

Validity of Multi-sentence Reading Comprehension  
Subtests in Aphasia Tests

L.E. Nicholas, D.L. MacLennan and R.H. Brookshire  
Veterans Administration Medical Center, Minneapolis, Minnesota

The idea for this study came from our subjective impression that many items in aphasia tests which were designed to assess multi-sentence reading comprehension could be answered correctly without reading the related passages first. This subjective impression probably is one that has been shared by many of you. When we searched the literature on testing reading comprehension of non-brain-damaged individuals, we found a number of studies from the 1960's and 1970's that addressed this issue. A quote from Tuinman (1972-73) summarizes this issue quite nicely.

"Reading comprehension was defined by Smith (1971) as a reduction in uncertainty. That is, a reader brings to a passage a certain amount of linguistic and conceptual information. This information is incomplete, however, and reading the passage reduces this incompleteness and thus reduces uncertainty. Tests of a reader's ability to comprehend printed material, therefore, should be tests of his ability to use the material to reduce uncertainty. This makes it necessary to assume that, prior to reading, some uncertainty exists. If a question about a reading passage can be answered by someone who has not read the passage, the question can be said to be entirely passage independent. It has been shown repeatedly that current standardized tests contain many questions with a high degree of passage independency. Therefore, when a student obtains a particular test score, one can not be certain to what extent that score is a valid measure of his ability to read those passages in the test."

In a 1974 study, Sidman asked 15 non-brain-damaged adults to answer questions from the reading comprehension subtest of Eisenson's Examining for Aphasia without reading the related paragraphs. In spite of the fact that none of the subjects had read the paragraphs, 17 of the 20 questions were answered correctly by at least ten of the 15 subjects. Since the 1970's several authors of standardized reading comprehension tests intended for the general population have taken the passage dependency of test items into account when developing new tests. It obviously is not feasible to make every reading comprehension test item entirely passage dependent, but these authors have tried to maximize the number of passage dependent questions to increase the validity of their tests.

In looking at aphasia test manuals, we were unable to find any reference to the issue of the passage dependency of test items in the multi-sentence reading comprehension subtests. Therefore, we designed this study to address these five questions: (1) Which multi-sentence reading subtest items are responded to correctly by a significant ( $p \leq .05$ ) number of aphasic and non-brain-damaged subjects when related passages are not available? (2) What are the overall passage dependency ratings for various multi-sentence reading subtests? (3) Do aphasic and non-brain-damaged subjects perform in a similar way on subtest items when related passages are not available? (4) Does the performance of aphasic subjects in this study depend upon their paragraph

Table 1. Descriptive information and Nelson Reading Test scores for aphasic subjects.

Subject	Age	Education	BDAE Severity	Nelson Vocab.* (n = 100)	Nelson Paragraph* (n = 75)	Nelson Total* (n = 175)
<b>Nonfluent</b>						
1.	51	14	3	56(8.3)	39(7.5)	95(7.9)
2.	59	18	1	34(5.6)	19(3.8)	53(4.8)
3.	51	12	3	49(7.1)	30(5.6)	79(6.4)
4.	58	12	4	53(7.8)	38(7.2)	91(7.4)
<b>Fluent-mixed</b>						
1.	64	12	5	49(7.1)	34(6.1)	83(6.7)
2.	59	9	5	27(4.8)	15(3.1)	42(4.1)
3.	65	16	5	92(10.5+)	45(8.9)	137(10.5+)
4.	70	12	2	31(5.0)	9(2.2)	40(3.9)
<b>Fluent-anomic</b>						
1.	73	16	5	56(8.3)	42(8.2)	98(8.3)
2.	61	12	5	62(9.7)	53(10.3)	115(9.7)
3.	73	17	5	61(9.1)	36(6.5)	97(8.2)
4.	64	17	5	93(10.5+)	49(9.5)	142(10.5+)
Mean	62.3	13.9	4	55.3(7.8)	34.1(6.6)	89.3(7.4)
Range	51-73	9-18	1-5	5.0-10.5+	2.2-10.3	3.9-10.5+

\*Nelson Reading Test scores are presented as number correct with grade equivalent in parentheses. The highest possible grade equivalent is 10.5.

Table 2. Age, education and Nelson Reading Test scores (and grade equivalents) for non-brain damaged and aphasic subjects.

	Age	Education	Nelson Vocab.* (n = 100)	Nelson Paragraph* (n = 75)	Nelson Total* (n = 175)
<b>NBD</b>					
Mean	61.3	13.9	89.0 (10.5+)	54.4 (9.2)	143.4 (10.4)
Range	55-72	10-18	78-96 (10.5+-10.5+)	30-73 (5.6-10.5+)	116-168 (9.8-10.5+)
<b>APH</b>					
Mean	62.3	13.9	55.3 (7.8)	34.1 (6.6)	89.3 (7.4)
Range	51-73	9-18	27-93 (5.0-10.5+)	9-53 (2.2-10.3)	40-137 (3.9-10.5+)

\*Nelson Reading Test scores are presented as number correct with the grade equivalent in parentheses. The highest possible grade equivalent is 10.5.

reading level? (5) Does the performance of aphasic subjects in this study depend upon the type of aphasia which they exhibit?

## METHOD

Subjects. Subjects were twelve aphasic adults and twelve non-brain-damaged adults. Mean age and education were essentially equivalent for the two groups. Aphasic subjects were independently classified by two of the three investigators into groups representing fluent, nonfluent, and mixed aphasia based on analysis of a taped conversational speech sample and a picture description. The severity of each subject's aphasia also was determined in a similar manner using the subjective rating scale from the Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1973). Aphasic subjects were at least one year post onset of their aphasia and had not been tested within the previous three months with the multi-sentence reading comprehension subtests investigated in this study. All subjects also were tested with the vocabulary and paragraph subtests of the Nelson Reading Test (Nelson, 1962) administered in standard fashion. Descriptive information and test scores for subjects are presented in Tables 1 and 2.

### Stimulus Materials and Procedures

Seventy-two test items designed to assess aphasic adults' reading comprehension of multi-sentence passages were administered to each subject. These items represented all the multi-sentence reading items from the following tests: Boston Diagnostic Aphasia Examination (BDAE) (Goodglass and Kaplan, 1983), Minnesota Test for the Differential Diagnosis of Aphasia (MTDDA) (Schuell, 1965), Examining for Aphasia (EFA) (Eisenson, 1954), Reading Comprehension Battery for Aphasia (RCBA) (LaPointe and Horner, 1979), and Western Aphasia Battery (WAB) (Kertesz, 1982). Each of these tests contains one multi-sentence reading comprehension subtest, except for the RCBA, which contains two. Subtest VII of the RCBA contains two-sentence passages and will be referred to as the RCBA short. Subtests VIII and IX of the RCBA contain approximately 50-word passages and will be referred to as the RCBA long.

All subtests were presented first in a "without passage" condition in which only the test items for each reading passage were presented. Immediately afterward, subjects received a "with passage" condition in which both the reading passages and the test items for each subtest were presented. (Examples of without passage and with passage items are presented in the Appendix.) The items from each test were presented sequentially as they appear in the original test and were presented as high quality photocopies of the original material. Half the subjects received the subtests in one randomly determined order and half received the subtests in the reverse order. All responses required pointing to the correct answer from the choices given in printed material. Subjects were instructed to read all material silently.

## RESULTS

Table 3 shows which multi-sentence reading subtest items were responded to correctly by a significant ( $p \leq .05$ ) number of aphasic and non-brain-damaged subjects when related passages were not available. There are some interesting response patterns displayed in this table. There are 33 instances in which subjects as a group responded with equal accuracy in both without passage and with passage conditions. In addition, there are 11 instances in which subjects as a group responded with greater accuracy in the without passage condition than in the with passage condition. Table 4 summarizes the data from Table 3.

Table 3. Number of correct responses for all test items with and without reading of passages by aphasic and non-brain-damaged subjects (n = 12 per item per condition).

	BDAE				MTDDA				EFA					
	NBD		APH		NBD		APH		NBD		APH			
	W/O	With	W/O	With	W/O	With	W/O	With	W/O	With	W/O	With		
1.	3	12	3	12	1.	6	10	4	7	1.	0	12	0	11
2.	12*	12	11*	11	2.	2	11	3	12	2.	4	12	3	11
3.	12*	12	12*	12	3.	7	6	8	4	3.	10*	11	10*	12
4.	12*	12	8*	7	4.	4	10	5	10	4.	10*	12	9*	12
5.	7*	10	7*	9	5.	11*	9	9*	6	5.	0	10	0	8
6.	8*	12	5	9	6.	5	11	8	12	6.	10*	11	6*	9
7.	9*	11	10*	12	7.	9*	11	6	8	7.	9*	10	6*	9
8.	11*	11	9*	6	8.	7	10	4	10	8.	11*	12	8*	9
										9.	12*	12	11*	11
										10.	9*	11	8*	9
										11.	6*	11	6*	9
										12.	0	9	2	9
										13.	9*	10	9*	9
										14.	7*	10	10*	7
										15.	9*	9	9*	9
										16.	2	12	3	11
										17.	12*	11	7*	10
										18.	3	10	8*	12
										19.	11*	12	7*	12
										20.	4	8	4	6

	RCBA (long)				RCBA (short)				WAB					
	NBD		APH		NBD		APH		NBD		APH			
	W/O	With	W/O	With	W/O	With	W/O	With	W/O	With	W/O	With		
1.	7*	12	6	12	1.	0	9	1	7	1.	8*	12	4	12
2.	7*	12	4	11	2.	8*	10	6	11	2.	12*	12	10*	12
3.	11*	12	11*	12	3.	2	9	2	7	3.	11*	12	10*	10
4.	5	11	6	11	4.	2	10	1	8	4.	7*	12	5	12
5.	5	12	10*	12	5.	6	11	10*	8	5.	12*	12	9*	11
6.	2	12	3	12	6.	6	12	5	11	6.	9*	11	7*	12
7.	12*	10	9*	11	7.	12*	12	11*	11					
8.	3	12	4	9	8.	12*	12	12*	12					
9.	12*	12	9*	9	9.	5	10	7*	11					
10.	12*	12	12*	11	10.	1	12	1	10					
11.	12*	12	12*	12										
12.	12*	12	11*	11										
13.	12*	12	9*	12										
14.	12*	12	11*	12										
15.	12*	12	9*	12										
16.	10*	12	12*	12										
17.	9*	11	4	9										
18.	12*	12	12*	12										
19.	12*	12	12*	12										
20.	2	12	3	12										

\* = items correctly answered significantly better than chance ( $p \leq .05$ ). The number of correct responses necessary to achieve this level varies across subtests depending upon the number of choices available for each item.

Table 4. Number of items which were correctly answered significantly better than chance ( $p \leq .05$ ) without reading passages.

Subtest	# of Items	# Better Than Chance	
		NBD	APH
BDAE	8	7	6
MTDDA	8	2	1
EFA	20	13	14
RCBA (LONG)	20	15	13
RCBA (SHORT)	10	3	4
WAB	6	6	4
Total	72	46	42

Except for the MTDDA and the RCBA short, more than half the items from these subtests were responded to correctly by a significant number of subjects in both groups without reading related passages.

Table 5 presents the proportion correct and the passage dependency indices (PDI) for each subtest for both groups of subjects. The PDI formula was chosen by Tuinman (1972-73) as the best measure of passage dependency from several formulas which he developed. The passage dependency index provides an overall measure of how subjects perform without reading related passages compared with their performance when reading passages are available. If subjects had zero correct without passages and 100 percent correct with passages, the PDI would be 1.00. There is no absolute value for a PDI that makes it acceptable, but values around .50 appear to be considered adequate. The values listed on this table for the Nelson Reading Skills Test were considered adequate by the authors of that test. Only the RCBA short comes close to these values. The MTDDA, which had a low proportion correct without passages, has a poor PDI because error rates for subjects were so high in the with passage condition that the differences between the with and without passage conditions are small. Chance probabilities of correct responses are also presented in Table 4 for each subtest. Both groups of subjects performed well beyond chance levels without reading related passages on all subtests except the MTDDA.

Table 6 reports the overall number of correct responses with and without passages and also shows the difference between the number of correct responses in these two conditions for each subject. Non-brain-damaged subjects correctly answered an average of 46 of the 72 test items in the without passage condition. Aphasic subjects correctly answered an average of 42 of the 72 test items in the without passage condition. All of the non-brain-damaged subjects and ten of the 12 aphasic subjects answered more than half the test items accurately without reading the related passages. A one way analysis of variance was performed on the difference scores for both subject groups. There was no significant group effect, suggesting that non-brain-damaged and aphasic subjects responded in similar fashion to the presence or absence of passages when they answered the test items. Additionally, of the test items answered correctly without passages, 38 of the same items were answered correctly by a significant number of subjects in both groups. This suggests that non-brain-damaged and aphasic subjects have a strong tendency to correctly answer the same items without access to the related passages.

To determine whether the performance of aphasic subjects depended upon their paragraph reading level as measured by the Nelson Paragraph Reading Test, a one way analysis of variance was performed on the difference scores for the high level and low level readers. There was no significant group effect suggesting that both high level and low level aphasic readers responded in similar fashion to the presence or absence of passages when they answered test items.

And finally, to determine whether the performance of aphasic subjects depended upon the type of aphasia which they exhibited, a one way analysis of variance was performed on the difference scores for nonfluent, fluent, and mixed aphasic subjects. There was no significant group effect, suggesting that all three groups of aphasic subjects responded in similar fashion to the presence or absence of passages when they answered test items.

In summary, these results show that more than half the test items in the multi-sentence reading comprehension subtests of aphasia tests are not suitable for assessing multi-sentence reading ability of non-brain-damaged or aphasic readers, because they have low passage dependency. Some test items can be

Table 5. Proportion correct and passage dependency indices (PDI) for non-brain-damaged and aphasic subjects on reading subtests (PDI = 1 - without/with: The higher the PDI the better the validity of the test).

	APH (n = 12)				NBD (n = 12)		
	Chance	$\bar{X}$ Prop Correct		PDI	$\bar{X}$ Prop Correct		PDI
		W/O	With		W/O	With	
BDAE	.25	.68	.81	.16	.77	.96	.20
MTDDA	.50	.49	.72	.32	.53	.81	.35
EFA	.25	.53	.81	.35	.58	.90	.36
RCBA (long)	.33	.70	.94	.25	.75	.98	.23
RCBA (short)	.33	.47	.80	.41	.45	.89	.49
WAB	.25	.63	.96	.34	.82	.99	.17

Nelson Reading Skills Test (1977)

Level B - Form 3 - PDI = .49

Level B - Form 4 - PDI = .45

(These PDIs were established using 4-6th grade readers.  
Lower PDIs might be expected with older readers.)

Table 6. Number of correct responses with and without reading of passages for aphasic and non-brain-damaged subjects (n= 72 per subject per condition).

Subject	Without	With	Difference
<b>Non-brain-damaged</b>			
1.	48	70	22
2.	49	71	22
3.	45	58	13
4.	40	59	19
5.	50	68	18
6.	47	68	21
7.	44	65	21
8.	46	61	15
9.	46	69	23
10.	46	72	26
11.	51	72	21
12.	45	66	21
$\bar{X}$	46.4	66.6	20.2
Range	40-51	58-72	13-26
<b>Aphasic</b>			
1.	41	65	24
2.	28	50	22
3.	44	58	14
4.	35	57	22
5.	43	62	19
6.	45	67	22
7.	49	68	19
8.	41	53	12
9.	46	62	16
10.	43	59	16
11.	45	65	20
12.	47	67	20
$\bar{X}$	42.3	61.1	18.8
Range	28-49	50-68	12-24

Subjects' numerical designations are the same as those in Tables 1 and 2.

answered correctly without the related passage because information contained in the passage is common knowledge. Other test items can be answered correctly because the items themselves identify the topic of the missing passage and answers can be inferred by eliminating choices that either do not pertain to the topic or conflict with information from previous test items. The two subtests which have the highest number of passage dependent items, the MTDDA and the RCBA short, sample a very limited range of reading comprehension performance. The MTDDA requires the reading of a single, very difficult 120-word passage, on which both groups of subjects made many errors even when the paragraph was available. The RCBA short was designed to assess the reading of two-sentence passages.

It would appear that to validly assess multi-sentence reading comprehension of aphasic adults, clinicians must use materials other than those contained in currently available aphasia tests. The Nelson Reading Skills Test, with its five passages at each of three difficulty levels and its relatively high passage dependency, would appear to be a good candidate. We are currently assessing the passage dependency of items in this test, for both non-brain-damaged and aphasic adults. This test also divides test items into three categories which represent three levels of reading performance--comprehension of literal information, translational or low-level inferential information, and higher level inferential information. At this time, we are evaluating the usefulness of this information in assessing the multi-sentence reading comprehension of aphasic adults.

#### REFERENCES

- Eisenson, J. Examining for Aphasia. New York: Psychological Corporation, 1954.
- Goodglass, H. and Kaplan, E. Assessment of Aphasia and Related Disorders. Philadelphia: Lea and Febiger, 1983.
- Hanna, G., Schell, L.M. and Schreiner, R. The Nelson Reading Skills Test. Chicago: Riverside Publishing Co., 1977.
- Kertesz, A. Western Aphasia Battery. New York: Grune and Stratton, 1982.
- LaPointe, L.L. and Horner, J. Reading Comprehension Battery for Aphasia. Tigard: C.C. Publications, 1979.
- Nelson, M.J. The Nelson Reading Test. Boston: Houghton-Mifflin Co., 1962.
- Schuell, H. Minnesota Test for the Differential Diagnosis of Aphasia. Minneapolis: University of Minnesota Press, 1965.
- Sidman, M. Testing for reading comprehension: A brief report on stimulus control. Journal of Applied Behavior Analysis, 7, 327-332, 1974.
- Tuinman, J.J. Inspection of passages as a function of passage-dependency of the test items. Journal of Reading Behavior, 5, 186-191, 1972-73.
- Tuinman, J.J. Determining the passage dependency of comprehension questions in 5 major tests. Reading Research Quarterly, 2, 206-223, 1973-74.

#### APPENDIX

Examples of without passage and with passage items

##### BOSTON DIAGNOSTIC APHASIA EXAMINATION

##### Without Passage

The aim of Civil Service classification is to . . .

- Achieve higher salaries
- Establish favoritism
- Effect a reduction in taxes
- Assure equal pay for equal work

With Passage

Favoritism used to be the rule in Civil Service, and many jobs paid more than they were worth. Civil Service reform has resulted in classifying positions according to their duties and responsibilities. The aim of Civil Service classification is to . . .

- Achieve higher salaries
- Establish favoritism
- Effect a reduction in taxes
- Assure equal pay for equal work

READING COMPREHENSION BATTERY FOR APHASIA - LONG

Without Passage

Soft goose feathers are called  
down up pillows

The material is used to fill  
weather sleeping bags time

A down-filled coat would be best in  
summer winter cars

Goose down equipment can be used for  
swimming music camping

With Passage

The soft and light feathers from the underside of a goose have so many uses. This material is called down, and it is used to fill some pillows. Goose down is used to fill sleeping bags and cold weather clothes, also. That is why you see so many naked geese walking around.

Soft goose feathers are called  
down up pillows

This material is used to fill  
weather sleeping bags time

A down-filled coat would be best in  
summer winter cars

Goose down equipment can be used for  
swimming music camping

DISCUSSION

- Q: What is the difference in behavior between the first 3 subjects in the fluent mixed category that had the aphasia severity ratings of 5 and all of the subjects in the fluent anomic category that had severity ratings of 5?
- A: The fluent mixed subjects had literal paraphasias in their conversational speech and their picture descriptions and the fluent anomic subjects did not. The fluent mixed aphasic subjects also had repetition problems and could be classified as exhibiting conduction aphasia.
- Q: Were there differences in auditory comprehension between those 2 groups?
- A: No, they all had good auditory comprehension, as you would expect from anomic and conduction aphasic patients.



- Q: Looking at your BDAE scores, except for about 4 subjects, this was a fairly mild group on that rating. Surprisingly, subject number 2 read better than some of the milder patients on the BDAE.
- A: The BDAE subjective rating scale score is mostly an expressive rating, so it is not a very good indicator of comprehension performance or the overall severity of aphasia. We did not have PICA or WAB overall scores for these subjects.
- Q: Would you care to speculate on whether more severely involved folks would be less able to respond correctly without reading the passages than milder folks?
- A: I suspect that they would be poorer but the subjects you are most interested in using these multi-sentence reading subtests items on tend to be higher level subjects because reading is an issue with them, whereas with lower level subjects you generally aren't working on that skill and aren't as interested in assessing it or looking at changes over time with it. But I suspect that as there is more brain damage there may be more cognitive kinds of problems so that they might not perform as well without passages.
- Q: I think that it is true that if you are using yes/no questions or multiple choice questions that subjects in fact don't even have to read the questions to answer with a certain probability. Did you make any attempt to separate out the effects of guessing randomly from the effects of deriving information from the questions without paragraph information?
- A: If subjects were guessing, they should have been at chance level. Since our results were significantly beyond chance level then I don't see that that is an issue. In Table 5 chance levels for the tests are given in the second column and in Table 3 those significance levels do take into account the chance probability of being correct. These were all significant at the .05 level beyond chance, which varied from subtest to subtest depending upon the number of foils.