Journal of the International Neuropsychological Society (2012), **18**, 375–378.

doi:10.1017/S1355617711001810

BRIEF COMMUNICATION

Reliable Change on the Boston Naming Test

Bonnie C. Sachs,[1] John A. Lucas,[1] Glenn E. Smith,[2] Robert J. Ivnik,[2] Ronald C. Petersen,[3] Neill R. Graff-Radford,[4] AND Otto Pedraza[1]

[1]Department of Psychiatry and Psychology, Mayo Clinic, Jacksonville, Florida
[2]Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minnesota
[3]Department of Neurology, Mayo Clinic, Rochester, Minnesota
[4]Department of Neurology, Mayo Clinic, Jacksonville, Florida

(RECEIVED July 6, 2011; FINAL REVISION December 1, 2011; ACCEPTED December 2, 2011)

Abstract

Serial assessments are commonplace in neuropsychological practice and used to document cognitive trajectory for many clinical conditions. However, true change scores may be distorted by measurement error, repeated exposure to the assessment instrument, or person variables. The present study provides reliable change indices (RCI) for the Boston Naming Test, derived from a sample of 844 cognitively normal adults aged 56 years and older. All participants were retested between 9 and 24 months after their baseline exam. Results showed that a 4-point decline during a 9–15 month retest period or a 6-point decline during a 16–24 month retest period represents reliable change. These cutoff values were further characterized as a function of a person's age and family history of dementia. These findings may help clinicians and researchers to characterize with greater precision the temporal changes in confrontation naming ability. (*JINS*, 2012, *18*, 375–378)

Keywords: BNT, RCI, Aging, Dementia, Serial Assessment

INTRODUCTION

Serial assessments are commonplace in neuropsychological practice and used to document cognitive trajectory in patients who undergo neurosurgical, behavioral or pharmacologic interventions; monitor the natural progression of neurodegenerative illness; and identify abnormal patterns of development in children and adolescents. A position paper by the American Academy of Clinical Neuropsychology (AACN; Heilbronner et al., 2010) highlights the ubiquity of serial cognitive assessments in both clinical and forensic practice, and emphasizes the need for additional research to characterize clinical change scores for many neuropsychological tests in common use.

The Boston Naming Test (BNT; Kaplan, Goodglass, & Weintraub, 1983) is one of the most widely used instruments for the assessment of confrontation naming ability (Rabin, Barr, & Burton, 2005). The reliability and validity of the instrument are well established (Strauss, Sherman, & Spreen, 2006) and normative reference values are available from a variety of sources, including Mayo's Older Americans Normative Studies (MOANS) and Older African American Normative Studies (MOAANS). Numerous studies have investigated changes in BNT performance across the lifespan (Albert, Heller, & Milberg, 1988; Au et al., 1995; Connor, Spiro, Obler, & Albert, 2004; Cruice, Worrall, & Hickson, 2000; Ivnik, Malec, Smith, Tangalos, & Petersen, 1996; Mitrushina & Satz, 1995; Welch, Doineau, Johnson, & King, 1996), including a relatively large study that detailed the effects of age, education, and gender on naming ability (Zec, Burkett, Markwell, & Larsen, 2007). However, few studies have systematically examined practice effects or reliable change on the BNT. For instance, a study of 122 older adults found no practice effect for healthy individuals tested annually over a 3-year time period (Mitrushina & Satz, 1995), a finding that was also obtained in a 4-year study of 91 healthy elderly Australian participants (Cruice et al., 2000). Two additional longitudinal studies examined BNT performance in older adults, but relied on the 85-item and 15-item versions of the test and did not provide measures of reliable change (Au et al., 1995; Kent & Luszcz, 2002). In a study of 353 individuals, Zec, Markwell, Burkett, & Larsen (2005)

Correspondence and reprint requests to: Bonnie C. Sachs, Department of Neurology, Virginia Commonwealth University, Richmond, VA 23298. E-mail: bsachs@vcu.edu

showed that a 4-point decline in BNT scores represents reliable change over a 9–15 month interval. As noted by the study authors, however, the BNT was administered without the provision of semantic or phonemic cues, which may limit the generalizability of their findings to standard clinical practice. Furthermore, no index of reliable improvement was provided.

The goal of the current study is to identify normative rates of change on the BNT in a large cohort of cognitively normal adults and provide estimates of reliable decline or improvement over a 9–24 month interval.

METHOD

Participants

Study participants included 844 Caucasian Americans over the age of 55 who took part in the Mayo Older Adult Normative Studies (MOANS). Study criteria and recruitment methodologies for MOANS have been detailed previously (Ivnik et al., 1990). In brief, all participants were community-dwelling adults who functioned independently and met the following inclusion criteria: (1) evaluation by a primary care physician within 1 year of study entry; (2) independence in instrumental daily activities as rated by an informant; (3) no active or uncontrolled medical, neurologic, or psychiatric condition that could adversely affect cognitive functioning; and (4) no use of psychotropic medications in sufficient amounts that could adversely affect cognition.

Only individuals who remained cognitively normal during all follow-up evaluations, and whose second evaluation occurred 9–24 months following their baseline exam, were included in this study. Many individuals had additional subsequent exams: 82.3% were seen a total of 3 times, 70.6% a total of 4 times, and 58.1% a total of 5 times.

All data were obtained in full compliance with a research protocol approved by the Mayo Clinic Institutional Review Board.

Statistical Analyses

Between-group differences were calculated using independent-sample *t* tests, and bivariate correlations were used to examine the relationship between demographic variables and BNT scores. A reliable change index (RCI) was calculated for all individuals with follow-up exams between 9 and 24 months, and then separately for those whose post-test exam occurred within 9–15 months or 16–24 months from the date of their baseline assessment. This latter approach provides reference values for individuals who return for follow-up approximately 1 *versus* up to 2 years after their initial exam.

A reliable change estimate, adjusted for practice effects, was calculated based on the standard error of measurement of the difference (SE_{diff}; equation 1), where SD_x and SD_y represent the standard deviation of baseline and follow-up BNT scores, respectively, and r_{xy} represents the correlation between baseline and follow-up scores. A 90% prediction interval was calculated by multiplying SE_{diff} by the corresponding value from the normal z-distribution (± 1.64). This value was then added to or subtracted from the mean practice effect, which was determined by subtracting the mean BNT score at baseline from mean BNT score at follow-up (BNT time2 – BNT time1). RCI cutoff values were then rounded to the nearest integer for ease of interpretation.

$$SE_{diff} = [(SDx^2 + SDy^2)(1 - r_{xy})]^{1/2} \quad (1)$$

RESULTS

Demographic characteristics and mean BNT data for our sample are presented in Table 1. Participants ranged in age from 56 to 99 years of age ($M = 75.9$; $SD = 7.6$) and, on average, had some college education ($M = 14.4$ years; $SD = 2.8$). Approximately two-thirds of the participants were female (65.5%). Average length of time between assessments was 14 months. Six hundred thirty-two participants were tested in the 9–15 month interval and the remaining 212 were tested between 16 and 24 months after their baseline assessment.

Mean BNT raw scores at baseline ($M = 53.4$; $SD = 5.6$) and follow-up ($M = 54.0$; $SD = 5.5$) were in the average clinical range (age-corrected scaled scores = 10) based on published normative data. Less than 1% of all participants had age-corrected scaled scores of 6 or less at baseline or follow-up. Mean practice effects did not differ between cognitively healthy men ($M = 0.63$; $SD = 2.5$) and women ($M = 0.53$; $SD = 3.1$) ($t(707.9) = .52$; $p = .61$, adjusted for unequal variances) and there was no significant relationship with education ($r = .05$; $p = .19$). Practice effects declined with age ($r = -0.11$; $p = .001$) and increased with a positive family history of dementia ($r = .11$; $p = .001$).

The indices of reliable change are summarized in Table 2. When examining all participants retested at any point between 9 and 24 months post-baseline, a decline of at least 4 points or an improvement of at least 5 points reflected reliable change. When examining only those who were retested approximately 1 year after baseline (9–15 months), a 4-point decline or a 5-point improvement also constituted reliable change. For those retested 16–24 months after baseline, a decline of at least 6 points or an improvement of at least 7 points constituted a reliable change. When considering age (median split), a decline of 4 points constituted reliable

Table 1. Demographic characteristics ($N = 844$) and Boston Naming Test (BNT) scores

	Mean	*SD*	Range
Age at baseline (years)	75.9	7.6	56–99
Education (years)	14.4	2.8	4–20
Test–retest interval (months)	14.2	3.1	9–24
BNT score (baseline)	53.4	5.6	24–60
BNT score (follow-up)	54.0	5.5	27–60
Sex (% Male)	34.5		

https://doi.org/10.1017/S1355617711001810 Published online by Cambridge University Press

Table 2. Reliable change indices for BNT scores.

	N	r_{xy}	SE_{diff}	90% PI	Practice effect	RCI cutoff
All participants (9–24 months)	844	.86	2.94	± 4.82	.56	−4 ≥ RC ≥ +5
Time interval (9–15 months)						
All participants	632	.84	2.66	± 4.36	.61	−4 ≥ RC ≥ +5
Age at baseline						
<75	316	.78	2.32	± 3.80	.72	−4 ≥ RC ≥ +5
≥75	316	.86	2.95	± 4.84	.49	−5 ≥ RC ≥ +6
Family history of dementia						
Yes	409	.83	2.57	± 4.22	.59	−4 ≥ RC ≥ +5
No	199	.86	2.78	± 4.56	.59	−4 ≥ RC ≥ +6
Time interval (16–24 months)						
All participants	212	.86	3.64	± 5.97	.43	−6 ≥ RC ≥ +7
Age at baseline						
<81	106	.84	3.14	± 5.14	.77	−5 ≥ RC ≥ +6
≥81	106	.84	4.06	± 6.66	.08	−7 ≥ RC ≥ +7
Family history of dementia						
Yes	95	.87	3.49	± 5.72	.53	−6 ≥ RC ≥ +7
No	115	.87	3.48	± 5.70	.16	−6 ≥ RC ≥ +6

Note. r_{xy} = test–retest reliability; SE_{diff} = standard error of the difference; Practice effect = (BNT time2 mean score –BNT time1 mean score); PI = 90% prediction interval; RCI = reliable change index adjusted for practice effect. Age at baseline based on median split.

change for adults younger than 75 years of age and tested within 9–15 months. For adults 75 years and older and tested within 9–15 months, or adults younger than 81 years of age tested within 16–24 months, a decline of 5 points constituted reliable change. A 7-point decline in adults 81 years and older and tested within 16–24 months represented reliable change. Within their respective retest intervals, RCI's were nearly identical for those with and without a family history of dementia.

We examined the frequency of individuals who exceeded the respective cut-off values for significant change at 9–15 and 16–24 month periods. We expected approximately 5% of follow-up scores to be above and approximately 5% below the 90% prediction interval. Indeed, 6.2% of healthy participants declined at least 4 points during the 9–15 month interval and 6.9% increased at least 5 points. Within the 16–24 month interval, 6.5% of cognitively healthy adults declined at least 6 points and 5.3% improved by 7 points or more.

DISCUSSION

The present study sought to improve the clinical utility of the BNT by providing reliable change estimates derived from a large cohort of adults aged 56 and over. All participants were considered cognitively healthy for the duration of the study, with more than half obtaining annual follow-up evaluations for 5 or more years. For all participants regardless of follow-up interval, a 4-point decline or a 5-point improvement on the BNT represented significant change. For those individuals retested approximately 1 year (9–15 months) after baseline, a 4-point decline or a 5-point improvement also represented statistically significant change in BNT scores. This overlap was not surprising because three-fourths of the total sample was retested within the 9–15 month interval. For those retested 16–24 months after baseline, a decline of at least 6 points or improvement of at least 7 points constituted reliable change. This broader range for reliable change scores in the 16–24 month interval is likely due to the smaller sample size and larger standard error of measurement of the difference in that group.

Increased precision in RCI values also may be obtained by considering the individual's age at baseline examination. Compared to their younger counterparts, older individuals tended to benefit less from repeated test exposure, as evidenced by smaller practice effects, and exhibited less stability in naming, as evidenced by a larger test–retest correlation coefficient and wider prediction intervals. Although all study participants remained clinically normal, it is also possible that some in the older age group may harbor preclinical neurodegenerative conditions that contribute to diminished score stability.

Although a family history of dementia represents an important risk factor for cognitive decline and subsequent dementia, there is no discernible difference in BNT change scores between cognitively normal individuals with or without a family history. It remains possible, however, that in the presence of a neurodegenerative process, BNT scores for those with a family history may decline faster than for those without any family history of dementia.

While the current study provides useful information for clinicians and researchers about the measurement of change on the BNT, there are several limitations that bear discussion. First, RCI values were calculated for individuals whose initial follow-up occurred 9–24 months after baseline exam; it is unclear whether these cutoff values apply at shorter or lengthier test–retest intervals. Second, we were only able to provide data for a single follow-up assessment rather than

https://doi.org/10.1017/S1355617711001810 Published online by Cambridge University Press

multiple repeated sessions. Third, the sample consisted only of Caucasian adults. Given the literature demonstrating unequal performance on neuropsychological instruments and differential rates of cognitive decline for Caucasian and African-American elders (Lucas et al., 2005; Manly et al., 1998; Pedraza et al., 2009; Wagner et al., 2007), caution is urged if applying these RCI values to non-Caucasian samples. Fourth, we were unable to obtain a sufficiently large clinical sample in which to test the current RCI's. These normative change scores need to be validated in independent clinical samples to establish the generalizability of the current parameters in various clinical populations.

These results are consistent with those reported by Zec and colleagues (2005), but extend their findings by providing RCI parameters for interval decline and improvement; evaluating change at two different follow-up time points; and determining the role of various factors such as age, education, and family history of dementia on reliable change. Moreover, our results are consistent with the needs expressed in the recent AACN position paper to define the normal boundaries of temporal fluctuations in cognitive test scores. It is hoped that these results may be useful for clinicians and researchers to identify individuals who begin to show early decline in naming ability, and perhaps increase our precision when evaluating language deficits in progressive conditions such as semantic dementia and Alzheimer's disease.

ACKNOWLEDGMENTS

There are no financial or other relationships that could be interpreted as a conflict of interest affecting this manuscript. This study was supported by NIH grants NS54722, AG016574, and AG06576.

REFERENCES

Albert, M.S., Heller, H.S., & Milberg, W. (1988). Change in naming ability with age. *Psychology and Aging*, *3*(2), 173–178.

Au, R., Joung, P., Nicholas, M., Obler, L.K., Kass, R., & Labert, M.L. (1995). Naming ability across the adult life span. *Aging and Cognition*, *2*(4), 300–311.

Connor, L.T., Spiro, A., Obler, L.K., & Albert, M.L. (2004). Change in object naming ability during adulthood. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, *59*, 203–209.

Cruice, M.N., Worrall, W.E., & Hickson, L.M.H. (2000). Boston Naming Test: Test results for healthy older Australians; A longitudinal and cross-sectional study. *Aphasiology*, *14*(2), 143–155.

Heilbronner, R.L., Sweet, J.J., Attix, D.K., Krull, K.R., Henry, G.K., & Hart, R.P. (2011). Official position of the American Academy of Clinical Neuropsychology on serial neuropsychological assessments: The utility and challenges of repeat test administrations in clinical and forensic contexts. *The Clinical Neuropsychologist*, *24*(8), 1267–1278.

Ivnik, R.J., Malec, J.F., Smith, G.E., Tangalos, E.G., & Petersen, R.C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R Reading, AMNART, Stroop, TMT and JLO. *The Clinical Neuropsychologist*, *10*, 262–278.

Ivnik, R.J., Malec, J.F., Tangalos, E.G., Petersen, R.C., Kokmen, E., & Kurland, L.T. (1990). The Auditory Verbal Learning Test (AVLT): Norms of ages 55 and older. *Psychological Assessment*, *2*, 304–312.

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

Kent, P.S., & Luszcz, M.A. (2002). A review of the Boston Naming Test and multiple-occasion normative data for older adults on the 15-item versions. *The Clinical Neuropsychologist*, *16*, 555–574.

Lucas, J.A., Ivnik, R.J., Willis, F.B., Ferman, T.J., Smith, G.E., Parfitt, F.C., … Graff-Radford, N.R. (2005). Mayo's Older African Americans Normative Studies: Normative data for commonly used clinical neuropsychological measures. *Clinical Neuropsychology*, *19*(2), 162–183.

Manly, J.J., Jacobs, D.M., Sano, M., Bell, K., Merchant, C.A., Small, S.A., & Stern, Y. (1998). Cognitive test performance among nondemented elderly African Americans and whites. *Neurology*, *50*(5), 1238–1245.

Mitrushina, M., & Satz, P. (1995). Repeated testing of normal elderly with the Boston Naming Test. *Aging Clinical and Experimental Research*, *7*, 123–127.

Pedraza, O., Graff-Radford, N.R., Smith, G.E., Ivnik, R.J., Willis, F.B., Petersen, R.C., & Lucas, J.A. (2009). Differential item functioning of the Boston Naming Test in cognitively normal African American and Caucasian older adults. *Journal of the International Neuropsychological Society*, *15*(5), 758–768.

Rabin, L.A., Barr, W.B., & Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology*, *20*(1), 33–65.

Strauss, E., Sherman, E.M.S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.). New York: Oxford University Press.

Wagner, M.T., Wymer J.H., Carlozzi N.E., Bachman D., Walker A., Mintzer J., & Alzheimer Study Group (2007). Preliminary examination of progression of Alzheimer's disease in a rural Southern African American cohort. *Archives of Clinical Neuropsychology*, *22*(3), 405–414.

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

Zec, R.F., Markwell, S.J., Burkett, N.R., & Larsen, D.L. (2005). A longitudinal study of confrontation naming in the "normal" elderly. *Journal of the International Neuropsychological Society*, *11*, 716–726.

Zec, R.F., Burkett, N.R., Markwell, S.J., & Larsen, D.L. (2007). A cross-sectional study of the effects of age, education, and gender on the Boston Naming Test. *The Clinical Neuropsychologist*, *21*, 587–616.