## Assessing the Assessments:
## The Adequacy of Standardized Tests of Aphasia

Craig W. Linebaugh

The George Washington University and Medical Center
Washington, DC

In 1954, Jon Eisenson's Examining for Aphasia (Eisenson, 1954) became the first test of aphasia to be made commercially available. In the 25 years since that momentuous event, some thirteen additional tests of aphasia have been published and numerous others have been described in the literature. Advances in our understanding of aphasia and measurement procedures, along with the preferences of the clinicians who employ these tests, have served to thin the forest, laying aside its weaker members so that the stronger could grow in the light of usage and systematic examination. In addition, new seedlings, hybrids featuring the combined traits of their predecessors, have sprung up and demanded our scrutiny.

The purpose of this presentation is to consider the currently available tests of aphasia with regard to a number of clinical applications. Generally, the information obtained from a test of aphasia is employed for one or more of six general purposes. These include (1) differential diagnosis, (2) localization of lesion, (3) determining the patient's level of functional communication, (4) prognosis, (5) focusing treatment, and (6) assessing recovery and the efficacy of treatment. Of course, not all tests of aphasia are structured to provide adequate samplings of communicative behavior to meet all of these purposes. This paper, therefore, will examine the merits of certain assessment instruments relative to each of these primary clinical applications.

### Differential Diagnosis

The clinician's first task when encountering a patient is to make a series of differentiations. Does the patient present some form of communicative impairment or are his communication skills within the range of normal individuals? If he does present some form of communicative impairment, is it language specific -- aphasia -- or is it a reflection of some other primary disturbance (e.g. dementia, psychosis)? Is the impairment limited to the primary language processes (morphology, phonology, syntax) or does it involve motor planning for speech -- apraxia of speech? These differentiations are critical not only for their contribution to a medical diagnosis, but for their implications for determining a patient's candidacy for treatment and the nature of that treatment.

Numerous factors may contribute to a differential diagnosis, and a test of aphasia which is to be used for this purpose must provide information relative to a substantial number of them. Among these factors is the fluency with which the patient produces speech. Thus, a test of aphasia should provide for the elicitation of verbal responses covering a wide range of length and complexity. Among the widely employed tests of aphasia, the Minnesota Test for Differential Diagnosis of Aphasia (MTDDA) (Schuell, 1965b) with its 15 subtests requiring oral responses in the "Speech and

Language Disturbances" section provides the widest range of verbal tasks. The Boston Diagnostic Aphasia Examination (BDAE) (Goodglass and Kaplan, 1972b) also includes a wide range of oral expressive tasks, plus the added feature of calling for specific judgments regarding melodic line, phrase length, and articulatory agility on its "Rating Scale Profile of Speech Characteristics." The Neurosensory Center Comprehensive Examination for Aphasia (NCCEA) (Spreen and Benton, 1969) and the Aphasia Language Performance Scales (ALPS) (Keenan and Brassell, 1975) provide less comprehensive samplings of oral expression, while the Porch Index of Communicative Ability (PICA) (Porch, 1967) includes four verbal subtests, 3 of which require only single word responses.

A second factor that is of importance for differential diagnosis is the relative impairment of specific communicative abilities. In order to provide adequate data for differentiating among various aphasic and non-aphasic communicative impairments, a test must include tasks in all language modalities covering a wide range of difficulty. Again the MTDDA provides the most comprehensive sampling of communicative skills. In addition, Schuell (1965a) provides data concerning the performance of each of her five primary diagnostic groups on each of the 47 subtests contained in the MTDDA. The BDAE also provides a wide range of task difficulty but with notable gaps in both the "Auditory Comprehension" and "Understanding Written Language" sections. These gaps are partially filled by the supplemental language tests described in the monograph which accompanies the BDAE (Goodglass and Kaplan, 1972a). Goodglass and Kaplan (1972a) do, however, provide extensive interpretive information, including profiles of various types of aphasic impairment. The NCCEA and the PICA also test a wide variety of communicative tasks, but with certain limitations with regard to range of difficulty. The developers of these tests appear to have sought to deal with the limited range of difficulty in different ways. Spreen and Benton, in formulating the NCCEA, included tasks of a relatively high degree of difficulty (e.g. Word Fluency, Sentence Construction, Parts E and F of Identification by Sentence) while excluding more traditional tasks such as picture description and paragraph comprehension. Porch took a unique approach with the PICA by employing a multidimensional scoring system to note more subtle impairments on less difficult tasks. The NCCEA and PICA differ in one other important way. Little interpretive information has been published for the NCCEA (Benton, 1967), whereas extensive study of various PICA profiles has been made. Profiles have been described corresponding to aphasia with and without various complications, bilateral brain damage (Porch, 1971b), apraxia of speech (Wertz, Rosenbek, and Collins, 1972), senile dementia (Watson and Records, 1978), right hemisphere communicative impairments (Deal et al., 1979), and even simulated aphasia (Porec and Porch, 1977).

A third factor of diagnostic significance is the pattern of responses on certain nonlanguage tasks. In general, subtests designed to assess non-language abilities are included in only the most comprehensive tests of aphasia. In particular, the MTDDA, BDAE, PICA, and NCCEA include a variety of tests of nonlanguage abilities. The nonlanguage subtests of the MTDDA include the matching of forms and letters, the imitation of gross movements of the speech mechanism, copying Greek letters, and reproducing a wheel. A complete section of the test is given to numerical relations. The BDAE includes tests of oral agility, rhythm matching, and sentence copying in the standard protocol. In addition, Goodglass and Kaplan (1972a) provide a number of supplemental nonlanguage tests including drawing to command, stick construction, three-dimensional block designs, numerical functions and a test for bucco-facial and limb apraxia. The PICA includes two matching subtests intended to assess visual perception, two copying subtests intended to determine if the patient can perceive a stimulus and duplicate it, and two pantomime subtests which assess gestural formulation and expression (Porch, 1971b). Finally, the NCCEA includes a sentence copying subtest and four control tests which are administered if items are failed on certain of the standard subtests. Two of the control tests are intended to assess manual stereognosis and two assess visual perception through object and letter matching. Note that for the most part, tests of nonlanguage abilities are included for the purpose of assessing specific processes which subserve language and the integrity of which is prerequisite to successful performance on language tasks. This limited use of nonlanguage tasks, however, is inadequate for differential diagnosis including the full range of acquired communicative impairments. The diagnosis of communicative impairments such as those related to right (nondominant) hemisphere involvement (Myers, 1978), confusion (Chedru and Geschwind, 1972), and dementia (Katzman and Karasu, 1975) requires the assessment of various nonlanguage abilities such as memory, visual perception, and intelligence. Currently available tests of aphasia examine these abilities tangentially at best, and while the instruments employed to evaluate them may better stand alone than being incorporated in tests of aphasia, the aphasiologist must nevertheless be familiar with their significance.

A fourth factor which is of particular diagnostic significance is the type of responses produced by the patient. If we are to fully understand the nature of a patient's communicative impairment, we must analyze in detail the responses produced by him and seek to infer the mediating processes by which they were produced. This type of analysis not only fosters our general understanding of language and its disorders but also permits a more precise focusing of treatment. It is unfortunate, therefore, that most tests of aphasia have discouraged this type of analysis by employing a plus-minus scoring system with the suggestion that the patient's responses be described or recorded verbatim. As Porch (1971a) has pointed out, responses need to be analyzed along a number of dimensions, and the use of descriptive or plus-minus scoring limits this analysis. Certain tests of aphasia (e.g. MTDDA, BDAE, NCCEA) do employ more discriminative scoring systems for particular subtests, but these also rely essentially on plus-minus scoring. There are, however, three notable exceptions among the commercially available tests of aphasia. The Language Modalities Test of Aphasia (LMTA) of Wepman and Jones (1961) employs a categorical scoring system for its oral and graphic subtests which places responses in specific categories (e.g. correct, phonemic or orthographic error, grammatical or syntactic error, semantic error, jargon) indicative of the psycholinguistic process in which the error occured. The PICA employs a 16-point multidimensional scoring system in which responses are scored for accuracy, responsiveness, completeness, promptness, and efficiency. This system has the added feature of being highly standardized and applicable to all the subtests within the PICA. It can also be readily adapted for use in scoring other tests of aphasia and responses elicited in treatment. The most recent highly discriminative scoring system to be described is that of the Revised Token Test (RTT) (McNeil and Prescott, 1978). This is a 15-point scoring system designed specifically for use with the auditory comprehension tasks of the RTT. In addition, McNeil and Prescott provide 11 scoring rules which serve to heighten the standardization of the system.

A fifth factor is that of the relevance of the patient's responses to the test stimuli. Halpern, Darley, and Brown (1973) described errors of relevance as "bizarre responses that appear unrelated to the stimulus" and of which the patient was seemingly unaware. They employed relevance as one of 10 categories of function used to assess acquired language impairments and found it, along with auditory retention, fluency, reading comprehension, and writing to dictation to be highly discriminative. Among their subjects, which included aphasic, apraxic, demented, and confused patients, only those with confusion showed a marked impairment with regard to relevance. DiSimoni, Darley, and Aronson (1977) extended the analysis of response relevance to include schizophrenic patients, who also showed a marked impairment in this area. More recently, Myers (1979) has reported the production of irrelevant responses by patients with right hemisphere involvement. For the most part, response relevance is not accounted for in a standardized way in tests of aphasia. Only the PICA and the RTT with their categories of "intelligible" responses account for relevance in scoring patient responses. This failure to account for response relevance is probably a result of the relative lack of irrelevant responses by aphasic patients (Halpern et al., 1973). Nevertheless, response relevance is a salient feature of patient performance, and it should be dealt with in a standardized manner.

Two additional factors which may be of diagnostic significance should be considered. DiSimoni et al. (1977), in describing the responses of their schizophrenic patients to the testing situation, noted a "tenuousness of rapport" marked by many questions regarding their performance and how the results were to be used. The other patient groups subjected to the same protocol (Halpern et al., 1973) displayed no such behavior. The second factor is that of the patient's awareness of his errors. The apraxic and confused patients described by Halpern et al. (1973) provide an instructive dichotomy relative to this factor. The former exhibited repeated efforts at self-correction, while the latter appeared to be completely unaware of erroneous and irrelevant responses. With regard to their place in tests of aphasia, these two factors have been largely ignored. The patient's response to the testing situation is included on no profiles generated by the currently available tests of aphasia. The patient's awareness of his deficit is only partially accounted for by the self-correction categories of the PICA and RTT. The inclusion of these factors in this discussion is not to suggest that they should be included on standardized tests of aphasia. Indeed, the patient's response to the test situation may best be determined through the astute application of clinical experience and insight, rather than by any standardized testing procedure. Both of these factors, however, may significantly affect patient performance, and must be considered when interpreting test results and making a differential diagnosis.

## Localization

The second purpose to which the results of standardized tests of aphasia may be applied is the determination of site of lesion. Since 1861 (Broca, 1861), it has been recognized that particular patterns of speech and language impairment may be related to involvement of specific regions of the central nervous system. With the development of sophisticated radiological techniques, the study of correlations between functional deficits and anatomical lesions has flourished. Benson (1967) reported a correlation of nonfluent aphasia with pre-Rolandic lesions and fluent aphasia with post-Rolandic lesions. A more extensive study correlating type of aphasia, as classified by the BDAE, with site of lesion, as determined by cranial computed tomography, was conducted by Naeser and Hayward (1978). Their findings generally were in agreement with the sites of lesion suggested by Goodglass and Kaplan (1972a). More recently, Mohr et al. (1978) differentiated between "dyspraxia of speaking aloud" and Broca's aphasia. The former correlates with lesions limited to the anterior operculum, while the latter correlates with more extensive lesions including the operculum, insula, and adjacent cerebrum.

Of the standardized tests of aphasia, correlations between specific patterns of response and anatomical lesions have been reported for only two. As reported above, the classifications of language impairment derived from the BDAE appear to correlate well with the predicted sites of lesion. Localization data have also been reported for the PICA (Porch, 1978). As Porch states, however, the information presented represents "initial efforts at localization with the PICA and ... must be considered as tentative." Finally, it should be emphasized that the determination of site of lesion is based on a pattern of functional deficits, rather than the scores from a particular test. With this in mind, it is apparent that any test of aphasia which provides an adequate sample of speech and language function can, with appropriate verification, be employed for localization of lesions.

## Determining the Level of Functional Communication

Traditionally, tests of aphasia have been developed with the purpose of isolating and assessing specific communicative abilities. Furthermore, efforts have been made to eliminate potentially confounding factors such as context. What this approach has served to do is remove speech and language from the natural communicative situations in which it is employed and thereby ignore the matter of functional communication. Two basic approaches to determining a patient's level of functional communication have been employed. The first involves predicting a patient's performance in natural communicative situations from his performance on standardized tasks employed in formal testing. Two examples of this approach are found in the BDAE and the PICA. Goodglass, Quadfasel, and Timberlake (1964) developed an aphasia severity rating scale which was incorporated into the BDAE. This scale has the unique feature of considering the communicative burden carried by the listener or examiner in determining the severity of the patient's communicative impairment. In addition, the scale was proven to be quite reliable, with two of three raters agreeing on 95 percent of cases.

In the interpretation of the PICA overall response levels (Porch, 1971b), characteristics exhibited by patients whose overall levels fall in a given range are described. Thus a patient whose overall response level falls above 14.00 is said to present "little or no difficulty in the usual communicative situation." On the other hand, a patient with an overall level between 8.00 and 10.00 "who uses his visual input well is often mistakenly attributed by his friends, family, and hospital personnel with having better auditory input than the tests reveal." Herein lies a recognition of the fact that the results of formal testing may underestimate a patient's functional communicative efficiency (Holland, 1977; Meek and Stewart, unpublished paper). It is interesting to note, however, that the PICA correlated with the Communication Abilities in Daily Living (CADL) (Holland, 1977), a test designed specifically to assess functional communication, with a correlation coefficient of .93. In addition, when correlated with criterion measures for patients' communicative ability in daily living, the CADL achieved a level of .60 and the PICA a level of .55, a difference of only .05. This high correlation between the two measures probably has its source in Porch's (1967) ascertion that the multidimensional scoring system of the PICA yields information that is directly related to the patient's communicative ability.

An alternative, and more direct, approach to determining a patient's level of functional communication is to assess his performance in something approaching natural communicative situations. This approach has considerably greater face validity than does the predictive approach, for as Holland (1977) has pointed out, the use of the predictive approach assumes that acontextual tests actually measure communication. The use of tests which directly assess functional communication, of course, eliminates this difficulty. To date, two such tests have been developed. One, the Functional Communication Profile (FCP) (Sarno, 1969) employs a 9-point rating scale to assess performance on 45 communicative tasks. Good reliability in the use of the rating scale has been demonstrated, as evidenced by an inter-judge correlation of .95 for the FCP overall score. An item analysis of the FCP, however, suggests that the test directly assesses functional communication in only a limited way. Several items on the Understanding section require that the rating be inferred or determined from the patient's report of his success on a given task. Ratings for the Reading section are based on the patient's performance on some standardized test, while the Other section, which includes writing, requires the elicitation of specific responses. For both of these sections, the ratings are based on information obtained from essentially acontextual communicative situations.

The second test of this type, the CADL (Holland, 1977) seeks to provide context in a more natural, yet controlled, manner. It involves the "manufacturing" of context by the use of various props and "scene-setting" statements. In addition, the CADL appears to be the best measure of functional communication developed to date as evidenced by its relatively high correlation of .60 with a criterion measure based on observations of patients in natural settings.

## Prognosis

Perhaps the most challenging and perplexing task with which aphasiologists are routinely faced is that of predicting the eventual extent of an individual patient's recovery. Traditionally, a factor approach to prognosis has been employed in which patient parameters such as age, time post onset, site of lesion, etc. have been singled out as predictors. Within this approach, the results of aphasia testing have been employed only to the extent that initial severity and the degree of impairment in auditory comprehension are potentially good indicators of expected recovery.

Three tests of aphasia have, however, specifically addressed the matter of prognosis, and some recent research has supplied data which allow the results of a fourth to be employed for this purpose. The authors of one of these tests, the ALPS, suggest using a combination of test results and history information. They provide specific, test-based guidelines only for their Talking scale, and here, given the large increments in communicative ability that exist between certain points on the scale, the predicted outcome for a given initial score extends over a wide range.

A markedly different approach to prognosis was employed by Schuell in developing the MTDDA. As noted above, the results of the MTDDA are used to differentiate among five major and two minor categories of aphasia. A general prognosis is then stated for each of the categories. In addition, a comparison of a given patient's scores on specific subtests with those of his diagnostic group may provide additional prognostic information. That is, those patients who exceed the mean performance of their diagnostic group on those subtests which correlate particularly well with the other subtests in a given section of the MTDDA may be expected to experience a more complete recovery. The categoric approach to prognosis may also be applied to the BDAE in light of recent findings by Kertesz and McCabe (1977). These authors, using the Western Aphasia Battery, which is a modification of the BDAE, described recovery patterns for Broca's, Wernicke's, conduction, transcortical, anomic, and global aphasic patients. It would appear that their classifications are compatible with those derived from the BDAE, and thus their data regarding recovery may be used in formulating prognoses for patients classified using either test.

A third approach to prognosis may be described as statistical in nature and is exemplified by the PICA. A number of measures derived from patient scores on the PICA have been employed to predict patient recovery. Porch (1971b) originally derived a set of predictive curves, called HOAP slopes, from overall percentile change in patients' performance between one and six months post onset. He has also used the "high-low gap," the percentile difference between the mean of the nine highest and the nine lowest subtest scores. In 1973, Porch, Wertz and Collins (1973) employed five variables—gestural, verbal, and graphic modality scores, the high-low gap, and the patient's age, all at one month post onset—in multiple linear regression analyses to predict overall scores at three and six months post onset. The multiple correlation coefficient obtained for the three-month score was .96 and that for the six-month score was .84. More recently, Porch (1978) has suggested the use of peak-mean difference in determining prognosis. This measure, which is the cumulative difference between the highest score and the mean for each subtest, was based on the notion that the peak score is the best indication of potential patient performance. Last year, Aten and Lyon (1978) reported on the use of five measures of within subtest variance, including two forms of peak-mean difference, as predictors of recovery. Their findings, which indicated only minimal predictive value for the measures employed, essentially reflect the state of our prognostic art. This is particularly true with regard to the individual patient. It is apparent that a great deal of work remains to be done in this area, including the identification of the most salient predictors of recovery and the most precise way of measuring them.

## Focusing Treatment

Perhaps the most important criterion for a test of aphasia is that it provide information which can be reliably employed in focusing treatment. Indeed, if a test meets this criterion, it probably will meet nearly all others and be applicable to all the other clinical purposes discussed in this paper. In formulating goals and selecting an appropriate treatment approach, several factors must be considered. In general, these factors may be divided into psycholinguistic and processing factors. In addition, the level of breakdown and the types and patterns of errors provide important information. While time does not permit a full delineation of the extent to which the various tests of aphasia provide information regarding each of these areas, certain factors within each area can be considered.

Psycholinguistic Factors. The literature on aphasia is replete with reports of the effects of various psycholinguistic variables on aphasic linguistic performance. For example, Schuell, Jenkins, and Landis (1961) reported that aphasic patients appear to comprehend words with high frequency of occurence more readily than those with lower frequency of occurence. Shewan and Canter (1971) described the effects of length and syntax, as well as vocabulary, on auditory comprehension. It is rather curious, therefore, that psycholinguistic variables have been employed in the development of tests of aphasia in so limited a manner. Of the generally familiar tests of aphasia, only four report the use of any psycholinguistic criterion for vocabulary selection. The vocabulary included in certain subtests of the MTDDA and the Auditory Comprehension Test for Sentences (ACTS) (Klor, 1977) was selected on the basis of frequency of occurence. Benton (1967) reported that the vocabulary for the NCCEA was selected on the basis of acquisition age, the age at which children were able to name an object under testing conditions. Finally, Goodglass and Kaplan (1972a) selected the vocabulary for the BDAE to represent six semantic categories, but with no apparent criteria for selection within the categories. One additional test, the experimental Boston Naming Test (Kaplan, 1977) also employed frequency of occurence as a criterion for item selection.

Other psycholinguistic variables have been employed with even less frequency in test design. In developing the ACTS, Shewan and Klor (Klor, 1977) also considered length and syntax. Length was controlled for both number of critical elements and number of syllables. Syntax was controlled with regard to the number of transformations in a given sentence. Certain subtests of the NCCEA were also controlled for length. The stimuli of the Sentence Repetition subtest were controlled with regard to the number of syllables, while the items of the Identification by Sentence (Token Test) were controlled for number of critical elements. Part F of the Identification by Sentence subtest varies the syntactic complexity of the stimuli, but in no systematic manner. The same is true of the original Token Test of DeRenzi and Vignolo (1962). McNeil and Prescott (1978), however, in the RTT, have remedied this problem by developing specific subtests to assess the comprehension of locative prepositions, left vs. right, and conditional statements.

One final comment regarding the use of psycholinguistic factors in developing tests of aphasia is in order. While psycholinguistic variables may provide useful guidelines in initial item development, they should not be the only criterion on which items are ultimately selected for inclusion in a given subtest. Porch (1971a), with regard to stimuli and, more recently, McNeil and Prescott (1978) with regard to the motor response, have stressed the importance of within-subtest homogeneity. That is, the items contained within a given subtest should be of equal difficulty. Exceptions to this rule may be permitted where an acceptable number of items are sampled at given increments of some variable (e.g. length, word frequency) within a subtest, but the principle remains the same. Items to be included in a given subtest or at a given level of complexity within a subtest (e.g. Identification by Sentence subtest of the NCCEA) must be selected on the basis of being essentially equal in difficulty, not on the basis of conformance to some psycholinguistic criterion.

Information Processing Factors. As Haaland (1979) has reported, a number of information processing factors play a significant role in language performance and must, therefore, be accounted for in the assessment of aphasia. One such factor which has been given particularly close scrutiny is that of stimulus and response modality. Not only have the traditional linguistic modalities of auditory and reading comprehension and oral and graphic expression been examined, but the effectiveness of presenting stimuli in various sensory modalities and the comprehension and expression of gestures have also been investigated. In developing tests of aphasia, the approach has essentially been one of assessing performance in various combinations of input and output modalities for the purpose of isolating modality specific deficits and identifying which of them may be utilized in facilitating communication. Emerick (1971), in particular, has focused on combinations of input and output modalities in his Appraisal of Language Disturbance (ALD). More traditionally, however, the input modalities employed have been the visual (objects, pictures), graphic, and auditory-verbal, while the most frequently employed output modalities have been the gestural (pointing), verbal, and graphic. Two notable exceptions deserve mention. One is the two function demonstration subtests of the PICA, and the other is the tactile naming subtests of the NCCEA.

As with the stimulus modality, the nature of the stimulus also has a bearing on patient performance. The literature regarding the physical characteristics of stimuli is somewhat ambiguous, but it would appear that the use of objects as opposed to black and white line drawings would be of no significant benefit to most patients (Corlew and Nation, 1975). Rather the selection of stimuli with regard to their physical characteristics is largely a matter of the abilities to be assessed.

While the physical characteristics of the stimuli may be of little significance, the complexity of the stimuli is an entirely different matter. The difficulty here, however, is arriving at a common means of describing complexity. In reality, no single measure of complexity can be applied to all the abilities that one may wish to sample. For example, in a test of sentence comprehension, complexity may be assessed in terms of the number of syllables (e.g. ACTS), transformations (e.g. ACTS), or critical elements (e.g. ACTS, BDAE, NCCEA) in a given sentence. On the other hand, complexity of stimuli in a word repetition task may be determined on the basis of number of syllables or phonetic context. As noted above, the critical factor is not the measure of complexity, but that the stimuli within a given subtest be homogeneous. Furthermore, maintaining as many common features as possible across subtests (e.g. PICA) will allow for a more valid assessment of relative deficits and a more precise identification of potential facilitators.

A third factor of great significance is that of retention and memory. As Cermak and Moreines (1976) have demonstrated, aphasic patients appear to have memory impairments related to their linguistic deficit. Conversely, as Haaland (1979) has observed, any test which requires patients to repeat or comprehend a stimulus is influenced by memory factors. Unfortunately, memory factors are not controlled in a systematic way in any test of aphasia. The influence of certain factors may be inferred, however, from particular tests. For example, the effects of proactive and retroactive interference may be inferred on the basis of which elements in a serial repetition or sentence comprehension task are most frequently missed. Sentence repetition tasks, particularly that of the BDAE with its high and low probability sentences, and comprehension tasks with carefully manipulated foils (e.g. ACTS) may provide information regarding the role of semantic encoding in memory. Nevertheless, tests of aphasia provide little information regarding memory and, perhaps even more importantly from the viewpoint of rehabilitation, largely fail to consider the effects of memory impairments on communication.

Three other information processing variables have been given some consideration in the development of tests of aphasia. One is processing time, which is measured as response latency. While latency has been widely recognized as an important factor in aphasic behavior, only the PICA, RTT, and BDAE incorporate it as a critical feature in response analysis. A second is the order of presentation of test items with regard to difficulty. Brookshire (1972) reported that the presence of difficult items in a list of otherwise easier items, reduced naming performance on the easy items. In this vein, tests of aphasia have been organized in order of increasing task difficulty. The exception is the PICA which was ordered according to decreasing task difficulty in order to minimize learning effects. Recently, Dumond, Hardy, and Van Demark (1978) compared the standard hard-to-easy order of presentation of the PICA subtests to a reversed easy-to-hard order and reported no significant difference in performance. The third factor which bears mention is that of structure. Most tests of aphasia impose a strict item-by-item structure on the patient. Again the PICA is an exception. On subtests I, II, A, and B the patient must formulate his own order of response. The effects of task structure may be analyzed by assessing the patient's response to the cues which may be provided on these four tests and by comparing performance on subtests II and III.

It should also be noted that there are numerous other information processing factors which are not now accounted for in tests of aphasia. Such factors as rate of presentation, exposure time, inter-stimulus interval, and error detection have been shown to play an important role in aphasic performance. Perhaps the wider application of information processing analyses such as those proposed by Haaland (1979) would allow for a fuller delineation of those processing factors which are influencing particular patients and to what degree.

<u>Level of Breakdown/Pattern of Response</u>. It is a well-established principle in aphasia rehabilitation that treatment should begin at that point where the patient begins to experience difficulty—the level of breakdown. More specifically, the level of breakdown might be described as that point on a continuum of tasks within each communicative modality where the patient is unable to perform at some predetermined level of accuracy. Other dimensions of patient performance (e.g. completeness, latency, efficiency) might also be considered when determining the level of

breakdown. It is readily apparent that if a test of aphasia is to provide information as to the level of breakdown, it must include tasks covering a wide range of difficulty in those communicative modalities it is intended to assess (Osgood and Miron, 1963). In addition, there should be a high degree of across-subtest homogeneity in order that the patient's performance in various modalities can be compared to determine potential facilitators.

Several tests of aphasia provide useful information with regard to determining the level of breakdown. The MTDDA, BDAE, and NCCEA share the common strength of assessing performance over a fairly substantial range of difficulty and the common weakness of inadequately sampling performance at given points within that range. The RTT and the ACTS both provide extensive information concerning various parameters of auditory comprehension at the sentence level. Two other tests which provide some guidance with regard to focusing treatment in the test manuals bear special mention.

The scales of the ALPS were constructed to sample performance over a wide range of message length and complexity. In addition, the authors provide data on several patients and demonstrate how the results can be employed in planning treatment. There are, however, two weaknesses in the construction of the ALPS which limit its usefulness for this purpose. One is the insufficient number of items at each level of difficulty. The other is that the increments between the ten levels in each modality are not equal. Indeed, at some points (e.g. between items 7 and 8 on the Talking scale, between items 6 and 7 and 8 and 9 on the Reading scale) the increase in task difficulty is extremely large.

The PICA also provides direction in planning treatment. The 18 subtests of the PICA are organized along a continuum of task difficulty and the level of breakdown can be determined by examining the "Ranked Response Summary" for that point where performance begins to decline. The limitation of the PICA is that the continuum of task difficulty does not extend sufficiently toward the "hard" end. This is particularly true with regard to auditory comprehension (Kearns and Hubbard, 1977) but applies to all modalities, in that stimuli and responses of no greater length and complexity than a simple sentence are employed. This limitation is partially overcome by the more precise delineation of responses via the multidimensional scoring system, but an extension of the task continuum would permit a more precise focusing of treatment, particularly for high-level patients.

In concluding this section on focusing treatment, a few comments regarding error types and patterns of response are pertinent. In the section of this paper dealing with differential diagnosis, the analysis of patterns of response in order to differentiate among various communicative impairments was emphasized. Inherent in this differentiation of impairments is a differentiation among mechanisms by which errors are generated. For example, the phonetic substitution errors of an apraxic patient are attributed to a breakdown in motor programming, while those of a fluent aphasic patient are due to a deficit in phoneme selection. Specific information processing deficits may also be inferred from patterns of response as demonstrated by Brookshire (1972) with regard to auditory comprehension. The value of this type of analysis relative to rehabilitation is, of course, that the identification of an error-generating mechanism or specific processing deficit permits a more precise focusing of treatment. In order for error patterns to emerge, however, a test must include a substantial number of items (McNeil and Prescott, 1973). Furthermore, the items within a given subtest should be homogeneous with regard to difficulty. Thus in developing or selecting a test of aphasia, these two psychometric criteria should be given careful consideration.

## Assessing Recovery and the Efficacy of Treatment

The final application of tests of aphasia to be discussed is that of assessing recovery and the efficacy of treatment. These uses place three special constraints on the tests to be employed. First, if performance is to be compared across communicative tasks and modalities, some form of standard score must be used. This is required in order to insure equivalence of scoring (Ludlow, 1977). At this time, only four tests of aphasia provide standard scores. These are the BDAE, which employs z-scores, and the PICA, NCCEA, and RTT, which employ percentiles.

A second constraint is the use of a scoring system which is sensitive to subtle changes in patient performance. On the basis of this criteria the PICA and the RTT are the tests of choice. The multidimensional scoring systems employed in these tests provide more precise information over a greater number of performance parameters than do the scoring systems of the BDAE and NCCEA.

The third constraint is that of a high degree of test stability over time. Thus tests to be employed for these purposes must have good test-retest reliability. To a large extent this is dependent on the extent to which the test administration has been standardized. Here again the PICA and the RTT are the tests of choice. The NCCEA is somewhat more standardized in its administration than the BDAE, but no reliability data have been published on it to date.

In selecting tests to be employed over repeated administrations for the purpose of assessing recovery or the efficacy of treatment, the PICA and the RTT emerge as the tests of choice. The RTT, of course, assesses only auditory comprehension, while the PICA presents the previously-discussed limited range of difficulty. Between the two, only the PICA possesses the advantage of assessing performance on a number of different tasks in order that an untreated, baseline ability can be monitored to assess spontaneous recovery or cross-modality generalization. In the final analysis, it may be stated that no perfect test exists for assessing recovery or the efficacy of treatment, and given the difficulties inherent in measuring these phenomenon, one is not likely to be forthcoming until all the variables contributing to them are much better understood.

## Conclusion

In summarizing the ideas presented in this paper, it may be said that the last 25 years have witnessed major advancements in the assessment of acquired language impairments. Both the better understanding of these impairments contributed by those interested in them and the application of psychometric principles have contributed to the improvement in our tests. It is important to recall, however, that in spite of these improvements, the tests of aphasia are no more than tools. They are no substitute for the knowledge and insight of the experienced clinician. In addition, substantial improvement, particularly in the areas of differential diagnosis, prognosis, and focusing treatment, remains to be made. With the careful consideration of variables such as context, cueing, learning potential, and recovery patterns, and the continued application of psychometric principles, this improvement can be achieved, and our tests of aphasia can be made even more useful.

## References


Aten, J.L. and Lyon, J.G. Measures of PICA subtest variance: A preliminary assessment of their value as predictors of language recovery in aphasic patients. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1978. Minneapolis: BRK Publishers (1978).

Benson, D.F. Fluency in aphasia: Correlation with radioactive scan localization. Cortex 3, 373-394 (1967).

Benton, A.L. Problems of test construction in the field of aphasia. Cortex 3, 32-58 (1967).

Broca, P. Remarques sur le siege de la faculte du langage articule, suivis d'une observation d'aphemie. Bulletin de Societe Anatomie (Paris) 6, 330-357 (1861).

Brookshire, R.H. Differences in auditory processing deficits among aphasic patients. Paper presented to the American Speech and Hearing Association convention, San Francisco (1972).

Brookshire, R.H. Effects of task difficulty on naming by aphasic subjects. Journal of Speech and Hearing Research 15, 551-558 (1972).

Cermak, L.S. and Moreines, J. Verbal retention deficits in aphasic and amnesic patients. Brain and Language 3, 16-27 (1976).

Chedru, F. and Geschwind, N. Disorders of higher cortical functions in acute confusional states. Cortex 8, 395-411 (1972).

Corlew, M. and Nation, J. Characteristics of visual stimuli and naming performance in aphasic adults. Cortex 11, 186-191 (1975).

Deal, J.L., Deal, L., Wertz, R.T., Kitselman, K., and Dwyer, C. Right hemisphere PICA percentiles: Some speculations about aphasia. Paper presented to the Clinical Aphasiology Conference, Scottsdale, AZ (1979).

DeRenzi, E. and Vignolo, L.A. The token test: A sensitive test to detect receptive disturbances in aphasics. Brain 85, 665-678 (1962).

DiSimoni, F.G., Darley, F.L., and Aronson, A.E. Patterns of dysfunction in schizophrenic patients on aphasia test battery. Journal of Speech and Hearing Disorders 42, 498-513 (1977).

Duffy, J.R., Keith, R.L., Shane, H., and Podraza, B.L. Performance of normal (non-brain-injured) adults on the Porch Index of Communicative Ability. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1976. Minneapolis: BRK Publishers (1976).

Dumond, D.L., Hardy, J.C., and Van Demark, A.A. Presentation by order of difficulty of test tasks to persons with aphasia. Journal of Speech and Hearing Research 21, 350-360 (1978).

Goodglass, H., Quadfasel, A., and Timberlake, W.H. Phrase length and the type and severity of aphasia. Cortex 1, 133-153 (1964).

Goodglass, H. and Kaplan, E. The Assessment of Aphasia and Related Disorders. Philadelphia: Lea and Febiger (1972a).

Goodglass, H. and Kaplan, E. Boston Diagnostic Aphasia Examination. Philadelphia: Lea and Febiger (1972b).

Haaland, K.Y. The "what" of speech and language evaluation. Paper presented to the Clinical Aphasiology Conference, Scottsdale, AZ (1979).

Halpern, H., Darley, F.L., and Brown, J.R. Differential language and neurologic characteristics in cerebral involvement. Journal of Speech and Hearing Disorders 38, 162-173 (1973).

Holland, A. Comment on "Spouses Understanding of the Communication Disabilities of Aphasic Patients." Journal of Speech and Hearing Disorders 42, 307-308 (1977).

Holland, A.L. Estimators of aphasic patients' communicative performance in daily living. Final report submitted to the National Institute of Neurological and Communicative Disorders and Stroke (1977).

Eisenson, J. Examining for Aphasia. New York: The Psychological Corporation (1954).

Emerick, L. Appraisal of Language Disturbance. Marquette, MI: Northern Michigan University Press (1971).

Kaplan, E. Boston naming test. Unpublished.

Katzman, R. and Karasa, T. Differential diagnosis of dementia. In Fields, W.S. (Ed.) Neurological and Sensory Disorders in the Elderly. New York: Stratton Intercontinental Medical Book Corporation (1975).

Kearns, K. and Hubbard, D.J. A comparison of auditory comprehension tasks in aphasia. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1977. Minneapolis: BRK Publishers (1977).

Keenan, J.S. and Brassell, E.G. Aphasia Language Performance Scales. Murfreesboro, TN: Pinnacle Press (1975).

Kertesz, A. and McCabe, P. Recovery patterns and prognosis in aphasia. Brain 100, 1-18 (1977).

Klor, B. Clinical implications for a test of sentence comprehension. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1977. Minneapolis: BRK Publishers (1977).

Ludlow, C. Recovery from aphasia: A foundation for treatment. In Sullivan, M. and Kommers, M.S. (Eds.) Rationale for Adult Aphasia Therapy. Omaha: University of Nebraska Medical Center (1977).

McNeil, M.R. and Prescott, T.E. Revised Token Test. Baltimore: University Park Press (1978).

Meek, S.D. and Stewart, D.B. Performance of severely aphasic patients in a formal versus informal testing situation. Unpublished paper.

Mohr, J.P., Pessin, M.S., Finkelstein, S., Funkenstein, H.H., Duncan, G.W., and Davis, K.R. Broca aphasia: Pathologic and clinical. Neurology 28, 311-324 (1978).

Myers, P.S. Analysis of right hemisphere communication deficits: Implications for speech pathology. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1978. Minneapolis: BRK Publishers (1978).

Myers, P.S. Profiles of communication deficits in patients with right cerebral hemisphere damage: Implications for diagnosis and treatment. Paper presented to the Clinical Aphasiology Conference, Scottsdale, AZ (1979).

Naeser, M.A. and Hayward, R.W. Lesion localization in aphasia with cranial computed tomography and the Boston Diagnostic Aphasia Exam. Neurology 28, 545-551 (1978).

Osgood, C.E. and Miron, M.S. (Eds.) Approaches to the Study of Aphasia. Urbana, IL: University of Illinois Press (1963).

Porch, B.E. Diagnosis of aphasia and brain dysfunction. Albuquerque, VA Satellite Television Presentation (1978).

Porch, B.E. Multidimensional scoring in aphasia testing. Journal of Speech and Hearing Research 14, 776-792 (1971a).

Porch, B.E. Porch Index of Communicative Ability. Palo Alto, CA: Consulting Psychologists Press (1967).

Porch, B.E. Porch Index of Communicative Ability. Volume 2: Administration, Scoring, and Interpretation. Rev. Ed. Palo Alto, CA: Consulting Psychologists Press (1971b).

Porch, B.E. Profiles of aphasia: Test interpretation regarding the localization of lesions. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1978. Minneapolis: BRK Publishers (1978).

Porch, B.E. The use of homogeneous subtest items in assessing aphasia. Paper presented to the American Speech and Hearing Association convention, Chicago (1971c).

Porch, B.E., Wertz, R.T., and Collins, M.J. Recovery of communicative ability: Patterns and predictions. Paper presented to the Academy of Aphasia, Albuquerque (1973).

Porec, J. and Porch, B.E. The behavioral characteristics of "simulated" aphasia. In Brookshire, R.H. (Ed.), Clinical Aphasiology, Conference Proceedings, 1977. Minneapolis: BRK Publishers (1977).

Sarno, M.T. The Functional Communication Profile. New York: Institute of Rehabilitation Medicine, New York University Medical Center (1969).

Schuell, H. Differential Diagnosis of Aphasia. Minneapolis: University of Minnesota Press (1965a).

Schuell, H. The Minnesota Test for Differential Diagnosis of Aphasia. Minneapolis: University of Minnesota Press (1965b).

Schuell, H., Jenkins, J., and Landis, L. Relationship between auditory comprehension and word frequency in aphasia. Journal of Speech and Hearing Research 4, 30-36 (1961).

Shewan, C.M. and Canter, G.J. Effects of vocabulary, syntax, and sentence length on auditory comprehension in adult aphasic patients. Cortex 7, 208-226 (1971).

Spreen, O. and Benton, A.L. Neurosensory Center Comprehensive Examination for Aphasia. Victoria, B.C.: University of Victoria (1969).

Watson, J.M. and Records, L.E. The effectiveness of the Porch Index of Communicative Ability as a diagnostic tool in assessing specific behaviors of senile dimentia. In Brookshire, R.H. (Ed.) Clinical Aphasiology, Conference Proceedings, 1978. Minneapolis: BRK Publishers (1978).

Wepman, J.M. and Jones, L.V. The Language Modalities Test for Aphasia. Chicago: Education-Industry Service (1961).

Wertz, R.T., Rosenbek, J.C., and Collins, M.J. Identification of apraxia of speech from PICA verbal tests and selected oral-verbal apraxia tests. Paper presented to the Clinical Aphasiology Conference, Albuquerque (1972).