Classifying the Aphasias: A Comparison of the Boston Diagnostic Aphasia Examination and the Western Aphasia Battery

Robert T. Wertz, Jon L. Deal, and Alice J. Robinson
Veterans Administration Medical Center, Martinez, California

Comparison of tests for aphasia is popular (Holland, 1980; Sanders and Davis, 1978; Wertz, Keith, and Custer, 1971). Typically such comparisons are done to validate a new test or to compare tests that differ in design and content. Results of comparisons usually yield significant correlations among the tests compared. An exception was our comparison (Wertz, Kitselman, and Deal, 1981) of the Boston Diagnostic Aphasia Examination (BDAE) (Goodglass and Kaplan, 1972) and the Western Aphasia Battery (WAB) (Kertesz, 1982). Agreement between the two measures in the classification of patients into aphasic types was 68 percent. However, we did not administer either the BDAE or WAB to our sample of aphasic patients. We selected fluency, auditory comprehension, repetition, and naming tasks from a battery of measures we had administered and used these to classify patients according to the BDAE types and the WAB taxonomy. Thus, the differences in classification between the two systems might have resulted from the methods we employed and might not represent differences in the two measures.

The purpose of this paper is to report a comparison of aphasic patients' performance on the BDAE and the WAB. Both tests were administered to all patients in our sample, and we attempted to classify each patient into a specific type of aphasia using each system's procedures for classification.

METHOD

Forty-five aphasic patients were tested with the BDAE and the WAB. All had suffered a single, left hemisphere thromboembolic cerebral vascular accident (CVA). Descriptive data (Table 1), indicate that our sample displayed a wide range in age, education, months postonset, and severity of aphasia. Patients less than three months postonset at the time of testing received both measures within a two-week period. Patients beyond three months postonset received both measures within a one-month period. Order of test administration was random across patients.

Table 1. Descriptive data for study patients at the time of testing.

VARIABLE	$\bar{X}$	Range	S.D.
Age in Years	57.89	26-84	10.99
Education in Years	12.02	8-16	2.07
Months Postonset	15.93	1-80	22.29
PICA Overall %ile	55.71	9-98	25.06

Both measures were administered using standardized procedures. Each patient was classified into an aphasia type category by his clinician, who used the system specified for each measure--cut-off scores for the WAB taxonomy, and Rating Scale Profile of Speech Characteristics ranges for each

BDAE type. Reliability of classification for each measure was obtained by having a second clinician independently classify each patient. Percent agreement between clinicians was 91 percent for the BDAE and 96 percent for the WAB. After reliability was computed, data on patients who generated disagreement were reevaluated with the assistance of a third clinician, and these patients received a final classification. Thus, all patients were classified into one of the eight WAB types--Global, Broca's, Isolation, Transcortical Motor, Wernicke's, Transcortical Sensory, Conduction, and Anomic. To this was added "not aphasic" for patients whose Aphasia Quotient exceeded 93.8, the WAB cut-off score, and "unclassifiable" for patients who could not be classified with the WAB criteria for classification cut-off scores (Kertesz, 1979, p. 58). Similarly, all patients were classified into nine BDAE types--Global, Broca's, Transcortical Motor, Wernicke's, Transcortical Sensory, Conduction, Anomic, Mixed, and Unclassifiable--using the profile ranges provided by Goodglass and Kaplan (1972).

RESULTS

Analysis of patient performance on the WAB was limited to those measures used to classify--fluency, auditory comprehension, repetition, and naming. As shown in Table 2, our patients displayed a wide range of severity in the Aphasia Quotient (from below one to 98), and in each of the four WAB subsections. A similar range of severity was observed on the BDAE. On the Severity Rating Scale, patient performance ranged from zero to five, and on the seven characteristics in the Rating Scale Profile, performance was sprinkled across the possible range. The correlation between the WAB Aphasia Quotient (AQ) and the BDAE severity rating was +.83, significant at $p < .001$. Similarly, both measures correlated significantly ($p < .001$) with the Porch Index of Communicative Ability (PICA) (Porch, 1967) Overall percentile--the WAB at +.89 and the BDAE at +.77.

Table 2. Group performance on the BDAE and the WAB.

MEASURE AND RANGE OF SCORE	$\bar{X}$	Range	S.D.
BDAE			
Severity Rating Scale (0-5)	2.47	0-5	1.31
Melodic Line (1-7)	4.69	1-7	1.86
Phrase Length (1-7)	4.40	1-7	2.00
Articulatory Agility (1-7)	4.11	1-7	1.69
Grammatical Form (1-7)	4.71	1-7	2.06
Paraphasia In Running Speech (1-7)	5.38	2-7	1.95
Word Finding (1-7)	4.56	2-7	1.63
Auditory Comprehension (1-7)	4.96	1-7	1.87
WAB			
Aphasia Quotient (0-100)	66.22	0.75-98.00	30.19
Fluency (0-10)	5.38	0-10	3.13
Comprehension (0-10	7.33	0.75-10	2.62
Repetition (0-10)	6.28	0-10	3.69
Naming (0-10)	5.78	0-10	3.46

Comparison among the subtests in each measure used to classify the type of aphasia, shown in Table 3, also yielded significant correlations ($p < .001$).

Total scores for fluency, auditory comprehension, repetition, and naming on the WAB and BDAE were used in this comparison.

Table 3. Correlations between BDAE and WAB subtests.

COMPARISON	r*
Fluency	.84
Auditory Comprehension	.88
Repetition	.94
Naming	.92

*All correlations significant at $p < .001$.

When patients were classified into aphasia types (Table 4), agreement between the two measures diminished. On the BDAE, our 45 patients were distributed among the various types except for transcortical motor and transcortical sensory. The majority of patients (62 percent) were unclassifiable. On the WAB, all types except isolation and transcortical sensory were represented, and the most frequent type was anomic (33 percent). All patients were rated aphasic on the BDAE. However five patients were rated not aphasic on the WAB, because their AQs exceeded the 93.8 cut-off score. Five patients (11 percent) could not be classified on the WAB, compared with 28 patients (62 percent) that could not be classified on the BDAE. Only two patients could not be classified on either measure.

Table 4. Classification of patients on the BDAE and the WAB.

TYPE	NUMBER OF PATIENTS	
	BDAE	WAB
Global	4	6
Broca's	6	6
Isolation	N/A*	0
Transcortical Motor	0	1
Wernicke's	2	4
Transcortical Sensory	0	0
Conduction	1	3
Anomic	3	15
Mixed	1	N/A*
Unclassifiable	28	5
Not Aphasic	0	5

*N/A = Not applicable

The percent agreement between the two measures for individual types of aphasia, shown in Table 5, ranged from 67 percent for Global to zero percent for transcortical motor, conduction, and not aphasic. Overall agreement in classification was 27 percent.

Table 5. Percent agreement in classification of aphasic types between the BDAE and the WAB.

TYPE	% AGREEMENT
Global	67
Broca's	33
Transcortical Motor	0
Wernicke's	50
Conduction	0
Anomic	13
Unclassifiable	40
Not Aphasic	0
Overall	27

A confusion matrix (or who was rated what on who's whom) was constructed. As shown in Table 6, two of the six WAB Global patients were mixed or unclassifiable on the BDAE. Four of the six WAB Broca's patients were unclassifiable on the BDAE. One WAB transcortical motor patient was unclassifiable on the BDAE. Two of four WAB Wernicke's patients were unclassifiable on the BDAE. None of the three WAB conduction patients were classified as Conduction on the BDAE. One was Broca's, and two were unclassifiable. Only two of the 15 WAB anomic patients were anomic on the BDAE. One was classified Broca's, and 12 were unclassifiable. Two of the five unclassifiable patients on the WAB were unclassifiable on the BDAE, but two were classified Broca's and one was classified conduction. Four of the five WAB "not aphasic" patients were labeled unclassifiable on the BDAE, and the other was classified anomic.

DISCUSSION

These two tests, designed to classify the aphasias, show good agreement in the way severity of aphasia is rated, and they show good agreement on the subtests used to classify the aphasias. However, they display only 27 percent agreement in the way patients are classified. Except for Global and Wernicke's, there was less than 50 percent agreement between the two measures in classifying all other types of aphasia.

Significant correlations between the WAB and BDAE severity ratings and test subsections--fluency, auditory comprehension, repetition, and naming--indicate that both measures are similar in the way they measure severity of aphasia and severity within the four modalities. However, if our data are representative, clinicians using one measure to classify the type of aphasia will classify patients differently from clinicians using the other measure move than 70 percent of the time. One wonders why two measures designed to do the same thing do that thing so differently.

An obvious difference between the WAB and BDAE is the tendency for the former to classify almost all patients and the latter's failure to classify approximately 60 percent of the patients. The discrete cut-off scores in the WAB criteria for classification, Kertesz tells us, are based on a review of performance by 150 patients. "These numbers represent ranges of scores chosen to classify all aphasics unequivocally" (Kertesz, 1979, p. 57). Our failure to classify five patients resulted from WAB naming scores that

Table 6. Disagreement in classification between the BDAE and the WAB.

TYPE	NUMBER OF PATIENTS												% AGREEMENT
	WAB	BDAE											
		Global	Broca's	Isolation	Trans. Motor	Wernicke's	Trans. Sensory	Conduction	Anomic	Mixed	Unclass.	Not Aphasic	
Global	6	4								1	1		67 %
Broca's	6		2								4		33 %
Isolation	0												-
Trans. Motor	1										1		0 %
Wernicke's	4					2					2		50 %
Trans. Sensory	0												-
Conduction	3		1								2		0 %
Anomic	15		1						2		12		13 %
Mixed	N/A												-
Unclass.	5		2					1			2		40 %
Not Aphasic	5								1		4		0 %
TOTAL	45	4	6	N/A	0	2	0	1	3	1	28	0	27 %

exceeded the cut-off for naming but other performance too low to meet the nonaphasic cut-off score. Nevertheless, almost all of our patients, 89 percent, were classified on the WAB.

Conversely, the BDAE failed to classify 62 percent of our sample. This would not surprise Goodglass and Kaplan (1983), because they state that BDAE ". . . scores do not objectively and automatically classify the patient" (p. 2). Further they report, "Estimates of the proportion of cases that can be unambiguously classified range from 30 percent in some centers to 80 percent in others" (p. 74). Eighty-two percent of the disagreement between measures in our sample resulted from the WAB classifying patients the BDAE did not.

The other sources of disagreement between the WAB and the BDAE classification resulted from (a) the WAB not classifying patients the BDAE did, (b) differences in the way each measure rates fluency, (c) differences in the severity of auditory comprehension deficit, and (d) differences in whether the patient was rated aphasic.

As mentioned earlier, the WAB's failure to classify some of our patients resulted from naming scores that exceeded the WAB naming cut-off score, but the patient was too severe on other measures to be classified nonaphasic. The BDAE classified some of these patients, in part, because the way naming is rated (as "Word Finding") on the BDAE Rating Scale Profile of Speech Characteristics is as much a measure of fluency as it is a measure of naming ability. The WAB primarily measures naming through confrontation naming.

Similarly, the WAB fluency scale differs somewhat from the way the BDAE rates fluency. The WAB employs a ten-point scale to determine fluency, while the BDAE utilizes four of the seven Rating Scale speech characteristics (melodic line, phrase length, articulatory agility, grammatical form) to separate the fluent from the nonfluent. Further, the WAB fluency scale permits nonfluent patients to display some paraphasias. The BDAE does not.

Our clinical impression is that the WAB auditory comprehension subtests may be less difficult than the BDAE auditory comprehension subtests, and this may result from differences in administration. For example, the BDAE moves among classes (nouns, letters, geometric shapes, colors) when testing auditory comprehension, but the WAB remains within class (tests noun recognition, then letter recognition, then geometric shape recognition, etc.). Staying within a class may result in better auditory comprehension scores.

Finally, the WAB uses a numerical criterion to separate the aphasic from the nonaphasic (AQ above 93.8 = not aphasic). The BDAE is less objective but, perhaps, more sensitive in its use of a severity rating scale and a profile of speech characteristics to determine who is and who is not aphasic.

Our experience with 45 patients indicates that the WAB and the BDAE show only 27 percent agreement in the way they classify the aphasias. Several reasons may account for the disagreement between the two measures, but the main source of discrepancy appears to be the WAB's tendency to classify most patients and the BDAE's tendency to classify less than 60 percent of the patients. If one seeks to classify a patient's aphasia, it is useful to have a measure that does that. On the other hand, if one seeks to classify a patient's aphasia, it is useful to have a measure that does that correctly. The WAB classifies more patients than the BDAE. Whether it does this accurately remains to be resolved. More importantly, if one is interested in classifying the aphasias, one should be able to do so accurately and reliably with either the BDAE or the WAB. Our data indicate that one cannot.

REFERENCES

Goodglass, H. and Kaplan, E., The Assessment of Aphasia and Related Disorders, Second Edition. Philadelphia: Lea & Feibiger, 1983.

Goodglass, H. and Kaplan, E., The Assessment of Aphasia and Related Disorders. Philadelphia: Lea & Feibiger, 1972.

Holland, A.L., Communicative Abilities in Daily Living. Baltimore: University Park Press, 1980.

Kertesz, A., Western Aphasia Battery. New York: Grune & Stratton, 1982.

Kertesz, A., Aphasia and Associated Disorders: Taxonomy, Localization, and Recovery. New York: Grune & Stratton, 1979.

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

Sanders, S.B. and Davis, G.A., A comparison of the Porch Index of Communicative Ability and the Western Aphasia Battery. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1978. Minneapolis, MN: BRK Publishers, 117-126, 1978.

Wertz, R.T., Keith, R.L., and Custer, D.D., Normal and aphasic behavior on a measure of auditory input and a measure of verbal output. Paper presented to the American Speech and Hearing Association, Chicago, Illinois, November, 1971.

Wertz, R.T., Kitselman, K.P., and Deal, L.A., Classifying the aphasias: Contributions to patient management. Paper presented to the Academy of Aphasia, London, Ontario, October, 1981.

DISCUSSION

Q: I wonder whether we take any of this--classification of patients or aphasic syndromes--too seriously. Are you saying we have an imprecise science and somebody other than ourselves has determined what the science is? Should we do something about all of this?

A: Some seem to take classification of patients into aphasic syndromes seriously. Others do not. The data we presented imply that the science of classifying patients is imprecise. The origins of the science are Neurology and Neuropsychology. If one takes classification seriously and one's ability to classify is imprecise, one might want to do something about it.

Q: What is a syndrome?

A: A group of signs and symptoms that indicate the presence of a specific condition or disorder.

Q: Do we have aphasia syndromes?

A: We have them. They exist in print. There are tests designed to detect their presence. The question is whether aphasic patients have them.

Q: Are they artifacts of tests?

A: They can be if one uses test performance to separate aphasic patients into syndromes. The point of our paper was that the BDAE and the WAB disagree on how they separate some patients into different syndromes. How you look for something probably influences what you find.

Q: If you exclude those patients that are unclassifiable on the BDAE but classifiable on the WAB, what is the agreement between the two tests?

A: Excluding the 28 patients, 62% of the sample, that were unclassifiable on the BDAE leaves 17 patients. Ten of these were classified the same on both tests, so agreement would be just under 60%.

C. We need to remember what our purpose is when we classify. A hundred years ago clinicians used classification to localize lesions. Today, we have some very sophisticated techniques--CAT, PETT, NMR--and no longer need the syndromes to localize lesions. We need to redefine why we use classification. As a neurologist, I can talk to a patient and decide he's got a Broca's aphasia. I may not know that from reading a two- or three-page report from a speech pathologist.

C: If you are suggesting that one of the present purposes of classification is a descriptive shorthand, fine. But, the results we presented indicate Goodglass and Kaplan's shorthand differs from that used by Kertesz.

C: I think Broca and Wernicke attempted to keep things simple. If you are going to separate boys from girls, you don't count kidneys and lungs and ribs, because those will lead you to believe that boys and girls are the same. You concentrate on differences. I think Broca and Wernicke looked for the differences. I wonder if we haven't, with our tests, started to count so many irrelevant-to-the-question-of-classification things, that we are amazed by most of our patients because they all have the same number of ribs.

C: I don't disagree with that, and I am not all that amazed by the similarities and the differences among my patients. I am amaged that two tests, designed to do the same thing, do that thing quite differently.