APHASIOLOGY, 2001, *15* (1), 85–92

Short Report
The 60-Item Boston Naming Test: Cultural bias and possible adaptations for New Zealand

Suzanne L. Barker-Collo
The University of Auckland, New Zealand

This study examined whether a sample of New Zealand university students performed differently on the 60-item Boston Naming Test (BNT) when compared to published North American norms. The BNT was administered to 58 New Zealand university students, and their performance was compared to published data from North America. Mean performance of the sample was significantly worse than published North American norms. In comparing proportion of errors made on BNT items, New Zealanders made 60% more error on the items *pretzel* and *beaver*, and 20% more errors on the items *globe*, *funnel*, and *tripod* than did North Americans. In addition, the New Zealand sample made 20% more errors on the item *asparagus* than North American and Australian samples. Within the New Zealand sample, Maori individuals performed significantly worse than European individuals. Items contributing to this difference included *canoe*, *beaver*, and *abacus*. It was concluded that in administering the BNT to New Zealanders attention should be given to the potential for cultural biases. Adaptations to better reflect New Zealand culture are suggested.

The Boston Naming Test (BNT; Kaplan, Goodglass, & Wintraub, 1983) is likely the most frequently administered confrontation-naming test in the western world. It provides invaluable diagnostic information in the detection of mild naming deficits and word-retrieval difficulties in patients with aphasia (e.g., Kohn & Goodglass, 1985; Sandson & Albert, 1987), as well as those suffering other forms of linguistic impairment as a result of the cognitive consequences of brain injury (e.g., Jordan, Cannon, & Murdoch, 1992), degenerative disorders such as multiple sclerosis and dementias (e.g., Beatty & Monson, 1989; Lindman, 1996), and stroke (Margolin, Pate, Friedrich, & Elia, 1990). Because the BNT is geared towards the identification of mild deficits, clinicians must consider factors that may influence scores on the BNT. One such factor is culture experience.

Despite its clinical utility, the content of the BNT reflects the cultural context in which it was developed, and may not be applicable to persons from other cultures. For example, Kim and Na (1999) indicate that many BNT items are not applicable to Korean

Address correspondence to: Dr. Suzanne Barker-Collo, Department of Psychology, University of Auckland, Private Bag 92019, Auckland, New Zealand. E-mail: s.barker-collo@auckland.ac.nz

http://www.tandf.co.uk/journals/pp/02687038.html DOI:10.1080/02687040042000124

populations, noting that some BNT items (e.g., trellis, beaver, pelican) are not familiar to Korean individuals, whereas most Koreans are able to easily recognise some of the most difficult BNT items (e.g., abacus). In examining the validity of applying North American BNT norms to older Australians, Worrall, Yiu, Hickson, and Barnett (1995) suggested that original BNT word-items that are not frequently used in the Australian English language (e.g., beaver) should be replaced with more culturally appropriate items (e.g., platypus). The authors concluded that use of the BNT outside North America requires that new normative data must be collected due to word-frequency differences across cultures (Worrall et al., 1995). In a similar examination of Australian adults Pozzebon (1990) reported that Australian clinicians often omit items such as ''beaver'' and ''pretzel'' from the BNT, or accept close approximations. However, while these approaches do acknowledge the cultural bias of BNT items, they run the risk of interfering with test reliability and interpretation of the results.

In New Zealand, recent efforts have begun to examine the cultural bias of assessments of cognitive functioning such as those for verbal memory. These preliminary investigations have revealed that New Zealand samples, particularly individuals of Maori ethnicity, perform differently than the normative data would anticipate, and that these differences in performance may be due to cultural bias of test items (e.g., Ogden & McFarlane-Nathan, 1997; Barnfield & Leathem, 1998).

The purpose of the present study was to determine whether a sample of New Zealand university students performed differently on the BNT when compared to published North American norms. In addressing this issue, an attempt was made to identify those items that are culturally biased through a comparison of errors made on BNT items by New Zealanders to data available from Canada and Australia.

METHOD

Participants

Participants included 58 New Zealand-born first- and second-year university students who ranged in age from 17.5 to 25.25 years, with a mean age of 20 .2 years; 43% percent of participants were male. Marital status reported by participants revealed that 84.5% were single, 8.6% married, 3.4% living in common-law relationships, and 3.4% were engaged. Cultural identity of 41 (70.7%) participants was Pakeha (i.e., European), while 15 (25.9%) were Maori, and 2 (3.4%) identified themselves as Pacific Island peoples. While all participants indicated that their first language was English, nine of the 15 Maori participants indicated that they spoke both Maori and English at home. One of the two Pacific Island peoples indicated that he was bilingual, speaking both English and Tongan. A self-report history of medical and psychiatric problems, including identification of prior history of concussion or loss of consciousness, brain injury, or previously identified learning disability was obtained from each potential participant. All persons with a prior history of psychiatric illness, neurological disease, head injury, or learning disability were excluded from the sample.

Procedure and measures

All participants were tested individually using the Boston Naming Test (BNT: Kaplan et al., 1983). Participants were administered all 60 items of the BNT, beginning with item one. Participants were given up to 20 seconds to name each object, with full credit accorded to self-corrections that occurred during this time period. If an incorrect response

was made, or if no response was made within this time period a stimulus cue was provided (e.g., ''it is an animal''). If provision of a stimulus cue did not result in a correct answer, a phonemic cue was provided (e.g., ''it starts with 'be' . . .''). All responses were scored according to the standard single-word scoring key presented in the BNT response booklet. The standard discontinuation criterion of failure to correctly name objects on six consecutive trials was used. None of the participants met criteria for discontinuation.

RESULTS AND DISCUSSION

The purpose of the present study was to determine whether a sample of New Zealand university students performed differently on the BNT than North American norms. Table 1 presents the mean and standard deviation scores obtained by the present sample, and those of two published North American studies. Although there are many studies that provide normative data for the BNT (e.g., Van Gorp, Satz, Kiersch, & Henry, 1996; Welch, Doineau, Johnson, & King, 1996), the studies of Tombaugh and Hubley (1997; Canadian) and Farmer (1990; United States) have been selected for comparison with the present sample. Selection of these samples was made on the basis of sample age and clinical usage. With regard to age, the majority of studies that report normative data for the BNT have focused on older adults. Thus, it was necessary that the samples selected included separate data for younger adults to ensure that valid comparisons could be made. With regard to usage, it should be noted that due to its inclusion in Spreen and Strauss's (1998) compendium of neuropsychological tests, Tombaugh and Hubley's (1997) normative data is likely the most commonly applied to younger adults in clinical settings.

Comparison of published BNT mean total scores (total spontaneous correct + total correct with stimulus cue) and standard deviations of Tombaugh and Hubley (1997; Canadian) and Farmer (1990; United States), to the present sample indicated that the three samples were significantly different, F (171) = 55.296, $p < .01$. Post-hoc analysis indicated that the present sample was significantly different from both of the North American samples ($p < .05$). Although Canadian and American cultures are not identical, they are considered more congruent that North American and New Zealand cultures. The two North American samples used for comparison in this study did not differ significantly in BNT performance. The performance of the present sample places the average New Zealand participant 1.2 standard deviations below the average performance of the closest North American normative sample (i.e., Tombaugh & Hubley 1997), placing the performance of the average New Zealander below the 10th percentile. Thus,

TABLE 1
Mean and standard deviations by age range from North American and New Zealand samples

	Age range	*BNT*[1] *mean*	*BNT standard deviation*
North American samples			
Tombaugh & Hubley 1997	25–34	56.0	2.90
Farmer, 1990	20–69	57.47	2.789
New Zealand sample	17.5–25.25	52.5	2.2

[1] BNT mean and SD are based on total number of spontaneous correct responses, plus correct responses following stimulus cues.

use of Tombaugh and Hubley's (1997) norms for this confrontation-naming test is likely to result in a much larger proportion of New Zealand individuals being identified as having deficits. Although clients' test-taking behaviours, responsiveness to cues, and an analysis of errors made is as integral a part of the interpretation of BNT performance as comparison to normative data, comparison to normative data continues to be one of the bases on which clinical decisions are made. It must, therefore, be concluded that until normative data for New Zealanders are available; caution should be used in applying existing norms to the identification of confrontation-naming deficits from BNT scores of New Zealand subjects. The finding that New Zealanders performed significantly worse than the other samples is particularly interesting given that the New Zealand sample was younger, and likely more educated, than the other samples. Although not all studies of the BNT have indicated the presence of significant age and educational effects, younger age and greater levels of education have been associated with improved BNT scores (Farmer, 1990; Tombaugh & Hubley, 1997; Welch et al., 1996). This would indicate that the differences found in the present study are in the opposite direction to what would be expected, given age and education levels, and that the cultural bias in performance may have been underestimated.

In addition to comparing overall performance of the New Zealand sample to existing North American data, this study attempted to identify BNT items most subject to cultural bias through examining the patterns of errors made by the New Zealand sample as compared to the pattern of errors observed in Canadian (Tombaugh & Hubley, 1997) and Australian samples (Worral et al., 1995), who are considered to be more culturally similar to New Zealanders than Canadians/North Americans that (see Table 2). Although the percentages correct responses of Australian and New Zealand samples are presented, it should be noted that the two samples did differ significantly in age. Thus, inclusion of the Australian sample in this comparison has not been seen as a comparison of absolute error frequencies, but is viewed as a means of identifying similarities in patterns of performance within the two samples that could be a result of similarities between New Zealand and Australian culture/experience.

As can be seen in Table 2, the proportion of New Zealand participants making errors on two BNT items (i.e., *pretzel*, *beaver*) were much higher than that of the Canadian

TABLE 2
Difficulty index expressed as proportion of correct responses to each BNT item

Item	*Canadian (n = 219)*	*Australia (n = 136)*	*New Zealand (n = 58)*
1. bed	100.0	100	100
2. tree	100.0	100	100
3. pencil	100.0	100	100
4. house	100.0	100	100
5. whistle	99.5	93.4	98.3
6. scissors	100.0	100	100
7. comb	100.0	100	100
8. flower	100.0	100	100
9. saw	100.0	100	100
10. toothbrush	100.0	100	100

(Continued)

TABLE 2
(Continued)

Item	*Canadian (n = 219)*	*Australia (n = 136)*	*New Zealand (n = 58)*
11. helicopter	99.1	97.8	100
12. broom	100.0	98.5	100
13. octopus	90.0	90.4	89.7
14. mushroom	99.5	99.3	98.3
15. hanger	100.0	100	100
16. wheelchair	100.0	100	100
17. camel	99.1	99.3	98.3
18. mask	98.6	90.4	89.7
19. pretzel	92.2	**29.4**	**27.6**
20. bench	99.5	97.8	98.3
21. racquet	100.0	99.3	100
22. snail	95.4	96.3	98.3
23. volcano	97.7	95.6	98.3
24. seahorse	84.9	81.6	84.5
25. dart	98.6	93.4	96.6
26. canoe	100.0	91.2	89.7
27. globe	96.8	**75.7**	**77.6**
28. wreath	99.5	87.5	93.1
29. beaver	97.5	**37.5**	**31.0**
30. harmonica	96.8	98.5	98.3
31. rhinoceros	90.4	88.2	94.8
32. acorn	93.6	68.4	79.3
33. igloo	99.1	84.6	91.4
34. stilts	95.0	90.4	91.4
35. dominoes	90.9	80.9	82.8
36. cactus	100.0	94.1	93.1
37. escalator	99.1	94.1	96.6
38. harp	97.3	97.8	98.3
39. hammock	94.1	87.5	93.1
40. knocker	97.7	69.9	81.0
41. pelican	92.7	97.1	98.3
42. stethoscope	95.0	92.6	93.1
43. pyramid	96.8	92.6	94.8
44. muzzle	92.7	77.2	91.4
45. unicorn	91.3	75.0	82.8
46. funnel	96.3	**76.0**	**75.9**
47. accordion	81.7	89.0	93.1
48. noose	91.3	89.7	89.7
49. asparagus	93.6	91.2	**70.7**
50. compass	69.0	58.8	60.3
51. latch	80.8	79.4	84.5
52. tripod	89.5	**57.4**	**55.2**
53. scroll	92.7	83.1	93.1
54. tongs	84.5	64.7	89.7
55. Sphinx	75.8	73.5	77.6
56. yoke	63.0	59.5	69.0
57. trellis	77.2	84.5	87.9
58. palette	69.0	80.9	77.6
59. protractor	39.7	27.9	27.6
60. abacus	57.5	60.3	65.5

Difficulty index: percent giving correct response spontaneously or with stimulus cue.

sample (i.e., >60% more errors), and were well below those of adjacent items. These items can be seen to reflect differences between New Zealand and Canadian experiences. For example, in considering the item ''pretzel'' it should be noted that pretzels are not a common food item in New Zealand. Examination of errors made on this item indicates that the majority of errors (i.e., ''knot'' or ''rope'') reflected objects with similar appearance that are common in New Zealand's nautical culture. Similarly, the majority of errors made when presented with ''beaver'' tended to reflect animals of similar characteristics that are more common to New Zealand and Australia (e.g., ''platypus'' or ''possum'').

The proportion of errors made was also higher, but to a lesser extent (i.e., >20% more errors), for three BNT items (i.e., *globe*, *funnel*, *tripod*). This pattern of performance was, again, similar to that of the Australian sample. In addition, the New Zealand sample made 20% more errors on the item *asparagus* than either of the other samples. This result was quite surprising, as asparagus is commonly used as a canned vegetable in New Zealand. It was hypothesised that, as New Zealanders are more likely to have experience of canned rather than fresh asparagus, the increased frequency of errors on this item may have been due to the differences in appearance of the two. In addition, because comparisons of frequency data were to older samples, it is possible that increased frequency of errors on the items asparagus, globe, funnel, and tripod could have been due to differences in experience of different age groupings. For example, it is possible that items such as globes and tripods that were once commonly seen in classrooms may have been replaced in recent generations by newer technology. One must consider whether a telescope with a tripod or a globe would be seen in classrooms of today which rely more and more on video and computer/internet images.

In summary, these findings indicate that some BNT items, particularly those for which large discrepancies were found, are culturally biased and should not be used in determining word-retrieval performance in New Zealand samples. It is suggested that replacing the items pretzel and beaver with biscuit and possum, respectively, modify the BNT. As BNT items are presented in order of frequency of usage, further research must be conducted not only to generate normative data for any modified version of the BNT, but also to determine the frequency of usage of these suggested replacements. Unfortunately, no data are currently available on frequency of word usage for BNT items in New Zealand.

Within the New Zealand sample, gender did not have a significant impact on overall BNT performance. However, in considering different cultural groups within the New Zealand sample it was found that Maori individuals performed significantly worse on the BNT than Pakeha/European individuals, $t(54) = 2.81$, $p < .01$. Items contributing significantly to this difference in performance included *canoe*, *beave*r, and *abacus* ($p < .001$). In examining individual responses to these three items it could be seen that the majority of errors on items *beaver* and *abacus* were due to ''don't know'' responses for both groups. In contrast, for the item *canoe*, while errors by Pakeha/European participants were due to ''don't know'' responses all of the errors made by Maori individuals were due to provision of the response ''waka'', a Maori word used to refer to a large war canoe. When data were re-analysed with the term ''waka'' considered a correct response, while overall performance on the BNT remained significantly different, $t(54) = 2.85$, $p < .01$, performance of Pakeha/European and Maori individuals on this item were not significantly different ($p > .05$). It should be noted that Maori and Pakeha/European groups within this sample did not differ significantly in age or years of education ($p > .05$), so these factors do not explain differences in performance. In addition, it should

be noted that while all participants indicated that their first language was English, 9 of the 15 Maori participants indicated that they spoke both Maori and English at home, and one of the two Pacific Island peoples indicated that he was bilingual, speaking both English and Tongan. As concluded by Hermans, Bongaerts, DeBot, and Schreuder (1998) bilingual speakers cannot suppress activation from their first language while naming pictures in a second language. In addition, bilingual individuals have been reported to produce greater variability in responses on the BNT in their non-English language, despite there being no significant differences in overall first- versus second-language performances (Kohnert, Hernandez, & Bates, 1998). Thus, the differences between groups may have been a result of bilingualism, rather than a result of culture itself.

Although the small number of Maori individuals in the sample, along with their high level of education, raises questions about the generalisability of these findings, the results indicate that while New Zealanders as a whole are the subject of cultural bias when presented with the BNT, cultural groups within New Zealand might be the subject of additional cultural biases. Given these findings, it is suggested that in administering the BNT to New Zealanders, particularly those who are not part of the Pakeha/European majority that more closely resembles its North American counterparts, added attention should be given to the potential for cultural biases. The identification of acceptable alternative responses to BNT items in the Maori language may provide additional validity to the interpretation and scoring of the BNT when administered to New Zealanders.

In conclusion, despite its clinical utility the content of the BNT reflects the cultural context in which it was developed, and may not be applicable to persons from other cultures, and it appears that there is the need to modify existing BNT items when used in populations outside North America. Although possible modifications to the BNT for use within New Zealand have been suggested (e.g., alteration of culturally biased words, provision of Maori alternatives), the applicability of any modified version of the BNT rests on the availability of normative data collected within New Zealand. Thus, it would seem that further study is required to specify the generalisability of these findings to New Zealanders with differing levels of education, and from different age groups to those included in this study. Further research should also be conducted to examine the impact of levels of acculturation on BNT performance. However, while such modifications may result in a version of the test that is *more* valid in New Zealand, it is unlikely that any single version of the BNT will be culturally appropriate to the diverse populations that comprise New Zealand society. It is unrealistic to think that a single version of any test could be appropriate across cultures or sub-cultures. Indeed, clinical decision making must always consider the diversity of experience that clients bring to the testing situation, and how this will impact on performance.

Manuscript received 4 July 2000
Manuscript accepted 9 September 2000

REFERENCES

Barnfield, T.V., & Leathem, J. M. (1998). Neuropsychological outcomes of traumatic brain injury and substance abuse in a New Zealand prison population. *Brain Injury*, *12*(11), 951–962.

Beatty, W.W., & Monson, N. (1989). Lexical processing in Parkinson's disease and multiple sclerosis. *Journal of Geriatric Psychiatry and Neurology*, *2*, 145–152.

Farmer, A. (1990). Performance of normal males on the Boston Naming Test and The Word Test. *Aphasiology*, *4*, 293–296.

Hermans, D., Bongaerts, T., DeBot, K., & Schreuder, R. (1998). Producing words in foreign language: Can speakers prevent interference from their first language? *Bilingualism, 1*(3), 213–229.

Jordan, F.M., Cannon, A., & Murdoch, B.E. (1992). Language abilities of mildly closed head injured (CHI) children 10-years post-injury. *Brain Injury, 6*, 39–44.

Kaplan, E.F., Goodglass, H., & Wintraub, S. (1983). *The Boston Naming Test. Experimental edition.* Philadelphia: Lea & Febiger.

Kim, H., & Na, D.L. (1999). Brief Report. Normative data on the Korean version of the Boston Naming Test. *Journal of Clinical and Experimental Neuropsychology, 12*, 127–133.

Kohn, S.E., & Goodglass, H. (1985). Picture naming in aphasia. *Brain and Language, 27*, 380–384.

Kohnert, K.J., Hernandez, A.E., & Bates, E. (1998). Bilingual performance on the Boston Naming Test: Preliminary norms in Spanish and English. *Brain & Language, 65*(3), 422–440.

Lindman, K.K. (1996). *Gender differences in dementia of the Alzheimer's type: Evidence for different semantic memory degradation.* Paper presented at the meeting of the International Neuropsychological Society, Chicago.

Margolin, D.I., Pate, D.S., Friedrich, F.J., & Elia, E. (1990). Dysnomia in dementia and in stroke patients: Different underlying cognitive deficits. *Journal of Clinical and Experimental Neuropsychology, 12*, 597–612.

Ogden, J.A., & McFarlane-Nathan, G. (1997). Cultural bias in the neuropsychological assessment of young Maori men. *New Zealand Journal of Psychology, 26*(2), 2–12.

Pozzebon, M. (1990). A critique of the Boston Naming Test—and some suggestions towards its modification. *Melbourne Papers in Applied Linguistics, 2*, 19–29.

Sandson, J., & Albert, M.L. (1987). Varieties of perseveration. *Neuropsychologia, 22*, 715–732.

Spreen, O., & Strauss, E. (1998). *A compendium of neuropsychological tests, Second Edition.* New York: Oxford University Press.

Tombaugh, T.N., & Hubley, A. (1997). The 60-item Boston Naming Test: Norms for cognitively intact adults ages 25 to 88 years. *Journal of Clinical and Experimental Neuropsychology, 19*, 922–932.

Van Gorp, W., Satz, P., Kiersch, M.E., & Henry, R. (1996). Normative data on the Boston Naming Test for a group of normal older adults. *Journal of Clinical and Experimental Neuropsychology, 8*(6), 702–705.

Welch, L.W., Doineau, D., Johnson, S., & King, D. (1996). Education and gender normative data for the Boston Naming Test in a group of older adults. *Brain and Language, 53*, 260–266.

Worrall, L.E., Yiu, E.M-L., Hickson, L.M.H., & Barnett, H.M. (1995). Normative data for the Boston Naming Test for Australian elderly. *Aphasiology, 9*(6), 541–551.