Prediction of Revised Token Test Overall, Subtest, and Linguistic Unit Scores by Two Shortened Versions

Joan C. Arvedson and Malcolm R. McNeil
University of Wisconsin, Madison, Wisconsin

Terri L. West
Whittier College, Whittier, California

Efficient yet valid and reliable assessment tools are at a premium to the clinical aphasiologist. This desire for efficiency without sacrificing validity and reliability prompted DiSimoni, Keith, Holt, and Darley (1975) to ask whether it was possible to make the Porch Index of Communicative Ability more efficient. In 1980, DiSimoni, Keith, and Darley reported the successful development of two shortened versions to predict overall PICA scores. They found that with the shortened versions there was little loss of information and a substantial decrease in test time, subject boredom, and fatigue.

Encouraged by their results and motivated by similar pragmatic concerns, we attempted to develop a shortened version of the Revised Token Test (RTT) (McNeil and Prescott, 1978) in order to predict the overall mean score, and if possible individual subtest and linguistic unit scores. The RTT is designed to assess the severity and nature of auditory verbal processing deficits and to monitor change over time. The standard test consists of ten subtests with ten equally difficult commands in each subtest. Performance on each element of each command is rated on a 15-point multidimensional scale. Administration time ranges from 13 to 75 minutes, averaging about 30 minutes. This time is considerable, given that the RTT assesses only one modality and a limited sample of auditory processing tasks. The possibility of obtaining predictive information with fewer test items in less time than that required for the standardized test seemed feasible, given that the RTT contains a large number of relatively homogeneous items. In order to test this hypothesis, a two-phase investigation was undertaken.

PHASE I

Method. In the first phase, multiple regression analyses were performed on the original standardization data of 30 adults with aphasia (McNeil and Prescott, 1978) in order to determine the plausibility of developing shortened forms. The regression analyses yielded many possible shortened forms, from which two were chosen.

Results. The first five items in each subtest correlated with the overall nearly perfectly ($r = 0.99$). Multiple regression further demonstrated that the mean for Subtest Six correlated more highly ($r = 0.95$) with the overall mean score than any other single subtest.

DiSimoni et al. (1975) had cautioned that one could not simply apply derived prediction equations to construct a shortened version of the PICA before the shortened versions were compared to and validated against the standard version. Thus, three forms of the RTT were prepared:

1) The STANDARD test consisted of ten items in all ten subtests;
2) The FIVE-ITEM test consisted of the first five items in each of the ten subtests;
3) The SUBTEST SIX version consisted of all ten items of Subtest Six.

PHASE II

Subjects. The second phase involved testing and comparing, objectively, the five item version, subtest six alone, and the standard test on a new subject sample. Twenty aphasic subjects participated in this study (Table 1).

Table 1. Biographical variables for twenty aphasic subjects with left hemisphere lesions.

Sex	Age	Months Post Onset	Aphasia Type	Lesion Site, Etiology, Speech/ Language and Motor Concomitants
M	39	7	Anomic	L parietal; vascular; hemiplegia.
M	70	1	Anomic	L hemisphere; vascular; hemiplegia.
M	72	2	Anomic	L hemisphere; vascular; no hemiplegia.
F	80	8	Wernicke	L temporal parietal; embolic; no hemiplegia.
F	74	2	Mixed	L frontal; vascular; hemiplegia
M	73	10	Anomic	L frontal parietal meningioma.
F	63	8	Anomic	L parietal aneurysm/subdural hematoma; post-surgical.
F	71	2	Anomic	L fronto-parietal; vascular; no hemiplegia.
M	67	6	Broca	L hemisphere; vascular; hemiplegia.
F	57	4	Wernicke	L posterior parietal; vascular; no hemiplegia.
F	70	1	Anomic	L hemisphere; vascular; hemiplegia.
F	72	13	Wernicke	L hemisphere; vascular; no hemiplegia.
F	66	2	Wernicke	L temporal, insular, and basal ganglia; vascular; no hemiplegia.
M	70	7	Mixed	L hemisphere; vascular; hemiplegia.
F	73	7	Anomic	L hemisphere; middle cerebral artery distribution; vascular; hemiplegia.
F	69	1	Mixed	L hemisphere; vascular; hemiplegia; dysarthria; apraxia of speech.
F	68	36	Mixed	L hemisphere; vascular; no hemiplegia; apraxia of speech.
F	21	2	Mixed	L temporal; glioma, craniotomy; hemiplegia; apraxia of speech.
F	82	240	Mixed	L hemisphere; vascular; hemiplegia.
F	74	2	Global	L parietal; vascular; no hemiplegia.

They ranged from mild to severely impaired and were heterogeneous by any classification system. Aphasia types were described on the basis of subjects' performance on the Boston Diagnostic Aphasia Examination or the Porch Index of Communicative Ability. All subjects were right-handed and had left hemisphere lesions as determined by at least two diagnostic measures (e.g., CT scan, EEG, Isotope scan, neurologic evaluation). Subjects were fourteen women and six men who ranged in age from 21 to 82, with a mean age of 66 years. Time post-onset ranged from one month to twenty years, with a mean of 18 months. Sixteen subjects were living in the community, three in nursing homes, and one was in a hospital at the time of testing. Two additional subjects were unable to complete the experiment for reasons unrelated to their ability to perform the tasks. The diagnosis of a left hemisphere lesion and aphasia without other perceptual or cognitive deficits, and passing the standardized RTT pretest were criteria for participation. Some subjects had an associated hemiparesis, dysarthria, or apraxia of speech.

Procedure. The standard RTT and the two shortened versions were administered in randomized order to all subjects, one form on each of three consecutive days at approximately the same time of day. All tests were administered and scored by an experienced and reliable RTT examiner (JCA). The alpha level was set at 0.05 for between-test differences.

Results. The results of the second phase of this investigation included Pearson Product Moment Correlation coefficients, which were computed between the standard and the two shortened test versions. Correlations were 0.98 for the comparison of the FIVE-ITEM version with the STANDARD overall (O.A.) mean. Because we were interested in making direct predictions from the two shortened versions to the standard version, and because large magnitude mean differences could exist in the presence of high correlations, we tested for significant mean differences (Table 2), with One-way Analysis of Variance with Repeated Measures, which revealed a significant difference ($F = 37.59$, $df = 2$, $p < 0.05$) between the STANDARD and the two shortened versions. Post-hoc analysis revealed no significant difference between the STANDARD and the FIVE-ITEM overall, but a significant difference between the STANDARD and SUBTEST SIX, and between the FIVE-ITEM version and SUBTEST SIX. No further discussion will be made of the latter, since we were interested in how each of those two versions could predict the STANDARD.

Table 2. Mean scores and standard deviations for Standard, Five-Item, and Subtest Six of the three RTT versions.

	Standard	Five-Item	Subtest Six
Mean score	12.33	12.28	11.19*
Standard Deviation	1.63	1.59	2.16

*Significant Difference from Standard, $p < 0.05$

Tests for linear relationships yielded the following predictive regression equations:

1. Predicted STANDARD mean score = 0.013 + (1.00 x FIVE-ITEM mean score).
2. Predicted STANDARD mean score = 4.230 + (0.724 x SUBTEST SIX mean score).

The FIVE-ITEM mean score predicted the STANDARD mean score almost perfectly. The SUBTEST SIX mean score underpredicted the STANDARD mean score. For the group, the linear regression equation from SUBTEST SIX predicted accurately. However, individual subjects showed greater differences from a perfect correlation than in the comparison of the FIVE-ITEM to STANDARD version. When the coefficient is close to one, the scores can be treated as simple difference scores.

We were also interested in the correlations for individual subtests among these test versions. Subtest correlation coefficients (Table 3) between the two short versions and the STANDARD revealed a larger range than for the overall test correlation. Correlations between the FIVE-ITEM and the STANDARD ranged from 0.83 to 0.94, with one exception. The Subtest IX correlation was 0.42 (that subtest contains adverbial clauses). As can be seen, correlations between the overall mean score of SUBTEST SIX and the STANDARD subtest means ranged from 0.65 to 0.97.

Table 3. Pearson Product Moment Correlation coefficients for Five-Item and Subtest Six correlated with the Standard.

	Standard										
Subtest	I	II	III	IV	V	VI	VII	VIII	IX	X	O.A.
Five-Item	0.89	0.84	0.83	0.89	0.94	0.91	0.94	0.90	0.42	0.89	0.98
Subtest Six	0.88	0.87	0.86	0.76	0.95	0.97	0.95	0.94	0.65	0.73	0.96

Linear regression equations were computed for predicting STANDARD individual Subtest mean scores from the FIVE-ITEM subtest scores (Table 4). Subtest V was the best predictor of the STANDARD version Subtest V score, with Subtest VII the next best.

Table 4. Means, Standard Deviations, and regression equations for subtests, and prediction of Standard Subtest Scores from Five-Item Subtest Scores.

	Standard		Five-Item			
Subtest	Mean	S.D.	Mean	S.D.	r	Regression equation
I	14.06	1.00	14.21	1.07	0.89	S = 2.21 + 0.834(5-Item)
II	13.39	1.25	13.27	1.30	0.84	S = 2.64 + 0.810(5-Item)
III	12.64	1.68	13.08	1.92	0.83	S = 3.15 + 0.726(5-Item)
IV	11.82	1.76	11.92	1.82	0.89	S = 1.60 + 0.857(5-Item)
V	11.89	2.38	12.12	2.25	0.94	S = 0.13 + 0.992(5-Item)
VI	11.66	2.27	11.09	2.47	0.91	S = 2.41 + 0.835(5-Item)
VII	11.94	2.42	11.92	2.46	0.94	S = 0.978 + 0.918(5-Item)
VIII	11.37	2.32	11.19	2.42	0.90	S = 1.73 + 0.861(5-Item)
IX	12.28	1.24	12.25	1.53	0.42	S = 8.12 + 0.340(5-Item)
X	12.23	1.61	11.71	1.88	0.89	S = 3.35 + 0.758(5-Item)
Overall	12.33	1.63	12.28	1.59	0.98	S = 0.013 + 1.00 (5-Item)

S = predicted Standard Score

Linguistic unit correlations between these versions are shown (Table 5). The FIVE-ITEM to STANDARD correlations ranged from 0.75 to 0.98. Lower correlation coefficients occurred for the adverbial clause than for other parts of speech. Correlations for SUBTEST SIX to STANDARD ranged from 0.89 to 0.96. It should be noted, however, that SUBTEST SIX does not contain all the linguistic units found throughout the entire STANDARD ten subtest version. Therefore, regression equations were not developed to predict STANDARD linguistic unit scores from SUBTEST SIX.

Table 5. Correlation coefficients for linguistic units in Five-Item and Subtest Six correlated with the Standard.

	Standard										
	Vb.	Size	Color	Shape	Impl	Size	Color	Shape	Place	Lf/Rt	Adv.
Five-Item	0.96	0.94	0.97	0.96	0.85	0.93	0.98	0.97	0.94	0.95	0.75
Subtest Six	0.94	0.89	0.93	0.95	--	0.93	0.96	0.93	0.90	--	--

Linear regression equations for prediction of STANDARD linguistic unit mean scores from the FIVE-ITEM version were developed (Table 6). Several of the regression equations contain coefficients close to one, with all but three at 0.9+.

Table 6. Means, Standard Deviations, and regression equations for linguistic units, and prediction of Standard scores from Five-Item linguistic unit scores.

Linguistic Unit	Standard Mean	Standard S.D.	Five-Item Mean	Five-Item S.D.	r	Regression equation
Direct Verb	12.98	1.25	12.91	1.21	0.96	S = 0.229 + 0.988(5-Item)
Size I Adj.	12.38	1.63	12.01	1.67	0.94	S = 1.34 + 0.919(5-Item)
Color I Adj.	12.35	1.64	12.39	1.56	0.97	S = 0.326 + 1.02 (5-Item)
Shape I Noun	12.18	1.69	11.98	1.69	0.96	S = 0.590 + 0.962(5-Item)
Implied Verb	12.63	1.32	12.71	1.54	0.85	S = 3.33 + 0.732(5-Item)
Size II Adj.	11.56	2.30	11.25	2.52	0.93	S = 1.96 + 0.854(5-Item)
Color II Adj.	11.63	2.40	11.59	2.48	0.98	S = 0.668 + 0.946(5-Item)
Shape II Noun	11.50	2.28	11.69	2.37	0.97	S = 0.572 + 0.935(5-Item)
Place Preposition	10.82	2.83	11.31	2.81	0.94	S = 0.082 + 0.949(5-Item)
Left-Right	10.48	3.26	10.33	3.27	0.95	S = 0.724 + 0.945(5-Item)
Adverbial Clause	12.79	1.23	12.49	1.51	0.75	S = 5.11 + 0.614(5-Item)

Administration time, in minutes, was reduced by approximately half with the FIVE-ITEM test (X = 17.5, S.D. = 5.6) compared with the STANDARD ($\bar{X}$ = 32.0, S.D. = 10.5). SUBTEST SIX required about 15% ($\bar{X}$ = 4.5, S.D. = 2.9) of the STANDARD test time.

DISCUSSION

The Revised Token Test was designed with ten homogeneous items within each of the ten subtests. This was done for two reasons. First, to increase the reliability of the observations, and second, to provide a format for capturing patterns of auditory processing disturbance as described by Schuell, Jenkins, and Jimenez-Pabon (1964), Porch (1967), Brookshire (1974), and others. Extensive research on these patterns over several years by McNeil and Hageman (1979) and others has led to the belief that the patterns do not exist as they were previously envisioned, and that only one pattern is reliably produced; that is, the intermittent one. With the abandonment of pattern quantification, it became reasonable to attempt to reduce the number of items per subtest.

The results confirmed our previous finding that the RTT is highly redundant. A short form of the RTT can provide accurate and useful data to predict STANDARD RTT mean scores. SUBTEST SIX underpredicts the overall mean score of the STANDARD version, and is limited in the repertoire of linguistic units compared with the STANDARD. Linear regression equations showed sufficient individual variability so that prediction of a STANDARD mean overall score is not recommended from SUBTEST SIX. The FIVE-ITEM version offered here can accurately predict the STANDARD mean overall score. In addition, subtest scores and linguistic unit scores (with the exception of Subtest IX) show sufficiently high correlations that the FIVE-ITEM version is recommended as a substitute for the STANDARD RTT. In fact, the FIVE-ITEM mean overall scores are so close to the STANDARD RTT that a direct substitution can be made. A regression equation is not needed to predict the overall mean score of the STANDARD version. With only twenty subjects who showed variability in scores on individual subtests and linguistic units, it is with caution that we recommend individual subtest and linguistic unit predictors. This FIVE-ITEM version reduces test time to about half, a fact that is not trivial in terms of patient energy or finances expended. The time required for scoring and profiling test results is also reduced accordingly.

ACKNOWLEDGMENT

The authors would like to thank the following institutions in Madison, Wisconsin, for assistance in acquiring subjects for this study: Visiting Nurse Service, Methodist Hospital, Methodist Health Center, Oakwood Retirement Center (Hebron Hall), and Madison General Hospital.

REFERENCES

Brookshire, R.H. Differences in responding to auditory verbal materials among aphasic patients. Acta Symbolica, 1, 1-18, 1974.

DiSimoni, F.G., Keith, R.L., Holt, D.C., and Darley, F.L. Practicality of Shortening the Porch Index of Communicative Ability. Journal of Speech and Hearing Research, 18, 491-497, 1975.

DiSimoni, F.G., Keith, R.L., and Darley, F.L. Prediction of PICA overall score by short versions of the test. Journal of Speech and Hearing Research, 23, 511-516, 1980.

McNeil, M.R. and Hageman, C.F. Auditory processing deficits in aphasia evidenced on the Revised Token Test: Incidence and prediction of across subtest and across item within subtest patterns. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1979. Minneapolis, MN: BRK Publishers, 1979.

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

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

Schuell, H., Jenkins, J.J., and Jimenez-Pabon, E. Aphasia in Adults: Diagnosis, Prognosis, and Treatment. New York: Hoeber Medical Division, Harper and Row Publishers, 1964.

DISCUSSION

Q: On your predictions from multiple regression, you are suggesting that you can predict the overall scores from these shortened tests, but that was taking the data from the original sample. Are you planning to do the next step and apply the multiple regression formula to a new sample to see if your predictions still hold up?

A: That's what we did. The second phase of this study was done with 20 subjects, who were a new sample. We gave each one of these subjects three versions, with each version on a different day.

Q: Did you do an item analysis to see how different the first item, for instance, is from the fifth item on each section?

A: That was done in the initial phase and there was no difference in item by item analysis. When the test was originally published, we did an item analysis, and there was no significant difference among any of the items in any of the subtests, except subtest six. There was an item in there that we couldn't account for and we attributed it to a chance goof-up, something funny in subtest six. The rest of them were homogeneous.

Q: Why did you pick subtest six?

A: Because it showed the highest correlation with the overall mean score of any of the individual subtests in the original regression analysis done on the original standardization sample.

Q: Do you feel like there is information you might lose by giving the five-item version as opposed to a ten-item version?

A: Not with the Revised Token Test, because there is so much redundancy in this test and a limited sample of behaviors that occur with the Revised Token Test. We do not feel that we are going to lose information with the five-item version. It would be very nice if others replicated that with more subjects.