Journal of the International Neuropsychological Society (2005), **11**, 716–726.

DOI: 10.1017/S1355617705050897

A longitudinal study of confrontation naming in the "normal" elderly

RONALD F. ZEC,[1,2,3] STEPHEN J. MARKWELL,[4] NICOLE R. BURKETT,[1] AND DEB L. LARSEN[1]

[1]Center for Alzheimer Disease and Related Disorders, Southern Illinois University School of Medicine, Springfield, Illinois
[2]Department of Neurology, Southern Illinois University School of Medicine, Springfield, Illinois
[3]Department of Psychiatry, Southern Illinois University School of Medicine, Springfield, Illinois
[4]Statistics and Research Consulting, Southern Illinois University School of Medicine, Springfield, Illinois

(RECEIVED June 4, 2005; ACCEPTED July 22, 2005)

Abstract

The longitudinal effects of age on confrontation naming using the 60-item Boston Naming Test (BNT) were studied in 541 "normal" elderly (ages 50–99). For participants with at least 4 annual assessments ($n = 238$), 150 were followed for ≥6 years, 81 for ≥8 years, and 43 for ≥10 years. A small practice effect (0.21 words, $p = 0.06$) and moderately high test-retest reliability were found when comparing the first 2 assessments, which were 9–15 months apart ($r = 0.76$, $n = 353$). Reliable change index scores indicated that an annual decline of ≥4 points on the BNT is needed for a statistically reliable decline in an individual. A gradient in the mean annual rate of change on the BNT was found with improvement in the 50s age group, no change in the 60s age group, and decline in the 70s and 80s age groups. When projected over 10 years, the magnitudes of the mean changes were relatively small, that is, a 1-word *improvement* for participants in their 50s and a 1.3-word *decline* for participants in their 70s. These findings demonstrate that lexical retrieval as measured by a visual object confrontation naming task is generally well preserved in aging with only subtle decline in the 7th and 8th decades of age. (*JINS*, 2005, *11*, 716–726.)

Keywords: Boston Naming Test, Aging, Cognition, Language, Lexical retrieval, Neuropsychology

INTRODUCTION

Almost all studies examining the effects of "normal" aging on confrontation naming ability as measured by the Boston Naming Test (BNT) have been cross-sectional, whereas only a few have been longitudinal in design (Au et al., 1995; Connor et al., 2004; Cruice et al., 2000; Kent & Luszcz, 2002; Mitrushina & Satz, 1995). A longitudinal research design is the method of choice for studying cognitive change as a function of age, yet because of practical considerations (e.g., expense in terms of time and other resources), the vast majority of publications on cognitive aging are cross-sectional age-comparative studies (Schaie & Hofer, 2001). This design offers the most direct means of assessing how confrontation naming ability (i.e., lexical retrieval) changes with advancing age.

Reprint requests to: Ronald F. Zec, Ph.D., Center for Alzheimer Disease and Related Disorders, Southern Illinois University School of Medicine, 751 North Rutledge, P.O. Box 19643, Springfield, IL 62794-9643. E-mail: rzec@siumed.edu

Methodological limitations of the published longitudinal studies on the effects of "normal" aging on the BNT include short follow-up time periods, small sample sizes, few follow-up assessments, different versions of the BNT being used, and not having administered a comprehensive cognitive assessment to identify and eliminate early dementia cases. The two longitudinal studies with a very short follow-up period found no decline on the BNT with age (Cruice et al., 2000; Mitrushina & Satz, 1995), whereas the three studies with longer follow-up periods did report age-related declines that were greatest in the older elderly (Au et al., 1995; Connor et al., 2004; Kent & Luszcz, 2002). Kent and Luszcz (2002) used a 15-item BNT, and although they concluded that naming ability declines in the older elderly based on 2-year follow-up data, no significant age effects were found when a third assessment at 8 years was included. However, there were considerably fewer participants tested at 8 years due to attrition, and it is possible that these participants may have been more cognitively intact than those who dropped out. Au et al. (1995) also reported age declines, but used the 85-item BNT and examined a

broad age range (30–79 years) with a small sample size and only three assessments. Only one of the three studies reporting age-related decline used the 60-item BNT (Connor et al., 2004), but there were relatively few participants who had multiple assessments.

None of these studies used a comprehensive cognitive evaluation to detect and exclude mild dementia cases. Three studies used the Mini Mental State Exam (MMSE); no participant scored below 27 in the Au et al. (1995) study or below 24 in the Mitrushina and Satz (1995) study. Kent and Luszcz (2002) used a population-based sample and eliminated 110 participants (11.8%) due to impaired MMSE scores. This crude procedure may have eliminated some dementia cases but it is not adequate for screening mild dementia, and it also can result in eliminating nondemented elderly who score in the lower end of the normal distribution.

One reason to study the longitudinal effects of "normal" aging on confrontation naming using the BNT is to improve our ability to reliably detect and differentiate the progressive decline due to a neurodegenerative disorder like Alzheimer's disease (AD) from normal age-related changes. The National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) clinical criteria for AD require that there be two or more areas of cognitive decline confirmed by neuropsychological tests, and the BNT was recommended as a useful measure of language skills (McKhann et al., 1984). The BNT is sensitive to the progressive word-finding difficulty in AD (Bayles et al., 1992; Bowles et al., 1987; Fisher et al., 1999). In a recent study, BNT impairment was found for all severities of AD, including amnestic mild cognitive impairment (MCI), but there were considerable individual differences, such that a clear majority of AD patients did not become impaired until the dementia was moderate to severe (Testa et al., 2004). Although initial BNT impairment was associated with increased risk of a subsequent AD diagnosis, it was not as useful a predictor as impairment in delayed recall (Testa et al., 2004).

In the present study, the longitudinal effects of age on visual object confrontation naming as measured by the BNT were studied in a large cohort of "normal" elderly that included subgroups of participants who were followed for ≥6, ≥8, and ≥10 years and who were tested on multiple annual visits. A comprehensive cognitive evaluation was administered at each assessment to identify and eliminate mild dementia cases, which was not done in previous longitudinal studies examining the effects of "normal" aging on the BNT. Test-retest reliability and practice effects were calculated for the first two annual BNT assessments. Reliable change index scores were calculated to determine the magnitude of the annual decline on the BNT that would be needed for there to be a statistically reliable decline in an individual. The mean annual rate of change on the BNT was determined for different age decades to study the effects of aging on BNT performance, and by inference, on retrieval from semantic memory.

We predicted an increasing rate of decline in confrontation naming on the BNT with successive age decades because of several *possible* factors that may negatively affect this ability (i.e., "normal" brain aging, increasing health problems, decreasing cognitive activity, and increasing incipient AD cases). We also predicted that the magnitude of the decline on the BNT would be relatively small, because this test requires retrieval from semantic memory, that is, from crystallized knowledge, which is relatively well preserved in aging (Horn, 1982a, 1982b; Horn & Donaldson, 1980).

METHOD

Research Participants

After excluding people with neurological conditions that could have possibly had a negative effect on cognitive functioning, there were 541 "normal" (i.e., nondemented) elderly volunteers (ages 50–99 years) who received at least two annual BNT assessments. In Table 1, demographic characteristics are presented for participants whose first and last assessments were compared, for participants whose first and second assessments were 9–15 months apart, and for participants with at least four annual assessments (i.e., the total group of participants, and participants with ≥6, ≥8, and ≥10 years follow-up data). This data represents 2155 annual BNT assessments.

Table 1. Demographics (mean and *SD*) for different groups of older adults (ages 50–99)[a]

Groups	*n*	Age (years)	Education (years)	MMSE	Gender (% female)
Participants whose 1st and last visits were compared	541	67.72 (8.55)	14.72 (2.93)	29.06 (1.16)	65%
Participants whose 1st and 2nd visits were compared (9–15 months apart)	353	67.25 (8.27)	14.63 (2.91)	29.10 (1.09)	67%
All participants with at least 4 visits	238	67.02 (7.50)	14.97 (2.67)	29.15 (1.11)	68%
Group with ≥6 years follow-up[b]	150	66.13 (7.21)	14.89 (2.62)	29.16 (0.94)	67%
Group with ≥8 years follow-up[b]	81	67.07 (6.75)	15.01 (2.62)	29.12 (0.96)	70%
Group with ≥10 years follow-up[b]	43	64.91 (6.82)	14.93 (2.20)	29.27 (1.01)	72%

Note. MMSE = Mini Mental State Exam.
[a]Demographic data for the first assessment.
[b]Participants who had 4 or more annual assessments.

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

This sample of convenience consisted of participants living independently in the community who were recruited into a study on cognitive aging from a variety of sources. Sources included spouses of patients evaluated at our memory clinic, newspaper advertisements, presentations given to various groups, and by word of mouth.

Virtually all participants were Caucasian. Previous studies indicate that African Americans score lower on the BNT (Manly et al., 1998), and thus we excluded from these analyses the two African Americans in our sample with longitudinal data. Another participant was excluded because English was not her first language. An additional 103 participants were excluded because they met exclusion criteria of having a neurological and/or psychiatric condition that was judged to put the participant at high risk for cognitive impairment at baseline. Fifty-five of these 103 participants were excluded because they met cognitive/clinical criteria at baseline for dementia or amnestic MCI (i.e., mild cognitive impairment in which the memory impairment was severe for their age) or declined to meet criteria during the course of the study. For eight additional participants, initial assessments were included, but later assessments were not because they later developed a neurological or psychiatric condition that could impair cognitive test performance (e.g., Parkinson's disease, stroke).

Procedure

The 60-item Boston Naming Test (Kaplan et al., 1983) *without cueing* was administered as part of a comprehensive neuropsychological battery in a study of cognitive aging. All 60 line drawings of objects were presented, that is, no discontinuation after a series of incorrect answers, and a total score of spontaneous correct responses was calculated. We did *not* employ the cueing procedures, which are part of the standardized BNT administration, in which incorrect answers are followed up with a semantic cue and, if still incorrect, with a phonemic cue. Without the cueing procedure, the BNT is relatively quick to administer to older people even if they have naming difficulty. The mean of spontaneous correct responses on the BNT (i.e., without cueing) was significantly lower in AD patients, including those with mild dementia, compared to elderly controls (Zec et al., 1992). Responses were scored according to instructions in the test manual, except that "French harp" and "false face" were accepted as regional substitutes for harmonica and mask. In addition, the 20-second time limit for a response was not strictly observed, but these "normal" participants rarely exceeded this time limit.

Statistics

Paired *t* tests were used to compare BNT performance on the first two annual assessments to determine practice effects. Pearson product-moment correlation coefficients assessed test-retest reliability between the first two assessments and between number of days between the first and last assessments by using the change score on the BNT. Reliable change index scores using one-tailed 95% confidence intervals were calculated to determine how large a difference is needed between two assessments to be 95% confident that the decline is significant, that is, it is not due to chance. Reliable change indices are based on the standard error of measurement of the difference between the first two BNT assessments (Becker & Markwell, 2000). Linear regressions were fit for each participant who had four or more assessments to determine the annual rate of change. Mean rates of change were computed for the total group and for different demographic groups. Independent groups *t* tests were used to compare the "decliner" and "nondecliner" groups in terms of BNT score, Alzheimer Disease Assessment Scale (ADAS) score, age, and education at baseline. Chi-square tests of independence were used to compare gender ratios. Results were considered statistically significant at $p < .05$.

RESULTS

Practice Effects and Test-Retest Reliability

A small practice effect on the BNT (0.21 and 0.19 words) that approached statistical significance ($p = .06$) was found for older adults (ages 50–99) between the first and second assessments, both when the test-retest interval was between 9–15 months and when comparing the first two assessments regardless of when the second occurred (Table 2). Test-retest reliability between the first two assessments was moderately high and statistically significant ($r = .76$ and .80 for the two test-retest intervals) (Table 2). These high test-retest reliabilities are especially impressive given the "ceiling" effect on the BNT for this normal sample in which the vast majority of scores fell in a relatively narrow range at the upper end of the distribution. By comparison, the test-retest reliability for the MMSE was very low and nonsignificant ($r = .19$, $p = .83$) due to restriction of range.

There was no change on the BNT (i.e., no *overall* aging effect) when comparing scores at baseline and the *last* assessment (Table 2). Furthermore, the correlation of the number of days between the first and last assessments with the change score on the BNT was *not* statistically significant ($r = .0571$, $p = .1850$). This indicates that the BNT score did *not* significantly decline as a *linear* function of aging (i.e., number of elapsed days) for the total group.

In summary, there was a nonsignificant trend for a very small practice effect when comparing the first two BNT assessments with moderately high test-retest reliability. There was no evidence of change on the BNT when the baseline and *last* assessment were compared.

Reliable Change Index

The NINCDS-ADRDA work group established clinical criteria for AD that require documentation of a progressive decline in cognitive functioning and point out that progressive worsening on neuropsychological tests can be estab-

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

Table 2. Practice effects and test-retest reliability on the Boston Naming Test (BNT) for older adults (ages 50–99)

Test-retest comparisons	*n*	BNT first visit[a]	BNT second visit[a]	Mean change score	Mean time interval	Test-retest reliability
1st and 2nd visit (9–15 months apart)[b]	353	55.78 (3.09)	55.99 (3.08)	0.21, $p = .06$	388.7 days (1.06 yrs)	$r = .76$, $p < .0001$
1st and 2nd visit (regardless of interval)[c]	540	55.53 (3.73)	55.73 (3.76)	0.19, $p = .06$	545.3 days (1.49 yrs)	$r = .80$, $p < .0001$
1st and last visit[c]	541	55.53 (3.73)	55.51 (4.00)	−0.02, $p = .85$	1593 days (4.36 yrs)	

[a]Mean (*SD*).
[b]There was only one participant in their 90s.
[c]There were only three participants in their 90s.

lished by comparison with a patient's previous performance (McKhann et al., 1984). Thus, it is important to know what would constitute a statistically reliable decline on the BNT. In Table 3, reliable change index (RCI) scores are presented for participants who had follow-up testing within 9–15 months of the baseline. These RCI scores indicate for a given individual how much the BNT score must decrease to be significant at the 95% confidence level. For the total group, an individual's BNT score must decrease by 4 or more points to be significant, that is, to be confident that the decrease was not due to chance. A minimum 4-point decline on the BNT is also needed to reach statistical significance for older adults at each of three educational levels, for both genders, and for participants in their 60s and 70s. A minimum 3-point decline is needed to be significant for participants in their 50s and a 5-point decline is needed for participants in their 80s. In summary, an annual decline of at least 4 points on the BNT is generally needed for there to be a statistically reliable decline for an individual participant.

Table 3. Reliable change index (RCI) scores (at the 95% confidence level) for older adults (ages 50–99)[a] who had a follow-up assessment within 9–15 months of the first assessment

Group	*n*	*SD*[b]	*r*[c]	*SEM*[d]	Reliable change index
Age group (years)					
50–59	59	2.90	0.83	1.19	2.80
60–69	169	2.56	0.69	1.41	3.31
70–79	92	3.32	0.80	1.50	3.52
80–89	32	4.11	0.78	1.91	4.59
Education (years)					
<12	16	4.49	0.93	1.23	3.04
=12	110	3.30	0.76	1.60	3.76
>12	227	2.71	0.72	1.44	3.36
Gender					
Females	235	3.17	0.76	1.57	3.66
Males	118	2.89	0.78	1.34	3.14
Overall	353	3.08	0.76	1.49	3.48

[a]There was only one participant in their 90s.
[b]*SD* = standard deviation for the mean score on the BNT from the second assessment.
[c]*r* = correlation coefficient (reliability) between first and second assessments.
[d]*SEM* = standard error of measurement of a single assessment.

Mean Annual Rates of Change for Participants with Four or More Assessments

In Table 4, the mean annual rate of change (ARC) on the BNT is presented for age decades (50–89 years), three educational levels, gender, and the total group of participants. Linear regressions were fit for each participant ($n = 238$) who had four or more annual assessments, to determine the mean ARC for the total group and for the different demographic groups. For the total group, the mean ARC on the BNT was −0.033 words for older adults who were tested a mean of 7.17 years after the first assessment (i.e., only a projected 0.33-word decline over a 10-year period). With respect to decade of age, the mean ARC on the BNT was small but positive for the 50–59 years age group (a projected 1.31-word improvement over 10 years). The mean ARC was near zero (no change with age) for the 60–69 years age group (i.e., 0.03 words over 10 years), and was small but negative for the 70–79 and 80–89 years age groups (i.e., a decline of 1.57 and 1.46 words over 10 years). With respect to educational level, there was a small mean annual decline with age in the 12 years educational group and an even smaller annual decline in the >12 years group, and almost no change in the <12 years group, but the sample size was small ($n = 5$). With respect to gender, there was a small mean annual rate of decline for females, but almost no change for males. In summary, the longitudinal BNT data indicates only a 0.33-word projected decline over 10 years for all participants, but a mean projected decline of 1.6 words for the 70–79 age group and 1.5 words for the 80–89 age group.

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

Table 4. Mean (*SD*) annual rate of change on the Boston Naming Test (BNT) based on linear regressions for older adults (ages 50–89) with four or more annual assessments

Group	*n*	Age (years)	Education (years)	BNT score at first visit	Mean years between first and last visit	Mean annual rate of change on the BNT
Age group (years)						
50–59	33	55.79 (2.57)	15.91 (2.49)	56.79 (2.46)	7.74	+0.131 (0.35)
60–69	124	64.52 (2.87)	14.71 (2.65)	56.57 (2.82)	7.27	+0.003 (0.36)
70–79	65	73.83 (2.83)	14.91 (2.80)	56.10 (2.92)	7.03	−0.157 (0.40)
80–89	16	81.94 (2.29)	15.25 (2.32)	52.56 (8.33)	5.82	−0.146 (0.72)
Education (years)						
<12	5	68.80 (5.81)	9.20 (0.84)	53.60 (3.21)	5.54	+0.015 (0.75)
=12	61	67.23 (6.56)	12.00	56.02 (3.43)	7.20	−0.054 (0.39)
>12	172	66.90 (7.87)	16.19 (2.06)	56.34 (3.61)	7.21	−0.027 (0.41)
Gender						
Females	161	67.06 (7.83)	14.63 (2.47)	55.99 (3.72)	7.25	−0.056 (0.38)
Males	77	66.94 (6.80)	15.66 (2.94)	56.65 (3.17)	7.01	+0.014 (0.46)
Overall	238	67.02 (7.50)	14.97 (2.67)	56.20 (3.57)	7.17	−0.033 (0.41)

Mean Annual Rates of Change for Participants with Different Follow-up Periods

The mean ARC on the BNT is presented in Table 5 for participants in their 50s, 60s, 70s, and 80s with ≥4, ≥5, and ≥6 visits and with ≥5, ≥6, ≥8, and ≥10 years follow-up testing. Only participants with ≥4 visits were included in these analyses. In general, participants in their 50s had a positive mean ARC (i.e., improved over time), whereas participants in their 70s had a negative mean ARC (i.e., declined over time). Participants in their 80s generally had a negative mean ARC, but the sample sizes were too small to be considered reliable. The mean ARC for participants in their 60s was small in magnitude (close to zero) and inconsistent in direction (3 positive and 4 negative annual rates of change) indicating little mean change over time. The magnitude of change for participants in their 50s was considerably smaller for the ≥8 and ≥10 year follow-up periods than for the shorter follow-up periods, but there were only 13 and 9 participants in these shorter follow-up groups. It is also possible that with the longer follow-up periods, participants in their 50s at the time of the later assessments would be well into their 60s, when there is little change on the BNT.

In summary, the mean annual rates of change indicate improvement in BNT scores over time for participants who were in their 50s at baseline, worsened scores for participants in their 70s and 80s, and very small inconsistent changes for participants in their 60s. Thus, there is a gradient in the mean ARC on the BNT with improving scores in the 50s age group, no change in the 60s, and declining scores in the 70s and 80s. However, when projected over 10 years, the magnitudes of the changes for participants in their 50s and 70s were relatively small, that is, approximately a 1-word *improvement* for participants in their 50s and a 1.3-word *decline* for participants in their 70s when averaged across all follow-up groups (Table 5).

Individual Annual Rates of Change for Participants with Ten or More Years Follow-up

The majority of participants who were followed for 10 or more years with at least 4 assessments (25 of 43 participants, i.e., 58%) displayed a *negative* ARC on the BNT (worsened scores over time), including 56%, 45%, and 82% of participants who were 50–59, 60–69, and 70–79 years old at baseline (Table 6). As discussed earlier, the *mean* ARC on the BNT for participants with ≥10 years follow-up data (Table 5) were very small (near zero) but positive for participants in their 50s and 60s, and small but negative for participants in their 70s. There was an outlier ARC score in the 70–79 years age group, and thus the median score of −0.056 words per year likely provides a more accurate measure of central tendency. In summary, a considerably higher percentage of participants in their 70s had a negative ARC on the BNT than participants in their 50s or 60s.

Baseline Demographic Comparison of BNT "Decliners" *versus* "Nondecliners"

In an attempt to identify variables that might contribute to a participant being a "decliner" *versus* a "nondecliner" on the

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

Table 5. Mean (*SD*) annual rate of change on the BNT based on linear regressions for older adults (ages 50–89) with ≥4, ≥5, and ≥6 visits and ≥5, ≥6, ≥8, and ≥10 years between the first and last assessment

Age groups	*n*	Age (years)	Education (years)	BNT score at first visit	Mean years between first and last visit	Mean annual rate of change on the BNT
≥4 visits						
50–59	33	55.79 (2.57)	15.91 (2.49)	56.79 (2.46)	7.74	+0.131 (0.35)
60–69	124	64.52 (2.87)	14.71 (2.65)	56.57 (2.82)	7.27	+0.003 (0.36)
70–79	65	73.83 (2.83)	14.91 (2.80)	56.10 (2.92)	7.03	−0.157 (0.40)
80–89	16	81.94 (2.29)	15.25 (2.32)	52.56 (8.33)	5.82	−0.146 (0.72)
≥5 visits						
50–59	27	55.81 (2.66)	15.63 (2.47)	56.74 (2.44)	8.25	+0.110 (0.21)
60–69	95	64.46 (2.84)	14.89 (2.46)	56.59 (2.92)	8.09	−0.048 (0.31)
70–79	46	73.57 (2.73)	14.74 (2.74)	55.80 (3.18)	8.28	−0.093 (0.40)
80–89	8	81.25 (2.05)	15.75 (2.49)	55.63 (2.97)	6.90	−0.283 (0.51)
≥6 visits						
50–59	24	56.17 (2.46)	15.79 (2.50)	56.58 (2.43)	8.30	+0.136 (0.20)
60–69	68	64.49 (2.79)	14.84 (2.52)	56.72 (3.03)	8.91	−0.042 (0.26)
70–79	39	73.54 (2.61)	14.77 (2.84)	56.33 (2.99)	8.55	−0.087 (0.36)
80–89	5	80.20 (0.45)	16.40 (2.61)	54.80 (3.42)	8.18	−0.033 (0.35)
≥5 years						
50–59	28	55.79 (2.62)	15.61 (2.51)	56.75 (2.40)	8.42	+0.090 (0.23)
60–69	96	64.43 (2.83)	14.66 (2.51)	56.53 (2.96)	8.29	−0.058 (0.29)
70–79	45	73.67 (2.75)	14.56 (2.71)	55.91 (3.20)	8.43	−0.110 (0.40)
80–89	8	80.88 (2.10)	15.25 (2.82)	53.63 (6.61)	7.71	+0.112 (0.55)
≥6 years						
50–59	25	55.96 (2.65)	15.48 (2.43)	56.60 (2.48)	8.75	+0.101 (0.22)
60–69	81	64.44 (2.85)	14.80 (2.59)	56.69 (2.91)	8.81	−0.041 (0.28)
70–79	36	73.69 (2.83)	14.58 (2.81)	56.22 (3.12)	9.13	−0.165 (0.37)
80–89	8	80.88 (2.10)	15.25 (2.82)	53.63 (6.61)	7.71	+0.112 (0.55)
≥8 years						
50–59	13	56.31 (2.63)	14.77 (2.55)	56.54 (2.99)	10.59	+0.027 (0.20)
60–69	42	64.24 (2.84)	15.07 (2.56)	56.68 (3.33)	10.47	+0.005 (0.23)
70–79	24	73.42 (2.70)	14.88 (2.88)	56.21 (3.28)	10.38	−0.171 (0.34)
80–89	2	80.00	17.00 (1.41)	55.50 (3.54)	10.04	−0.244 (0.39)
≥10 years						
50–59	9	56.33 (2.83)	15.22 (2.64)	57.11 (3.14)	11.44	+0.020 (0.22)
60–69	22	63.59 (2.79)	15.05 (2.06)	56.36 (4.16)	11.97	+0.041 (0.15)
70–79	11	73.18 (2.09)	14.36 (2.29)	54.73 (4.00)	12.01	−0.114 (0.30)
80–89	1	80.00	16.00	58.00	11.45	−0.518

BNT, we compared the "decliner" and "nondecliner" groups for each of the three follow-up periods (i.e., ≥6, ≥8, and ≥10 years) on age, education, gender, BNT score, and ADAS total error score at baseline (Table 7). The "decliner" groups for all three follow-up periods were somewhat older and less educated and had somewhat higher mean total ADAS error scores than the "nondecliner" groups. The age and education difference was statistically significant for the ≥6 years follow-up group, and the mean total ADAS error score was significant for the ≥10 years follow-up group. Thus, differences in age or education may partially explain why some participants tended to decline on the BNT while others did not. For all three follow-up periods, "decliners" had a *higher* mean BNT score at baseline than "nondecliners," and the differences were statistically significant for the ≥6 and ≥8 year follow-up periods. In summary, the "decliner" groups for all three follow-up periods had *better* BNT scores at baseline than the "nondecliner" groups despite being somewhat older and less educated, and also having somewhat *poorer* mean ADAS total error scores.

Comparison between First and Last Assessments for Individual Participants

Individual and mean difference scores on the BNT for participants with ≥10, ≥8, and ≥6 years between their first and last assessment are presented in Tables 8, 9, and 10. There was generally a small *mean improvement* on the BNT between the first and last assessment for participants in their 50s and 60s, but a *mean decline* for participants in their 70s regardless of time between assessments. The number of participants in the 80s age group was too small to

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

Table 6. Annual rate of change on the Boston Naming Test (BNT) based on linear regressions for each older adult with 10 or more years between the first and last assessments[a] (positive values indicate improvement and negative values indicate decline)

Annual Rate of Change on the BNT for Individuals in Different Age Groups			
50–59 years	60–69 years	70–79 years	80–89 years
+0.294	+0.381	+0.193	−0.518
+0.214	+0.272	+0.110	
+0.171	+0.227	−0.029	
+0.142	+0.215	−0.033	
−0.008	+0.148	−0.049	
−0.024	+0.111	−0.056	
−0.089	+0.089	−0.063	
−0.090	+0.070	−0.072	
−0.430	+0.053	−0.087	
	+0.038	−0.214	
	+0.031	−0.959	
	+0.001		
	−0.004		
	−0.004		
	−0.008		
	−0.016		
	−0.018		
	−0.026		
	−0.045		
	−0.097		
	−0.137		
	−0.375		

[a]Only participants with 4 or more assessments were included in this analysis.

Table 8. Frequency[a] of difference scores between the first and last assessments on the Boston Naming Test (BNT) for each older adult with ≥10 years follow-up ($n = 44$)

	Age Group (years)			
Difference[b]	50–59 ($n = 9$)	60–69 ($n = 22$)	70–79 ($n = 12$)	80–89 ($n = 1$)
+8		1		
+7				
+6				
+5				
+4	2	4		
+3		1	1	
+2	1		2	
+1	2	3	1	
0	1	5	4	
−1	1	5		
−2	1	2	2	
−3			1	
−4		1		1
−5	1[c]			
−6				
−7				
−8				
−9			1	
Mean difference	+0.444	+0.773	−0.667	—

[a]Frequency of participants who displayed a given difference score.
[b]Positive values indicate improvement and negative values indicate decline.
[c]Participants represented below the dashed lines had a statistically significant decline on the BNT based on the Reliable Change Index (RCI).

Table 7. Baseline demographic comparison of Boston Naming Test (BNT) "decliners" *vs.* "nondecliners" [mean (*SD*)] based on the annual rate of change for older adults (ages 50–89) who had ≥6, ≥8, and ≥10 years between the first and last assessments

Time between initial and last assessments	*n*	Age (years)	Education (years)	Gender (% female)	BNT	ADAS[a]
≥6 years						
Nondecliners	70	64.33 (6.78)	15.39 (2.62)	43:27 (61.4%)	55.73 (3.70)	3.40 (1.84)
Decliners	80	67.70 (7.25)	14.45 (2.56)	57:23 (71.3%)	56.98 (2.62)	3.62 (1.45)
p value[b]		.0039*	.0286*	.2030	.0207*	.4368
≥8 years						
Nondecliners	36	64.44 (5.67)	15.44 (2.79)	23:13 (63.9%)	55.69 (3.48)	2.78 (1.52)
Decliners	45	67.38 (7.30)	14.67 (2.45)	34:11 (75.6%)	57.11 (2.88)	3.43 (1.38)
p value[b]		.0514	.1860	.2532	.0484*	.0637
≥10 years						
Nondecliners	18	63.06 (4.57)	15.17 (2.23)	13:5 (72.2%)	55.06 (4.28)	2.36 (1.65)
Decliners	25	66.24 (7.88)	14.76 (2.20)	18:7 (72.0%)	56.92 (3.48)	3.76 (1.41)
p value[b]		.1032	.5557	.9872	.1228	.0108*

[a]Total error score for the Alzheimer Disease Assessment Scale (ADAS).
[b]*p* values based on *t* tests for age, education, BNT, and ADAS, and on chi-square for gender.
*Statistically significant differences.

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

Table 9. Frequency[a] of difference scores between the first and last assessments on the Boston Naming Test (BNT) for each older adult with ≥8 years follow-up (n = 78)

	Age Group (years)			
Difference[b]	50–59 (n = 13)	60–69 (n = 39)	70–79 (n = 24)	80–89 (n = 2)
+8		1		
+7				
+6				
+5				
+4	2	5		
+3	1	1	1	
+2	2	6	2	
+1	3	6	3	
0	2	8	7	1
−1	1	6	2	
−2	1	3	3	
−3			2	
−4		2	1	1
−5	1[c]		1	
−6				
−7				
−8				
−9			1	
−10				
−11		1		
−12			1	
Mean difference	0.769	0.256	−1.417	−2.000

[a]Frequency of participants who displayed a given difference score.
[b]Positive values indicate improvement and negative values indicate decline.
[c]Participants represented below the dashed lines had a statistically significant decline on the BNT based on the Reliable Change Index (RCI).

Table 10. Frequency[a] of difference scores between the first and last assessments on the BNT for each older adult with ≥6 years follow-up (n = 150)

	Age Group (years)			
Difference[b]	50–59 (n = 27)	60–69 (n = 79)	70–79 (n = 36)	80–89 (n = 8)
+8		1		1
+7				
+6				
+5				
+4	3	6	1	
+3	4	3	1	
+2	3	10	2	1
+1	7	8	5	1
0	4	17	10	1
−1	4	13	2	1
−2	1	11	5	1
−3		4	4	1
−4		3	2	1
−5	1[c]	1	1	
−6				
−7				
−8			1	
−9		1	1	
−10				
−11		1		
−12			1	
Mean difference	0.963	−0.190	−1.389	0.125

[a]Frequency of participants who displayed a given difference score.
[b]Positive values indicate improvement and negative values indicate decline.
[c]Participants represented below the dashed lines had a statistically significant decline on the BNT based on the Reliable Change Index (RCI).

draw conclusions for this group. One third of the participants who were followed for ≥10 years and ≥8 years, and 41% who were followed for ≥6 years displayed lower BNT scores on their last assessment compared to their first. However, based on the reliable change index, only a small minority of participants overall and in each age decade group displayed *statistically reliable* declines on the BNT between their first and last assessment, regardless of follow-up time period (6.8%, 10.3%, and 8.7% for the ≥10, ≥8, and ≥6 year follow-up periods, respectively). The percentage of participants with *statistically significant* lower scores on the last assessment increased with successively higher decades of age from 50 to 79 years for the ≥6 and ≥8 year follow-up groups, but not for the ≥10 year follow-up group.

DISCUSSION

Longitudinal effects of age on visual confrontation naming of line drawings of objects as measured by the BNT were studied in a large cohort of "normal" elderly that included subgroups of participants who were followed for ≥6, ≥8, and ≥10 years. Test-retest reliability was moderately high, whereas practice effects were very small. Reliable change index scores indicated that an annual decline of at least 4 points on the BNT is generally needed for a statistically reliable decline in an individual, which is useful information when following suspected early dementia cases.

Our prediction of an increasing rate of decline in confrontation naming ability as measured by the BNT with successive age decades was supported by the results of this study, but the age-related change was nonlinear rather than the linear relationship expected. A gradient in the mean annual rate of change on the BNT was found as a function of age with improvement in the 50s age group, no change in the 60s age group, and decline in the 70s and 80s age groups. Nonlinear changes in cognition with age have been previously reported (Albert et al., 1987).

Our additional prediction that the mean magnitude of the decline on the BNT with age would be relatively small was also confirmed. For example, when projected over 10 years, there was approximately a 1.0-word *improvement* for participants in their 50s and a 1.3-word *decline* for participants in their 70s. We predicted that the size of the decrement on

https://doi.org/10.1017/S1355617705050897 Published online by Cambridge University Press

the BNT with age would be small because the BNT measures retrieval from semantic memory (i.e., crystallized knowledge), which is relatively well preserved in aging (Horn, 1982a, 1982b; Horn & Donaldson, 1980).

The finding in the present study of age-related declines on the BNT in the older elderly is generally consistent with the findings from three previously published longitudinal studies that had relatively long follow-up periods (Au et al., 1995; Connor et al., 2004; Kent & Luszcz, 2002). The age decline on the BNT in the present study was relatively small, which is consistent with the subtle decline reported by Connor et al. (2004). Larger BNT declines with age were reported by Au et al. (1995), but that study had a very small sample size. In our study, a small improvement over time was found in the 50s age group and no change in the 60s age group. These results differ from the findings of Au et al. (1995) and Connor et al. (2004) who report some mean decline in these younger age groups (albeit less of a decline than in the older elderly). This may be a result of the differences in the participant samples (e.g., educational levels, overall health). For example, although the mean educational level was high in all three studies, it was somewhat higher in the present study (~15 years) than in the other two studies (~14 years), which may have contributed to the differences in findings. Alternatively, methodological differences may account for the different findings. For example, there were more frequent repeated annual assessments in the present study, and thus perhaps a greater practice effect masked a subtle mean decline in this younger elderly age range. In any case, the present study and the three longitudinal studies with relatively long follow-up periods report BNT declines in the older elderly.

There are several *possible* variables that may be contributing to the decline in naming ability in the older elderly, including "normal" brain aging, increasing health problems, decreasing cognitive activity, and increasing incipient AD cases. Changes in brain structure and function that occur in "normal" aging may be the primary cause of the decline in confrontation naming. Among the known brain changes in normal aging are significant neuronal loss in subcortical nuclei that influence production of neurotransmitters and project widely to the cortex, including the substantia nigra, basal forebrain, and locus coeruleus, which give rise to dopaminergic, cholinergic, and noradrenergic neuronal systems, respectively (Albert, 2002; Kemper, 1993). In addition, there is substantial structural change in cortical white matter, including alterations in myelin, abnormal inclusions, and substantial increase in lipofuscin (Nielsen & Peters, 2000). In the hippocampus, there is significant neuronal loss in the hilus (CA4) and subiculum and significant changes in *N*-methyl-D-aspartate (NMDA) receptor levels and late phase long-term potentiation (Morrison & Hof, 1997). One or more of these known brain changes in "normal" aging (or perhaps yet to be discovered changes) may be responsible for the accelerated decline in confrontation naming with age found in this study.

Disease is another possible contributing factor to the age decline in confrontation naming. Cumulative medical problems and associated medication use increase on average with age (e.g., high blood pressure, diabetes) and could contribute to cognitive declines (Arvanitakis et al., 2004; Glynn et al., 1999) including potentially on the BNT. A third possible contributing factor to cognitive decline is disuse due to decreasing levels of environmental complexity with age, for example, retirement from occupations to typically less cognitively challenging activities (Schaie, 1980; Schooler, 1990; Zec, 1995).

In addition to aging per se, disease, and disuse, another possible explanation for the decline on the BNT in the 70s and 80s age groups is that there may have been a greater number of participants in the incipient stages of dementia in these older age groups. This might be expected given that the prevalence of AD increases with advancing age with rates doubling approximately every 5 years over the age of 65 (Jorm et al., 1987). However, a strength of the present study is that all participants were assessed with a comprehensive neuropsychological battery for dementia, and those who met cognitive criteria for dementia or amnestic MCI (i.e., severe memory impairment) at baseline and/or at follow-up were excluded from the data reported in this article. Nonetheless, it is still possible that some participants in the incipient/prodromal stages of dementia were not excluded. However, the BNT tends not to be very sensitive to the very early clinical/cognitive changes in AD patients (Testa et al., 2004). Consequently, there is likely to be only very subtle effects on BNT performance in prodromal AD participants and this would be expected to have little or no effect on the mean BNT score in our study.

Given these major possible causal/contributing factors to an increasing rate of cognitive decline with age, the magnitude of the changes in confrontation naming with age found in this study were relatively small (e.g., a projected 1.3-word decline over 10 years on the BNT for the 70s age group). Thus, one major conclusion of this study is that in "normal" aging, confrontation naming as measured by the BNT undergoes a surprisingly small decline given possible undetected prodromal AD cases, known and unknown changes in brain structure and function associated with "normal" aging, increases in medical problems, and decreases in challenging cognitive activities. These findings demonstrate that retrieval from semantic memory (i.e., lexical retrieval) as measured by a visual object confrontation naming task is generally well preserved with age in that there is only a subtle decline in the 7th and 8th decades of age.

It is interesting and surprising that the "decliner" groups for all three follow-up periods had higher (i.e., better) BNT scores at baseline than the "nondecliner" groups despite being somewhat older and less educated, and having somewhat higher (i.e., poorer) mean ADAS total error scores. Because the mean BNT score was higher at baseline in the "decliner" groups, despite these participants being somewhat older, less educated, and having a poorer ADAS total

score, the longitudinal decline on the BNT in the "decliner" groups may in part represent regression to the mean. However, the improvement on the BNT in the "nondecliner" groups may also in part represent regression to the mean. Thus, regression to the mean likely occurs in both directions resulting in declining and improving scores that tend to cancel each other out. However, there is a "ceiling" effect on the BNT resulting from many normal elderly scoring nearly perfectly, which likely results in regression to the mean favoring declining scores on the BNT, that is, there is little room for improvement for many participants. Consequently, regression to the mean may account for some of the overall decline on the BNT (see Schaie & Hofer, 2001). Further evidence supporting regression to the mean is the finding that the mean annual rate of change for the ADAS total error score was in the direction of improvement for both the "decliner" and "nondecliner" groups for all three follow-up periods (i.e., >6, >8, and >10 years) (Table 7). Thus, the declining BNT scores in the "decliner" group was *not* part of a general cognitive decline in this group as measured by the total ADAS score.

The limitations of this study include that the findings are based on a Caucasian sample in the Central Midwest with a high mean educational level for whom English was their first language. Consequently, these findings cannot be generalized to other racial groups, samples with low educational levels, or to those for whom English is not their first language. Furthermore, these findings are limited to visual object confrontation naming as measured by the BNT. Increasing complaints of word-finding difficulties with age (Albert et al., 1988) is generally consistent with the findings in this study of a decline in BNT performance in the 70s and 80s age groups, although the magnitude of this decline was smaller than expected given these complaints. Perhaps object naming is better preserved than other types of confrontation naming, such as retrieving people's names (Cohen & Faulkner, 1986).

There are both inherent strengths and weaknesses to using a longitudinal design (Schaie, 2002; Schaie & Hofer, 2001). A major advantage is that cohort effects (i.e., year of birth) usually are less of a potential confounding factor than in studies that use a cross-sectional design to examine the effects of age on cognitive functioning. Another advantage of the longitudinal design is that each participant serves as his/her own control. The major disadvantage of this design is a potential selective attrition bias introduced by participants who do not continue with the follow-up assessments. Typically, a greater percentage of participants who are declining cognitively will drop out of the study and thereby minimize the effects of age on cognitive functioning. As has been reported for other cognitive measures, larger BNT declines have been found with increasing age when using cross-sectional *versus* longitudinal data (Connor et al., 2004; Cruice et al., 2000). Cross-sectional age comparisons on the BNT will be discussed in a subsequent article on this data set and compared with the findings from these longitudinal analyses.

ACKNOWLEDGMENTS

This work was supported by Southern Illinois University School of Medicine, Center for Alzheimer Disease and Related Disorders (CADRD). Many students, researchers, faculty, and staff of the CADRD have made important contributions to this longitudinal cognitive aging study.

REFERENCES

Albert, M.S. (2002). Memory decline: The boundary between aging and age-related disease. *Annals of Neurology*, *51*, 282–284.

Albert, M., Duffy, F.H., & Naeser, M. (1987). Nonlinear changes in cognition with age and their neuropsychologic correlates. *Canadian Journal of Psychology*, *41*, 141–157.

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

Arvanitakis, Z., Wilson, R.S., Bienias, J.L., Evans, D.A., & Bennett, D.A. (2004). Diabetes mellitus and risk of Alzheimer disease and decline in cognitive function. *Archives of Neurology*, *61*, 661–666.

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

Bayles, K.A., Tomoeda, C.K., & Trosset, M.W. (1992). Relation of linguistic communication abilities of Alzheimer's patients to stage of disease. *Brain and Language*, *42*, 454–472.

Becker, R. & Markwell, S. (2000). Problems arising from the generalising of treatment efficacy from clinical trials in Alzheimer's disease: Mistaking statistical significance as clinical significance. *Clinical Drug Investigations*, *19*, 33–41.

Bowles, N.L., Obler, L.K., & Albert, M.L. (1987). Naming errors in healthy aging and dementia of the Alzheimer type. *Cortex*, *23*, 519–524.

Cohen, G. & Faulkner, D. (1986). Memory for proper names: Age differences in retrieval. *British Journal of Developmental Psychology*, *4*, 187–197.

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

Fisher, N.J., Rourke, B.P., & Bieliauskas, L.A. (1999). Neuropsychological subgroups of patients with Alzheimer's disease: An examination of the first 10 years of CERAD data. *Journal of Clinical and Experimental Neuropsychology*, *21*, 488–518.

Glynn, R.J., Beckett, L.A., Hebert, L.E., Morris, M.C., Scherr, P.A., & Evans, D.A. (1999). Current and remote blood pressure and cognitive decline. *The Journal of the American Medical Association*, *281*, 438–445.

Horn, J.L. (1982a). The aging of human abilities. In B.B. Wolman (Ed.), *Handbook of developmental psychology*, pp. 847–870. Englewood Cliffs, NJ: Prentice-Hall.

Horn, J.L. (1982b). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F.I.M. Craik & S. Trehub (Eds.), *Aging and cognitive processes*, Vol. 8, pp. 237–278. New York: Plenum Press.

Horn, J.L. & Donaldson, G. (1980). Cognitive development: 2. Adulthood development of human abilities. In O.G. Brim &

I. Kagan (Eds.), *Constancy and change in human development*. Cambridge, MA: Harvard University Press.

Jorm, A.F., Korten, A.E., & Henderson, A.S. (1987). The prevalence of dementia: A quantitative integration of the literature. *Acta Psychiatrica Scandinavica*, *76*, 465–479.

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

Kemper, T.L. (1993). The relationship of cerebral cortical changes to nuclei in the brainstem. *Neurobiology of Aging*, *14*, 659–660.

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

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*, 1238–1245.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlin, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Heath and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*, 939–944.

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

Morrison, J.H. & Hof, P.R. (1997). Life and death of neurons in the aging brain. *Science*, *278*, 412–419.

Nielsen, K. & Peters, A. (2000). The effects of aging on the frequency of nerve fibers in rhesus monkey striate cortex. *Neurobiology of Aging*, *21*, 621–628.

Schaie, K.W. (1980). Intellectual change in adulthood (trans., Zeit). *Gerontology*, *15*, 373–384.

Schaie, K.W. (2002). The impact of longitudinal studies on understanding development from young adulthood to old age. In W. Hartup & R.K. Silbereisen (Eds.), *Growing points in developmental science: An introduction* (pp. 307–328). Cambridge, UK: Psychology Press.

Schaie, K.W. & Hofer, S.M. (2001). Longitudinal studies in research on aging. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (5th ed.), pp. 53–77. San Diego, CA: Academic Press.

Schooler, C. (1990). Psychosocial factors and effective cognitive functioning in adulthood. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed.), pp. 347–358. San Diego, CA: Academic Press.

Testa, J.A., Ivnik, R.J., Boeve, B., Petersen, R.C., Pankratz, V.S., Knopman, D., Tangalos, E., & Smith, G.E. (2004). Confrontation naming does not add incremental diagnostic utility in MCI and Alzheimer's disease. *Journal of the International Neuropsychological Society*, *10*, 504–512.

Zec, R.F. (1995). Neuropsychology of aging. *Experimental Gerontology*, *30*, 431–442.

Zec, R.F., Vicari, S., Andrise, A., Belman, J., & Landreth, E. (1992). A comparison of confrontational and generative naming in Alzheimer patients and elderly controls. New York Neuropsychology Group Meeting, New York. *The International Journal of Neuroscience*, *63*, 245.