herself was not impressed with the role of short examinations in the diagnosis of aphasic disorders. The large number of subtests includes many functions that exceed what some authors would consider the assessment of speech and language functions, and extend into material that has been a traditional component of many intelligence tests. Schuell's factor analysis may, on closer inspection, seem to reflect a major first "general" factor that is closely related to the g obtained in factor analyses of intelligence tests. Schuell and Jenkins (1959), however, considered this factor a general language factor, supporting their assumption about the unitary nature of language.

In summary, the MTDDA is an extensive examination. At the time, the test represented a major breakthrough in the development of comprehensive aphasia test instruments that met the requirements of both standardization and objectivity. Great care was taken in its construction, which used both clinical expertise and empirical technique. Potential users of the MTDDA should consider whether its length will be prohibitive in clinical settings. The user should also examine the theoretical bases of the MTDDA relative to the user's own conception of the nature of aphasic deficits, and balance its breadth against practical needs.

MULTILINGUAL APHASIA EXAMINATION

The benefits of having equivalent versions of a single aphasia examination for several language communities has been well stated by Benton (1967, 1969). The Multilingual Aphasia Examination (MAE) has been developed through the efforts of Benton and his collaborators to meet the requirements of a multilingual examination (Benton, Hamsher, Rey, & Sivan, 1994). The MAE is available in English (Benton et al., 1994) and Spanish (MAE-S; Rey & Benton, 1991). Chinese, French, German, Italian, and Portuguese versions of this test are being prepared. The different language versions of the MAE are functionally equivalent in content rather than simple translations. For example, COWA uses letters that have corresponding levels of difficulty in each language rather than identical letters. Hence, performance of the task in each language is functionally equivalent.

The MAE, a shortened and highly modified relative of the NCCEA, consists of seven subtests: Visual Naming, Sentence Repetition, COWA, Spelling, a version of the TT, Aural Comprehension of Words and Phrases, and Reading Comprehension of Words and Phrases. Two MAE rating scales are included. The first is a scale of speech articulation based on verbal performance throughout the test session. Ratings range from 0 ("speechless or usually unintelligible speech") to 8 ("normal speech"). The second scale encodes writing praxis, scored when possible by performance
on tasks of writing to dictation (from the MAE Spelling subtest); scores range from 0 ("illegible scrawl") to 8 ("good penmanship").

A practical and distinctive feature of the MAE is that alternate versions of Sentence Repetition, COWA, Spelling, and TT are available for repeat assessment of the patient. Hermann and Wyler (1988) provided a research example of the utility of MAE alternate forms in an examination of language behavior before and after temporal lobectomy in epileptic patients. Visual Naming requires naming of line drawings (whole objects and object details), which is more difficult than the naming of actual objects. COWA presents three letters of progressing association difficulty. The Spelling subtest permits oral spelling, writing to dictation, or block-letter spelling. The TT is composed of 22 commands at two levels of complexity, and is scored pass-fail. Aural and Reading Comprehension are administered in a multiple-choice format. Scoring adjustments for age and educational level are provided in the test manual.

The MAE manual provides standardized test instructions and normative information from a sample of 360 normal Iowa adults (aged 16 to 69 years) without a history of or evidence of neurological disability. A second validation sample of 50 aphasic patients is included, which may be used to discern aphasic subtypes. Normative information for 229 children (aged 6 to 12 years), based on a study by Schum, Sivan, and Benton (1989), is included in the manual. The MAE-S manual includes normative information from a sample of 234 normal Spanish-speaking adults (aged 18 to 70 years) from Texas and Puerto Rico. The Mayo Older Americans Normative Studies (MOANS; Ivnik, Malec, Smith, Tangelos, & Petersen, 1996) reported results with a large sample of normal adults between 55 and 97 years of age for two MAE subtests (COWA and TT). Because the Naming subtest may be especially sensitive to cultural experience, separate normative data have been obtained for urban inner-city Black residents (Roberts & Hamsher, 1984). The latest edition of the manual reports unpublished data by R. L. Schum and A. B. Sivan about the MAE performance of 60 healthy control subjects, aged 70 to 89 years. Stable performances were noted for subjects in their 70s, but relative performance decrements were found for those in their 80s, particularly on the Sentence Repetition and TT subtests. These studies were conducted with well-educated adults; the authors speculate that subjects with less education may show a steeper decline with age. Rather than looking for poor performance in depressed elderly, La Rue, Swan, and Carmelli (1995) emphasized good COWA performance in elderly adults scoring high on a measure of "zestfulness," suggesting that this is an example where COWA levels may show better than expected performance.

A study of discriminative validity between 115 normal and 48 aphasic
subjects with six of the MAE tests found that with the suggested cutoff scores, between 2.6 and 7.0% of normals, and between 14.4 and 64.6% of aphasics were misclassified by individual subtests. With failure on one subtest as a cutoff, 15% of controls and no aphasics were misclassified; with failure on two or more subtests, the misclassification rates were 3 and 4%, respectively (Jones & Benton, 1995). The TT proved to be the most divergent between the two groups; the two comprehension subtests were the least discriminative. Patients with left temporal lobe epilepsy also showed significant impairment on the MAE compared with epileptics with right hemisphere impairment (Hermann, Seidenberg, Haltiner, & Wyler, 1992; Hermann & Wyler, 1988). Concurrent validity of the Visual Naming subtest, which has 30 items, and the 60-item BNT (.86) was reported by Axelrod et al. (1994) in a diagnostically mixed sample of 100 adult patients with neurological or psychiatric histories.

Research shows that failure on the MAE Token Test is a sensitive indicator of the presence of acute confusional states (delirium) in nonaphasic medical inpatients (G. P. Lee & Hamsher, 1988). Levin and colleagues (Levin, Grossman, & Kelly, 1976; Levin, Grossman, Sarwar, & Meyers, 1981) have used the MAE to examine the linguistic performance of patients with closed head injuries. They documented a high frequency of naming errors, defective associative word-finding, and TT, and revealed the correlation of these with the severity of brain injury. A factor analysis of 16 aphasia battery subtests (including subtests of the MAE, NCCEA, and WAB) given to healthy Taiwanese volunteers (Hua, Chang, & Chen, 1997) suggested a major factor of verbal comprehension (including TT, Sentence Repetition, Digit Repetition, Visual Naming, Reading, and Aural Comprehension). A second factor was labeled effortful writing, and a third factor involved mainly verbal expression and word production.

In summary, it is hoped that the successful deployment of the MAE in a number of language communities will facilitate direct cross-community comparisons of case and sample data. Regardless of this research goal, clinical use of the English-language MAE suggests that it is an effective instrument which requires a relatively brief (usually under 45 min) administration time. In addition, the use in general clinical practice of individual subtests (viz., Visual Naming, TT, COWA) can serve as a good exploratory examination as to the presence of language deficits.

NEUROSENSORY CENTER COMPREHENSIVE EXAMINATION FOR APHASIA

The Neurosensory Center Comprehensive Examination for Aphasia (NCCEA) (Spreen & Benton, 1977; Spreen & Strauss, 1991) was designed to provide a comprehensive assessment of language comprehension, lan-
guage production, reading, and writing. Other stated goals of the NCCEA are to provide subtests that are sufficiently complex so that the clinician can obtain a relatively exact measure of performance level; to standardize and score performances such that necessary corrections for age, sex, and education can be made; to include nonlinguistic subtests to ensure valid interpretation of performance deficits on language tests as either linguistic in nature or due to other dysfunction; and to include specific subtests that could be used in aphasiology research (Benton, 1967).

The NCCEA is composed of 20 subtests that focus on the language function stated above and four “control” subtests of visual and tactile functioning. The test is designed to yield a description of the patient’s profile of abilities and disabilities. NCCEA subtests include stimulus presentation in the visual, auditory, or tactile modality. The subtests measure visual object naming, description of object use, tactile object naming for each hand, sentence repetition, sentence construction, object identification by name, oral reading of names and sentences, oral reading of names and sentences for meaning, writing of object names, writing to dictation, copying sentences, and articulation. The sequential order of subtests provides a meaningful grouping into tests of name finding, immediate verbal memory, verbal production and fluency, receptive ability, reading, writing, and articulation. Eight of the 20 language subtests and three out of the four control subtests require the use of four sets of eight common objects displayed on trays. These objects are presented in order of difficulty, from least to most difficult. The four sets are matched for item difficulty and are equivalent in mean and distribution of difficulty for aphasic patients (mean percent correct is approximately 63% for all trays) and young children (mean acquisition age of names approximately 5:8 years); they are rotated throughout the battery.

Several subtests provide a set of items for initial testing, as well as a second set of items to be used only if errors occurred on the first set. This feature tends to shorten test administration in areas in which a patient has no difficulties. The second set of items provides more detailed quantitative information on problem areas. Isolated errors due to poor attention or other irrelevant causes will be reduced in importance if the second set of items is then passed correctly. The complete test takes between 45 and 120 min to administer.

The range of item difficulty is limited. In an attempt to avoid highly specialized or low-frequency vocabulary, the authors used only common objects for their object naming, identification, and similar tasks. As a result, the test has a rather low ceiling on some of the subtests, with the effect that very mild aphasic symptoms in highly educated patients may be missed. Other subtests, however, are “open ceiling” tests for which this limitation does not apply.
Scores on the NCCEA are determined by response correctness. Incorrect responses and mispronounced correct responses are recorded verbatim to yield qualitative performance information. An individual's performance on the NCCEA, when corrected for the effect of age and educational level, can be converted into percentile scores to yield relative levels of performance on each subtest and can be ranked on three different profile sheets, comparing the patient's performance to samples of normal adults (Profile A), aphasic patients (Profile B), and nonaphasic brain-damaged patients (Profile C). The aphasic and nonaphasic brain-damaged samples consist of consecutive referrals for neuropsychological evaluation in acute-care hospital settings.

Because the NCCEA was primarily designed to assess patients with aphasia or aphasia-type complaints, patients without language problems and normal controls tend to obtain ceiling scores. Therefore, the test cannot be used to measure language ability in normal adults, although the language development in children has been successfully measured with most subtests up to a ceiling age from 8 to 13 years (Gaddes & Crockett, 1975).

An empirical study with 353 children, aged 5:5 to 13:5 (Crockett, 1974), found that seven factors described the content of the NCCEA in that population: reading/writing, verbal memory, name finding, auditory comprehension, syntactic fluency, reversal of digits, and repeating digits.

One-year retest reliability in older adults for selected subtests has been reported as satisfactory (Word Fluency, .70; Visual Naming, .82; TT, .50; Snow & Tierney, 1988).

Construct validity was examined in two studies by Crockett (1976, 1977). The first examined the discrimination of groups of aphasic patients based on ratings of verbal productions and divided on the basis of the Howes/Geschwind two-type, and the Weisenberg/McBride three-type typologies. Neither of the two models showed significant multivariate differences. The second study showed significant multivariate differences on the NCCEA between four types of aphasia empirically derived from ratings of verbal production by hierarchical grouping analysis. Two of the four types appeared to be similar to Howes’ two types, a third appeared to reflect Schuell’s single dimension of language disorder, and a fourth seemed to be characterized primarily by memory impairment. Concurrent validity with the WAB was demonstrated by Kertesz (1979). Concurrent validity for changes in language functioning during therapy was reported by Kenin and Swisher (1972).

Predictive validity was established in a study by Lawriw (1976), who also presented a successful cross-validation between patient groups from Iowa City, New York City, and Victoria, British Columbia. Kenin and Swisher (1972) and Ludlow (1977) investigated patterns of recovery from aphasia; improvement was best reflected in writing from copy and in tests
of comprehension, whereas expressive performance showed the least improvement. Single-word reception or production was more readily recovered than that of longer verbal units. The authors mentioned that reading, writing, and oral production items were not sufficiently difficult for patients at an advanced stage of recovery. M. T. Sarno (1984b) described significant differences between aphasic, dysarthric/subclinical, and subclinical aphasic patients on Visual Object Naming, Sentence Repetition, Word Fluency, and the TT. Patients with AD scored significantly lower in the areas of verbal expression, auditory comprehension, repetition, reading, and writing compared with age-matched, nonneurological controls (Murdoch, Chenery, Wilks, & Boyle, 1987). M. T. Sarno and Levita (1981) described recovery from global aphasia during the first 3 months after stroke.

The 1977 NCCEA normative data for an aphasic reference group are based on 206 unselected referrals to hospital and clinic services in Iowa City, New York City, and Victoria. Although the concept of “averaging” across aphasic patients, also used in the BDAE, disregards the different types of aphasia, this procedure allows the profiling of individual patients against that reference group; that is, an individual patient’s subtype will stand out more clearly. However, when patients from another referral source (e.g., patients in rehabilitation or patients with residual aphasia) are seen, the reference group may no longer be appropriate.

Normative data for most subtests remain stable through the age span up to age 64. Results of studies with a geriatric population on some of the tests (Montgomery, 1982) showed only a minor decline of 1 or 2 points, which has been incorporated into the age and education correction rules. A study by Tuokko and Woodward (1996) of elderly subjects in Vancouver, British Columbia, for example, showed a mean for Visual Naming of 16 (ceiling score) for subjects below 60 years of age to be maintained up to age 79, and drop to a mean of 15.44 for subjects 80 years and older. Similarly, Description of Use showed a ceiling score of 16 up to the age of 74; the mean for subjects 75 years and older was 15.78. Tactile Naming (right hand) showed a mean of 15.66 for subjects up to the age of 79, and 14.89 for subjects 80 years and older. Tactile Naming (left hand) showed a mean of 16 for subjects under 60, and means between 15.14 and 15.57 for subjects between the ages of 60 and 79 years, but a mean of 12.44 for subjects 80 years and older.

Normative data for children between the ages of 6 and 13 are presented. The means were merged from studies by Gaddes and Crockett (1975) and Hamsher (1980), as the differences between the two sources (Victoria and Milwaukee) were minimal.

The test has also been adapted into Italian, Japanese, and Spanish. Writing from Dictation and Writing from Copy were used in a Chinese study (Hua et al., 1997).
In summary, the NCCEA provides a comprehensive assessment of language functions for aphasic patients without the use of a specific language model and without applying a specific approach to delineate diagnostic types of aphasia. Psychometric development of the test has been slow, and few new studies from the last decade are available. The development of three different profile sheets for score evaluation is a distinct asset. On the other hand, the low ceiling of some of the subtests suggests that some aspects of language functioning in mildly or borderline aphasic patients cannot be adequately measured.

PORCH INDEX OF COMMUNICATION ABILITY

The Porch Index of Communication Ability (PICA) (Porch, 1967, 1973, 1981) is designed to assess verbal, gestural, and graphic responsiveness subsequent to brain damage. Unlike classificatory instruments, the PICA is designed to categorize the nature of the aphasic's ability to respond, modality of response, and quality of response to task demands. A prime use of PICA has been in assessing patient performance on multiple occasions postonset to determine recovery trends. The PICA has 18 subtests: four verbal, eight gestural, and six graphic. A high degree of homogeneity among subtests is established through the repeated use of 10 common, everyday objects of equal difficulty (e.g., a key, a cigarette) for a majority of the subtests. This allows examination of fluctuations of performance over time (but may also introduce a practice effect). Subtests were created to conform to a model of language functioning involving several possible input modalities and outgoing responses in several possible output modalities. Five modalities are assessed: auditory comprehension, visual comprehension, written expression, verbal expression, and pantomime. The test takes 1/2 to 2 hr to administer.

A persistent problem in recording responses involves assessing and quantifying the given response as one of a wide variety of possible responses. As a compromise between two possible extremes (i.e., longhand notation of response characteristics and simple pass-fail dichotomies), the PICA uses a 16-point multidimensional scoring system, an attempt by the author to integrate the strengths of the two approaches while minimizing their weaknesses.

A given response to a PICA test item is evaluated along five dimensions: accuracy, responsiveness, completeness, promptness, and efficiency. A scoring system that considers all possible permutations of these five dimensions is, for all practical purposes, impossible and meaningless. Hence, 16 categories have been identified that represent various relevant combinations of the five dimensions, resulting in a 16-point ranked scale from "no response" (1) to "complex response" (16). For example, any at-
tempt by the patient to perform on the task is scored at least a 6; all accurate responses are scored at least an 8. Additional points are given for a correct response after repeated instructions, for self-corrected responses, responsive ease, completeness, promptness, and efficiency. Porch (1967) reported the viability of the rank ordering of the 16 categories by pointing out the high agreement between PICA category ordering and ranking of categories by 12 speech pathologists. The individual item scores (180 possible) are transformed into an overall performance score, several modality scores, and individual subtest scores. The overall performance score is considered as the best single index of the patient's general communicative ability. Modality scores yield information on the relative capacity for verbal, gestural, and graphic communication. Use of mean values require that statistical and conceptual assumptions of equal intervals between category levels be met. Whether this assumption is legitimate for the PICA has been the subject of debate (e.g., Lincoln & Pickersgill, 1981; A. D. Martin, 1977; McNeil, 1979). Although single-item responses can be categorized on the 16-point scale (i.e., a score of 12 represents an incomplete response), mean values cannot be so categorized. Hence, mean scores cannot categorize the method by which the patient generally communicates, but can only represent a performance level relative to other normative values or to the patient's derived scores.

Data indicating that the PICA shows high interrater reliability as well as high test-retest reliability are provided in the manual. Holland (1980) reported concurrent validity of the PICA as .93 with the CADL, .86 with the FCP, and .88 with the BDAE. Lendrem and Lincoln (1985) found that the PICA given at 4 weeks postonset successfully predicted spontaneous recovery at 6 months in 32 male stroke victims between 48 and 80 years of age. Age, but not type of aphasia, was related to rate of recovery in 87 stroke victims between the ages of 38 and 92 years (Lendrem & McGuirk, 1988). However, Lincoln and McGuirk (1986), examining 124 patients between the ages of 38 and 92, 4 weeks and 34 weeks after stroke, noted that groups with and without treatment did not reach the level of recovery predicted by the PICA. They concluded that such predictions are not accurate enough for clinical practice.

Percentile data for all principal transformed scores from normative samples of 357 left hemisphere damaged patients, 96 right hemisphere, and 100 bilaterally damaged patients are presented in the manual (Porch, 1981). In addition to providing relative information for these transformed scores, percentiles are also used to determine a given patient's "aphasia recovery curve," by factoring in the overall test percentile, the mean percentile of the nine highest scored subtests, and the mean percentile of the nine lowest scored subtests. Predictions on the scope of recovery can be attempted
from this curve (Porch, Collins, Wertz, & Friden, 1980). Studying 110 aphasic patients with the PICA, Marshall and Tompkins (1983) found that age, health, etiology, and time since onset were good predictors of improvement during the course of treatment. Training patients in functional communication efficacy, on the other hand, did not show improvements on the PICA, although improvement was found on the CADI (Aten, Caligiure, & Holland, 1982).

DiSimoni, Keith, Holt, and Darley (1975) found a high degree of redundancy among PICA subtests and concluded that a shortened form of the test may be more useful. DiSimoni, Keith, and Darley (1980) described two short versions of the PICA that require only one-third of the time used for the full-length test. A preliminary Portuguese (Brazilian) version is available (Gunther, 1981).

In summary, the PICA is a well-developed and standardized test instrument that has been extensively used in rehabilitation settings to track recovery, although we found few research studies with the test in recent years. The multidimensional scoring system has become the most criticized aspect of the test, only partly remedied in the second edition. Two other shortcomings of the PICA include the paucity of sampling auditory comprehension and the misleading labeling of several subtests as gestural when they entail other specific behaviors. McNeil (1979) suggested that such criticism should not turn clinicians away from the PICA as a test instrument, but rather should make them more cautious interpreters of PICA results.

WESTERN APHASIA BATTERY

The Western Aphasia Battery (WAB) (Kertesz, 1979, 1982) is a close relative of the BDAE and shares with it the diagnostic goal of classifying aphasia subtypes and rating the severity of the aphasic impairment. The examination comprises four language and 3 performance domains. Syndrome classification is determined by the pattern of performance on the four language subtests, which assess spontaneous speech, comprehension, repetition, and naming. Weighted performance on these language subtests yields an overall measure of severity of aphasia, the Aphasia Quotient (AQ). Stepwise regression analysis has shown that, of the AQ constituents, the Information Content rating is most highly correlated with the AQ (Crary & Rothi, 1989). The three performance areas—reading and writing, praxis, construction, and the Raven's Colored Progressive Matrices—yield a second summary measure: the Performance Quotient (PQ). Finally, the AQ and the PQ are summed to form a Cortical Quotient (CQ). Criteria for the classification of eight classic aphasic syndromes are described based on the language subtest performances of 375 aphasic patients with various eti-
ologies (mostly CVA) and 162 normal individuals. Classification is forced into one of these syndromes; unclassifiable aphasia subtypes or mixed presentations are not addressed. In this respect, WAB classification is more rigid than in any other diagnostic battery. This rigidity may have relatively more appeal to researchers than clinicians.

Spontaneous speech is assessed both in response to questioning and in the patient's description of a line drawing, similar to the BDAE. Speech is rated on two 10-point scales: information content and fluency (the fluency scale incorporates both grammatical competence and the presence of paraphasias). Comprehension is assessed by yes/no questions that may be answered in either verbal or nonverbal fashion, by word recognition, and by performance to sequential commands. Repetition has 15 items that are scored as correct, in phonemic error (partial credit), or as an error. Naming is composed of object naming (without cuing or, if necessary, with tactile and/or phonemic cuing), word fluency (animals), sentence completion, and responsive speech. Test items were selected to provide a wide enough range of difficulty for assessing all levels of severity.

An uncommon feature of this test's structure is the dissociation from language performance of reading and writing abilities, which, along with nonverbal measures, form part of the PQ. Shewan (1986) reunited the spoken language section (i.e., the AQ tests) with reading and writing as part of a scale called the Language Quotient (LQ), and provided a detailed account of reliability and validity for this addition to the original WAB format. The LQ is weighted so that 60% reflects spoken language performance, and 40% reflects written language performance. Shewan's report emphasized the relation between the LQ measure and the severity of the aphasic disorder. Crary and Rothi (1989) demonstrated internal consistency for the 10 language subtests with protocols of 100 aphasic patients: All subtests correlated highly with the AQ.

WAB standardization information and reliability, and validity data were provided by Kertesz and Poole (1974) and then updated by Kertesz (1979) and Shewan and Kertesz (1980). The WAB clearly meets standard rules of test construction, although it ranked last in a review of nine aphasia tests by Skenes and McCauley (1985). The WAB manifests good internal consistency and high interrater and intrarater reliabilities. High test–retest reliability has been reported for a sample of 38 chronic aphasic patients. Successful criterion validity has been described by the author. Aphasics were differentiated from non-brain-damaged adults in their WAB performance; the AQ distinguished aphasics from non-brain-damaged controls. Construct validity was assessed in a sample of 15 patients who were examined with both the WAB and the NCCEA; there were high intercorrelations between corresponding subtests ranging from .82 for spontaneous speech
subtests to .95 for comprehension subtests. The WAB line drawing, used to elicit spontaneous speech, generated more enumerations, produced at a slower rate, than the pictures in the BDAE or the MTDDA (Correia, Brookshire, & Nicholas, 1990). McClenahan, Johnston, and Densham (1992) compared the accuracy of estimation of comprehension problems, using the WAB and the FCP, with estimates made by doctors, nurses, and relatives. The comprehension sections of the two tests usually overestimated the patient's ability compared with the judged estimates, but was more accurate for patients with mild problems. Interestingly, length of relationship or educational background did not affect the correctness of judgment, but high or low level of confidence in the judgment by the individual making it did have an influence. One study examined the validity of cutoff scores on the CQ (Fromm, Greenhouse, Holland, & Swindell, 1986; Fromm & Holland, 1989); a cutoff score of 90 points showed good sensitivity and specificity. The authors state that the language subtests can be administered in approximately 1½ hr, but the full WAB might require at least two 2-hr sessions to complete. The test manual is far less detailed than Kertesz' book (1979), which remains a good, detailed introduction to the WAB.

The WAB has established itself as a useful classificatory research instrument, helped by its inclusive objective classification rules and its summary measures. The main subtypes obtained with the WAB are global, Broca's, Wernicke's, conduction, and anomic aphasias (Shewan & Kertesz, 1980). Studies with the test include novel cluster analytic taxonomies of aphasic syndromes of different etiologies over time (Kertesz & Phipps, 1980), and the evolution of aphasic syndromes during the course of recovery (Kertesz, 1981) and during therapy (R. Lesser, Bryan, Anderson, & Hilton, 1986; Shewan & Kertesz, 1984). Crary et al. (1992) used cluster analysis with a small group of 47 patients and found that cluster membership corresponded only poorly (30% of cases) to the classification suggested by the WAB classification rules. Other research investigated the role of activation of the nondominant hemisphere during word repetition (as shown in PET studies; Ohyama et al., 1996), the relation between aphasia and nonverbal intelligence (Kertesz & McCabe, 1975), the relation between language and praxis (Gonzales-Rothi & Heilman, 1984; Kertesz & Hooper, 1982), comparative diagnostic classification between a Portuguese version of the WAB and a Portuguese aphasia examination (Ferro & Kertesz, 1987), and the efficacy of aphasia treatment. A study with 193 Norwegian aphasics found good agreement (85%) between classifications made with the WAB and the Norwegian Basic Aphasia Test, but only after a large number of mixed or otherwise unclassifiable patients were excluded (Sundet & Engvik, 1985); a cluster analysis of this group showed four types of aphasia—global, Wernicke, Broca, and anomia—each at a major and minor im-
pairment level. Butterworth, Howard, and McLoughlin (1984) found that semantic errors were not related to diagnostic grouping but to severity of aphasia. Semantic errors occurred on both naming and auditory comprehension, suggesting that both have a common underlying deficit.

Like the BDAE, the WAB offers a measure of spontaneous speech. The WAB measure of spontaneous speech, however, appears to be less comprehensive than the BDAE method; for example, fluency, grammatical competence, and the extent of paraphasic errors are combined into a single scale on the WAB, whereas they are assessed independently on the BDAE. Shewan and Donner (1988) also noted that the WAB spontaneous speech subtest does not provide comprehensive information compared with other tests designed to evaluate this aspect of language. The repetition test does not appear to be as encompassing or as well structured as other repetition tasks (e.g., NCCEA or MAE Sentence Repetition). Controlled associative fluency is measured as category (animal), not phonetic (letter), naming.

Mark and Thomas (1992) found auditory-verbal comprehension, as measured on the WAB, to be strongly related to outcome. This study did not find strong relationships between neuroradiological measures and outcome, whereas Metter and colleagues (1990; Metter & Jackson, 1992) found significant relations between outcome and PET scan findings in the left temporo-parietal cortex (left angular, supramarginal, postero-supero-temporal gyri). L. E. Nicholas et al. (1986) found that passage comprehension could be answered correctly better than chance without reading the sentences which the items purported to test, a finding that also applied to passage comprehension in the BDAE and the MTDDA.

In summary, the primary purpose of the WAB, like the BDAE, is diagnostic: the classification of aphasic performances into traditional aphasic syndrome subtypes. Explicit decision rules about which classification applies in an individual case are provided, but the test operates on the assumption that all cases can be clearly classified as one of eight basic types. Such clear-cut classification has limited meaning for the "mixed" aphasias that occur much more often in clinical practice than this classification system suggests. The WAB has also been used successfully in studies of recovery and treatment. It offers an additional choice for aphasia assessment which is on par with contemporary research in the field, and has found continuing use in research studies. The inclusion of a "cortical quotient" is new and unusual in an aphasia battery. The concept of a CQ, using a mixture of language and performance measures (including the Raven Progressive Matrices), seems to compete with the traditional concept of IQ (general intellectual functioning) without providing the solid psychometric and theoretical foundations for an intelligence test.