

C H A P T E R

10

**The Boston Naming Test:
Revised Administration and
Scoring Procedures and
Normative Information for
Non-Brain-Damaged Adults**

Linda E. Nicholas
Robert H. Brookshire
Donald L. MacLennan
James G. Schumacher
Shirley A. Porrazzo

The 60-item Boston Naming Test (BNT) (Kaplan, Goodglass, and Weintraub, 1983) is a confrontation naming test frequently used in our clinic. It is a fairly informative measure of the word retrieval performance of brain-damaged patients. The normative information published by Van Gorp, Satz, Kiersch, and Henry (1986) expanded the usefulness of this measure by providing quantitative information on the BNT performance of 78 non-brain-damaged adults between the ages of 59 and 95 years. The norms reported by Kaplan and colleagues were only for subjects up to the age of 59, and these subjects had actually been tested with an earlier 85-item version of the BNT. Their scores were then reanalyzed, looking only at the 60 items contained in the 1983 version. No information is available about item-by-item error rates or about the kinds of incorrect responses non-brain-damaged subjects give on this test. We also felt that several aspects of the test's administration and scoring procedures were not satisfactory.

First, we felt that we were not always administering the test and scoring patient's responses in the same way, either from examiner to examiner or from test to test by the same examiner. The administration and response scoring instruction provided with the test were not detailed enough to help resolve questions. No interjudge or intrajudge reliability data have been reported for the BNT.

Second, some test pictures frequently elicited responses that seemed reasonable names for the pictures but were not considered correct for test purposes. We wondered if synonyms for some words might be given often enough by non-brain-damaged adults for them to be considered normal responses to those pictures.

Third, some of the stimulus pictures were either visually confusing or too difficult even for many non-brain-damaged adults to name correctly. We wondered if there were items that would be missed by a large percentage of non-brain-damaged adults and if those items would occur only toward the end of the test.

Fourth, we wondered if we might learn more about our patients' word retrieval abilities if they were prompted to produce another response following certain types of incorrect responses. To address these questions and concerns, we did the following:

1. Developed standard administration and response coding and scoring procedures for the BNT.
2. Developed standard prompts to be given following certain types of incorrect responses.
3. Tested 60 healthy non-brain-damaged adults between the ages of 40 and 78 years (mean = 56.1, SD = 9.27).
4. Assessed the intrajudge and interjudge reliability of our multiple-category coding and scoring procedures.

5. Calculated item-by-item error rates.
6. Determined the types of incorrect responses that subjects gave.
7. Analyzed how subjects responded when prompts were given following incorrect responses.
8. Revised our prompts and response coding and scoring procedures based on the results of this study.
9. Assessed the reliability of our revised coding and scoring procedures.

The final revisions of our administration and response coding and scoring procedures are in Appendices A and B. The original coding procedure allowed us to differentiate among 10 types of patient responses with designated examiner prompts to follow most responses. The revised version contains only eight categories of patient responses. The reason for these changes will be explained later.

NORMATIVE INFORMATION

Our normative data generally are consistent with those by Van Gorp and colleagues (1986). The mean number of items correct for the 78 adults in the Van Gorp study was 54.31, and the mean for our 60 adults was 54.50 (SD = 3.52, range = 40-59). Van Gorp and his colleagues did not find a significant correlation between BNT total score and either age or education, although they noted greater variability in the naming performance of the older age groups. The correlation between our subjects' BNT total score and age was not significant either. The correlation between total score and education was significant ($p \leq .01$) but weak ($r = .3$), accounting for only 9 percent of the variance. Despite the lack of a significant correlation between total score and age, Van Gorp and his colleagues reported standard deviations and cutoff scores for unimpaired performance by age group. We felt that this weakened the value of their cutoff scores. Cutoff scores for unimpaired performance in the Van Gorp study ranged from 38 for the oldest group to 51 for the youngest group. The cutoff score for unimpaired performance (two standard deviations below the mean) based on our 60 subjects was 48.

ITEM-BY-ITEM ERROR RATE

Overall, 91 percent of test items were named correctly. Only 20 of the 60 test items were named correctly by all 60 subjects (Table 10-1). These

TABLE 10-1. NUMBER OF INCORRECT RESPONSES TO BOSTON NAMING TEST ITEMS

<i>Item</i>	<i>Number incorrect</i>	<i>Item</i>	<i>Number incorrect</i>
1. Bed	0	31. Rhinoceros	10*
2. Tree	0	32. Acorn	4
3. Pencil	0	33. Igloo	1
4. House	6*	34. Stilts	0
5. Whistle	0	35. Dominoes	4
6. Scissors	1	36. Cactus	0
7. Comb	0	37. Escalator	0
8. Flower	1	38. Harp	0
9. Saw	0	39. Hammock	1
10. Toothbrush	1	40. Knocker	6*
11. Helicopter	2	41. Pelican	6*
12. Broom	0	42. Stethoscope	2
13. Octopus	0	43. Pyramid	3
14. Mushroom	4	44. Muzzle	7*
15. Hanger	0	45. Unicorn	7*
16. Wheelchair	0	46. Funnel	3
17. Camel	1	47. Accordion	5
18. Mask	0	48. Noose	8*
19. Pretzel	0	49. Asparagus	4
20. Bench	0	50. Compass	32*
21. Racquet	0	51. Latch	26*
22. Snail	0	52. Tripod	11*
23. Volcano	3	53. Scroll	6*
24. Seahorse	5	54. Tongs	11*
25. Dart	1	55. Sphinx	12*
26. Canoe	1	56. Yoke	16*
27. Globe	5	57. Trellis	7*
28. Wreath	0	58. Palette	18*
29. Beaver	11*	59. Protractor	39*
30. Harmonica	8*	60. Abacus	27*

* Items named incorrectly by 10 percent or more of subjects.

were scattered throughout the first 40 items. Another 20 were named incorrectly by less than 10 percent of our subjects. These were scattered throughout the first 50 test items. The remaining 20 test items were named incorrectly 10 percent or more of our subjects. These items are asterisked in Table 10-1. Ten errors were in the first 50 items, and 10 were in the final 10 items on the test. The increase in errors across the test was not regular, but the largest number of high error items were concentrated at the end of the test.

TYPES OF INCORRECT RESPONSES

Half of all incorrect responses were related names (Table 10-2). Related Name and Don't Know responses accounted for nearly 80 percent of incorrect responses. Of 3,600 responses given by our subjects, only 4 were initially coded as Visual Misperception and 3 were coded as Wrong Part. This would seem to speak well for the clarity of the drawings. However, we feel that some Related Name responses also might have been due to lack of clarity in the drawings or to subjects responding before they had looked carefully at the details of the drawings (such as saying *schoolhouse* or *apartment building* for *house*, *tripod* for *easel*, *hippopotamus* for *rhinoceros*).

FREQUENTLY GIVEN OTHER NAMES FOR ITEMS

Ten items were called by a consistent other name by 10 percent or more of our subjects (Table 10-3). Only one was a true synonym (*mouth organ* for *harmonica*). Three were merely less precise terms than the one specified in the test, and six were outright errors. Examiners should be aware that such responses may be given commonly by adults who have no history of language or cognitive impairment. If this test were to be revised, we feel that confusing items such as *compass* and *protractor* should be eliminated and other currently incorrect responses should be considered acceptable, such as *mouth organ* for *harmonica*, *lock* for *latch*. The latter response was given by one-third of our subjects.

TABLE 10-2. TYPES OF INCORRECT RESPONSES ON THE BOSTON NAMING TEST

<i>Type</i>	<i>Number</i>	<i>Percent</i>
Total incorrect	326	9.1
Related name	164	50.3
Unrelated name	16	4.9
Multiple attempts	10	3.1
Off task	29	8.9
Visual misperception	3	0.9
Wrong part	3	0.9
Mispronunciation	5	1.5
Don't know	94	28.8
No response	2	0.6

TABLE 10-3. FREQUENTLY GIVEN OTHER NAMES FOR TEST ITEMS

<i>Type</i>	<i>Frequency</i>
Synonym	
Mouth organ (harmonica)	7
Less descriptive	
Rope (noose)	9
Lock (latch)	20
Harness (yoke)	10
Error	
Muskrat (beaver)	6
Hippopotamus (rhinoceros)	7
Compass (protractor)	12
Protractor (compass)	8
Easel (tripod)	6
Easel (palette)	7

RESPONSES FOLLOWING INITIAL INCORRECT RESPONSE PLUS PROMPT

Prompting subjects to give an additional response following an Off Task or Don't Know response was not very effective in eliciting correct names (only 5 and 8 percent of the time, respectively) (Table 10-4). However, prompting for an additional response following Multiple Attempts and Related Name responses was quite effective in eliciting correct names (62 percent for Multiple Attempts and 39 percent for Related Name). There were too few Visual Misperception and Wrong Part responses to permit any conclusions about the effectiveness of prompts following these responses. With this system, prompts were not given following two types of responses, Unrelated Name and Mispronunciation.

INTERJUDGE AND INTRAJUDGE RELIABILITY

Five speech-language pathologists from our clinic each administered, coded, and scored approximately 12 of the 60 BNTs used in this study. One of these examiners (reliability judge 1) then coded the responses of all 60 subjects. A second examiner coded 10 randomly chosen tests from the 60 subjects. Items for which subjects gave only the designated BNT

TABLE 10-4. CORRECT RESPONSES FOLLOWING INITIAL INCORRECT RESPONSES AND PROMPT

<i>Response type</i>	<i>Number of prompts</i>	<i>Correct after prompt</i>	
		<i>Number</i>	<i>Percent</i>
Related name	135	53	39.3
Multiple attempts	21	13	61.9
Off task	21	1	4.8
Don't know	63	5	7.9
Visual misperception	4	1	25.0
Wrong part	3	1	33.3

name were not included in reliability calculations because these were easy scoring judgments and would have inflated reliability values. Overall interjudge and intrajudge point-to-point reliability for coding responses was high (Table 10-5). Interjudge reliability (between the original examiner and reliability judge 1) ranged from 85.9 percent to 95.2 percent, with an average of 89.1 percent. Intrajudge reliability (reliability judge 1) was 97.6 percent. The agreement between the two reliability judges was 93 percent.

CATEGORY RELIABILITY

Point-to-point category reliability levels were more variable than those across judges (Table 10-6). Category reliability ranged from 31.8 to 100 percent. The reliability for Unrelated Name and Don't know categories was poor even though these categories occurred quite frequently. The majority of disagreement were between Unrelated Name and Related Name and between Don't Know and Off Task categories. Wrong Part, Visual Misperception, Mispronunciation, and No Response were low occurrence categories and therefore not much can be said about the reliability of them. We did feel that chance reliability calculations for these multiple category data were helpful in determining which reliability values should be considered acceptable. The overall reliability of the coding system using a matrix agreement system that adjusts for chance was 91.2 percent.

REVISIONS BASED ON OUR RESULTS

Because Don't Know and Off Task responses could not be differentiated readily from one another and because prompts following both of these

TABLE 10-5. INTRAJUDGE AND INTERJUDGE RELIABILITY FOR CODING RESPONSES TO BOSTON NAMING TEST ITEMS

<i>Examiner</i>	<i>Agreements</i>	<i>Disagreements</i>	<i>Reliability (%)</i>
1	164	4	97.6
2	119	6	95.2
3	94	12	88.7
4	85	14	85.9
5	103	16	86.6

Note: Reliability for Examiner 1 is intrajudge reliability. Those for the other four examiners are interjudge reliability.

TABLE 10-6. POINT-TO-POINT RELIABILITY FOR CODING CATEGORIES

<i>Category</i>	<i>Agreements</i>	<i>Disagreements</i>	<i>Reliability (%)</i>
Correct name	301	5	98.4
Multiple attempt	17	4	81.0
Related name	150	8	94.9
Unrelated name	7	15	31.8
Off task	23	4	85.2
Visual misperception	3	1	75.0
Wrong part	3	1	75.0
Mispronunciation	5	0	100.0
Don't know	49	15	76.5
No response	2	0	100.0

types of responses were ineffective in eliciting correct responses, we combined them into one category, called Associated Responses, which does not receive a prompt. We also combined the Unrelated Name and Related Name categories into one category called Other Name because of the difficulty we had in differentiating these categories from one another. We retained the Related Name category prompt ("Tell me another name for that") for this combined category because it had been effective in eliciting correct responses nearly 40 percent of the time. We felt that these changes would make the coding and prompting procedures easier to use and more reliable.

RELIABILITY OF OUR REVISED PROCEDURES

Two speech-language pathologists (one who was involved in the original study and one who had given the standard BNT but had not seen our revised BNT coding and scoring procedures) listened to 10 audiotaped administrations of the BNT and coded and scored them with the revised procedures. The naive scorer was given only the printed administration and response coding and scoring instructions found in Appendices A and B. She was given no additional instruction or assistance as she learned the procedures and coded and scored the tapes. Overall interjudge reliability for coding responses was 90 percent. Category-by-category reliability ranged from 80 to 100 percent for categories that occurred more than twice in the sample.

In summary, we feel that these revised and expanded administration instructions and response coding, scoring, and prompting procedures make the Boston Naming Test a more reliable and informative measure of confrontation word retrieval performance. These revised procedures do not, in our opinion, compromise the validity of the norms provided by Kaplan and colleagues (1983) and Van Gorp and his colleagues (1986) because they are consistent with the procedures apparently employed when these normative subjects were tested. If one wishes to use any of the norms published (Kaplan et al., 1983; Van Gorp et al., 1986; or those reported herein), all BNT items must be administered, and only the responses specified in the BNT scoring booklet can be accepted as correct. However, when the results of such tests are interpreted, the clinician should consider that some BNT test items consistently elicit names other than those specified as correct in the test manual and that other items are vague or confusing to non-brain-damaged adults.

REFERENCES

- Kaplan, E., Goodglass, H., and Weintraub, S. (1983). *The Boston Naming Test*. Philadelphia: Lea & Febiger.
- Van Gorp, W., Satz, P., Kiersch, M., and Henry, R. (1986). Normative data on the Boston Naming Test for a group of normal older adults. *Journal of Clinical and Experimental Neuropsychology*, 8, 702-705.

APPENDIX A

INSTRUCTIONS FOR ADMINISTERING THE BOSTON NAMING TEST

Begin with Item 1 and continue through Item 60, unless the patient is in distress or refuses to continue. Give the following instructions:

Tell me the *name* of each of these pictures. Some pictures may have more than one name. Tell me the most common name for each one. If you are not sure, just give your best guess.

Write down the patient's responses in as much detail as necessary. Code the response and (when appropriate) give prompts based on the final utterance in the patient's response (e.g., "Let's see, that looks like a church. No. *It's a house.*").

Provide only one prompt (when appropriate) per patient response to a test item, except for responses that follow Visual Misperception plus Stimulus Prompt, and Wrong Part plus Prompt. These responses are treated as first responses and may receive an additional prompt (see Appendix B).

If the patient does not appear to understand the task or seems to forget the task, the general prompt "Tell me the *name* of the picture," may be given. However, do not give this prompt if the patient appears to understand the task but is either editorializing about the task or attempting to self-cue by talking about the item. In this case, no prompt is given, and the response is coded as an associated response (AR). (See Appendix B).

Allow the patient 20 seconds to respond to each picture. If a patient continues to respond past 20 seconds, say "Let's go on," and do not give an additional prompt. Code the final response.

If a prompt fails to elicit the correct response, a phonemic cue (the initial phoneme[s] underlined on the score sheet) may be given at the examiner's discretion to obtain additional information about the patient's ability to retrieve the word. However, the examiner should realize that providing phonemic cues may provide information to the patient that could affect his or her performance on subsequent administrations of the test.

Allow the patient 5 seconds to respond following a prompt or phonemic cue. Responses following Visual Misperception plus Stimulus Prompt and Wrong Part plus Prompt are treated as first responses, and patients are therefore allowed 20 seconds to respond to these prompts.

After completing the test, score each item + or -. (See Appendix B.) You will need to modify column headings on the score sheets to allow for marking of response codes and +/- scores.

APPENDIX B

RESPONSE CODING AND SCORING PROCEDURES FOR THE BOSTON NAMING TEST

<i>Code</i>	<i>Patient response</i>	<i>Example of response</i>	<i>Examiner response</i>
CN (Correct Name)	Picture name as in BNT test booklet.	House	Accept and go on. Score +.
	Multiple-word or compound-word responses containing the BNT name with same meaning as BNT name.	Large house	Accept and go on. Score +.
	<i>Exception.</i> If a multiple-word or compound-word response contains the BNT name, but it has a different meaning, prompt and code it as an other name.	Doghouse Apartment house	See ON, below. Score -.
ON (Other Name)	A name (or multiple names) that does not contain the BNT name and does not seem to be a visual misperception.	Building Dwelling School A kind of factory	"Tell me <i>another</i> name for that." Score -, even if the correct name is given after the prompt.
MA (Multiple Attempts)	Multiple attempts in which the BNT name is given, but is <i>not</i> the last response given.	House, building	"Tell me the <i>best</i> name." Score + if the correct name is given after the prompt.

(continued)

Appendix B. (continued)

<i>Code</i>	<i>Patient response</i>	<i>Example of response</i>	<i>Examiner response</i>
MA (continued)	<p><i>Exception 1.</i> If the response contains multiple attempts, and the BNT name is the last one, code it as a correct name.</p> <p><i>Exception 2.</i> If the response contains multiple attempts that <i>do not</i> include the BNT name, code the final response and prompt as appropriate.</p>	Building, house	Accept and go on. Score +.
AR (Associated Response)	If the response consists of description of use, personal reactions to the item, telling what the item is not, or an indication that the patient does not know the name of the item, score it as an associated response.	<p>People live in it. That's a big one. It's not a store. It's like an apartment. I can't think of it.</p>	No prompt. Score -.
VM (Visual Mis-perception)	Name that suggests that the patient has misperceived the test picture.	Coffin (for tongs) Umbrella (for mushroom)	<p>Stimulus prompt next to test item in BNT booklet. Treat the response to the stimulus prompt as a first response and prompt as appropriate.</p> <p>Correct responses following a stimulus cue are scored +.</p>

<i>Code</i>	<i>Patient response</i>	<i>Example of response</i>	<i>Examiner response</i>
WP (Wrong Part)	Name for wrong part of the picture.	Doorknob (for knocker)	"No, this." an point to part to be named. Correct responses following a prompt are scored +.
MP (Mispronunciation)	Articulatory errors, literal paraphasias, unintelligible words, or neologisms that do not result in other real words. Singular form for words whose standard form in plural (e.g., scissor, tong).	Crushmoom Abiscus (abacus) Scissor	Accept and go on. For later scoring, score as + if a naive listener would recognize it as the BNT name.
NR (No Response)	No response in 20 s.	—	"Give it a try." Score —, even if the correct name is given after the prompt.

Source: Speech Pathology Section, Minneapolis Veterans Administration Medical Center, November 1987, revised May 1988.