Neuropsychological test performance in a cognitively intact sample of older Japanese American adults

Susan M. McCurry\textsuperscript{a,*}, Laura E. Gibbons\textsuperscript{b}, Jay M. Uomoto\textsuperscript{c}, Mary Lou Thompson\textsuperscript{b}, Amy B. Graves\textsuperscript{d}, Steven D. Edland\textsuperscript{b}, James Bowen\textsuperscript{b}, Wayne C. McCormick\textsuperscript{b}, Eric B. Larson\textsuperscript{b}

\textsuperscript{a}Department of Psychosocial and Community Health, University of Washington, Box 357263, Seattle, WA 98195-7263, USA
\textsuperscript{b}University of Washington, Seattle, WA 98195, USA
\textsuperscript{c}Seattle Pacific University, Seattle, WA, USA
\textsuperscript{d}University of South Florida, Tampa, FL, USA

Accepted 30 December 1999

Abstract

The purpose of this paper was to present population-based data showing the effects of age on cognitive test performance in a sample of older Japanese American adults. In addition, the relative effects of education, gender, and primary spoken language were compared to effects that have been reported in the literature for majority culture older adults. Subjects included 201 non-demented Japanese American adults age 70 and older currently enrolled in the Kame Project, a prospective study of aging and dementia in King County, WA. Cognitive tests included the Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) neuropsychological assessment battery, WAIS-R Digit Span and Digit Symbol subtests, Trail Making Test, Purdue Pegboard, and Finger Tapping. Older age was associated with significantly ($p < 0.05$) lower scores on all tests; less than high school education was associated with lower scores on all tests except Digit Span, Finger Tapping, and the Purdue Pegboard. Women and English-speaking participants scored higher than men and Japanese speakers on various tests of memory, attention, and visuomotor ability. These data reinforce the importance of using appropriately corrected norms when interpreting results of cognitive screening tests with

\textsuperscript{*} Corresponding author. Tel.: +1-206-685-9113; fax: +1-206-685-9551.

E-mail address: smccurry@u.washington.edu (S.M. McCurry).

\textsuperscript{1} Portions of this paper were originally presented at the Gerontological Society of America 51st annual meeting, November 20–24, 1998, Philadelphia, PA.
minority culture older adults. © 2001 National Academy of Neuropsychology. Published by Elsevier Science Ltd.

Keywords: Neuropsychological; Cognitive; Japanese American

1. Introduction

Although increasing numbers of older adults are referred for geropsychological assessment, there remains a lack of normative data relevant to older individuals (La Rue & Markee, 1995). It is well known that psychometric test scores are affected by age and education (Heaton et al., 1991). There is also growing recognition of the impact of ethnic and socioeconomic diversity on tests of cognitive or psychological function (Loewenstein et al., 1993; Geisinger, 1994). Such multicultural considerations may be particularly important among minority culture older adults who are screened for possible cognitive impairment, since unrecognized language and cultural biases can lead to the false-positive identification of dementia (Gurland et al., 1992). It is expected that by the year 2030, one-quarter of older adults in the US will be ethnic minorities (APA, 1997). Thus, it will become increasingly important to have additional information on the use and interpretation of neuropsychological tests with culturally diverse groups.

Normative studies of psychometric test performance are expensive and difficult to conduct, and as a result, published data for older adults are often based on small sample sizes, clinical populations, or samples of convenience. One alternative source of test performance data comes from population-based epidemiological studies of aging and dementia, which include cognitive or functional assessment measures as part of their screening and disease verification protocols. These data can be used to develop instrument norms so long as appropriate adjustments are made for any selection biases in the sampling design (Thompson et al., submitted for publication). Collaborative international studies offer particularly rich opportunities to examine the use of cognitive assessment instruments with older adults in diverse cultural settings (Launer, 1992; Tuokko et al., 1995; Larson et al., 1998).

The Consortium to Establish a Registry for Alzheimer’s Disease (CERAD) neuropsychological test battery (Morris et al., 1989) is one of the most widely used protocols for cognitive screening in epidemiological studies of aging. The CERAD battery is considered an internationally compatible minimal data set for clinical dementia assessments (Stahelin et al., 1997). Performance on the CERAD tests has been described for a variety of ethnic and cultural groups, including African American, Hindi, Spanish, French, and Jamaican cohorts (Mahurin et al., 1992; Demers et al., 1994; Guruje et al., 1995; Welsh et al., 1995; Ganguli et al., 1996; Unverzagt et al., 1999). The purpose of the present study is to describe the use of an expanded version of the CERAD battery with a non-demented, community-dwelling cohort of older Japanese Americans. Data are presented for two age groups: 70 to 79, and 80 years and older. For each age group, the effects of key confounding variables (education, gender, and language) on cognitive test performance are discussed. It is, to our knowledge, the first such data available for persons of Japanese background in the US. As such, relevant norms for this particular ethnic population can potentially benefit both research and clinical work in dementia.
2. Methods

2.1. Subjects

Data included in this study were drawn from the Kame Project, a prospective study designed to examine the prevalence and incidence rates of dementia among older Japanese American adults in Seattle-King County, WA. The study population included persons age 65 or older, who were of at least 50% Japanese heritage. Between May 1, 1992 and May 1, 1994, a total of 1993 eligible, community- and institution-dwelling individuals participated in the baseline screening for prevalent dementia. Of these, 453 were sampled for follow-up clinical assessment based upon their age and performance on the Cognitive Abilities Screening Instrument (CASI) (Teng et al., 1994), a 100-point cognitive screening test developed for cross-national studies of dementia. A two-stage stratified design was used for the follow-up sampling, where the strata were composed of three CASI classes ( < 81, 81–86.9, 87+) and five age groups (65–69, 70–74, 75–79, 80–84, 85+). Details regarding community population characteristics and the original sampling algorithm have been reported elsewhere (Graves et al., 1996).

A final sample of 384 subjects ultimately participated in the follow-up clinical assessment. Because only three people were sampled from the youngest, highest functioning age group, estimates for the 65–69 age group are not stable. The present paper will present data for the sample of non-demented subjects aged 70 years or older.

2.2. Procedures

All subjects who participated in the follow-up clinical evaluation underwent additional interviews with proxy informants, a physical and neurological examination, and a standardized neuropsychological and functional screening battery that was administered by a trained psychometrician and interpreted by a geriatric psychologist. Completed evaluations were presented and discussed at diagnostic consensus meetings attended by a neurologist, geriatrician, psychologist, epidemiologist, and research staff. Of the 384 subjects who were evaluated, 235 were not demented; of these, 201 were aged 70 and older, and are included in the present analysis (Table 1). The average subject age was 79.6 years (range = 70–101); average education level was 11.0 years (range = 5–20). Interviews (including both test instructions and subject responses) were conducted in either English or Japanese, based on the subjects’ primary language and speaking preference. In this sample, slightly less than half of interviews were conducted in Japanese (41.8%) or a mixture of Japanese and English (6.2%).

2.3. Neuropsychological testing

Testing was designed to assess the basic cognitive functions affected by aging and dementia, including aspects of memory, attention/concentration, language, visuospatial functioning, and psychomotor speed. The core of the cognitive screening battery was the CERAD neuropsychological test protocol (Morris et al., 1989). Additional tests were chosen for their proven validity and reliability, and their wide use in cognitive screening batteries.
The issue of cultural bias has been discussed relative to existing neuropsychological tests (Wong et al., in press) and as such, tests were also chosen that possessed characteristics that minimized such bias. Tests that possessed significant English verbal content or ones that assumed knowledge of Western or American culture were not included. Those that were considered for study were weighted toward non-verbal constructs, or where a Japanese translation was accurate enough to fairly reflect the content of the test (e.g., word list recall tasks). Japanese versions of all tests were developed as part of a cross-national study of dementia in Japanese older adults (Larson et al., 1998). Test administration instructions and materials were originally translated by bilingual interviewers, then back-translated into English by two professional translators for content comparisons against standard English instructions and materials, as well as between Japanese versions used at the three participating sites (Honolulu, HI; Tokyo and Hiroshima City, Japan).

2.3.1. CERAD neuropsychological assessment battery

As a measure of language fluency, subjects were asked to name as many animals as possible in 1 min. The subject’s score is the total number of animals named during this interval. Confrontational naming was assessed by showing subjects 15 line drawings from the Boston Naming Test that represent objects with a range of naming difficulty and frequency of occurrence in the English language.

Subjects’ ability to learn new verbal information was assessed over three learning trials with a 10-word list; in each trial, subjects read the list of common nouns aloud, then immediately remembered as many words from the list as possible. The summed total of the three learning trials is the memory acquisition score. After a brief delay with a figure construction distraction task, subjects were again asked to recall the 10-item word list, which is the memory recall score. A verbal memory percent savings score was also computed to show the relative amount of verbal information retained over the delay interval (\([\text{Verbal recall score/Learning Trial 3}] \times 100 = \% \text{ Savings}\)).

Between the immediate and delayed word list recall, subjects copied four line drawings (circle, diamond, intersecting rectangles, cube) as a measure of constructional praxis. Subjects
were also asked to recall and reproduce the four line drawings immediately following the word list recall, providing a construction recall score. A visual memory percent savings score was subsequently computed to show the relative amount of visual information retained over the delay interval ([(Construction recall score/Construction praxis] × 100 = % Savings).

2.3.2. Additional neuropsychological tests
Three tests of sustained attention were completed by subjects. In the WAIS-R Digit Span subtest (Wechsler, 1981), subjects repeated strings of digits presented orally both forwards and backwards; the test is considered a measure of immediate memory capacity, attention, and concentration. The WAIS-R Digit Symbol subtest (Wechsler, 1981) is a test of concentration and visual scanning that requires subjects to match random numbers with an associated nonsense symbol, copying the symbols as quickly as possible in a space below each listed number. Both Digit Span and Digit Symbol are presented in this paper using both raw and age-corrected scaled scores. The Trail Making Test — Part A (Reitan & Wolfson, 1985) is a timed test during which subjects draw a line connecting randomly positioned numbers in consecutive order (i.e., 1–2–3, etc.). In this study, Trails A was administered in the standard fashion except that the test was discontinued once subjects reached five errors, or if the test was not completed within 5 min.

Motor speed and dexterity were evaluated using the Purdue Pegboard and the Finger Tapping Tests. For the Purdue Pegboard, subjects completed two 30-s trials with each hand; for Finger Tapping, an average score for five 10-s trials was computed for each hand (Spreen & Strauss, 1991).

Some additional tests of cognitive functioning were included in the assessment interview that are not described here, including memory for paired associates, similarities, comprehension, mental arithmetic, and vocabulary questions. Tasks were excluded either because of missing data (for example, the vocabulary test was not given to Japanese-speaking participants), or the small range of scores for a given task (for example, the arithmetic test included only four items).

2.4. Statistical analysis
To adjust for the stratified sampling design, each sampled individual’s test score had to be weighted by the inverse of the sampling fraction for that stratum. For example, there were 80 subjects at baseline aged 75–79 with CASI 81–86.9; 24 of these subjects were sampled in the clinical assessment, so the scores for these individuals were hence weighted by 80/24. All subjects who had a CASI score ≤ 81 or who were over age 85 were selected for sampling, so their scores received a weight of 1. The numerator of each sampling fraction was based on the 384 subjects who agreed to participate, rather than the 455 who were selected for sampling.

The mean, standard deviation, median, and 25th and 75th percentiles of each test score were estimated for the two age groups. These analyses were carried out using the SAS procedure PROC UNIVARIATE which required rounding of the weights to the nearest integer. The median and quartile scores were estimated from the weighted empirical distribution function. If a quartile fell between two test scores, we chose the convention of assigning the lower (or slower) score.
The overall sample was too small to generate reliable norms on subgroups other than age (for example, age by education, or age by gender). However, the independent effects of age, education, gender, and language on mean test scores were evaluated as follows. For each test, the appropriate link function was selected. Weighted ordinal logistic regression was used for constructional praxis. For all other tests, normality assumptions were tenable for the original data or transformed data (only Trails A time was transformed), and weighted linear regression was used. In all modeling, the integral sampling weights were used. Next, the four main

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
<th>25th</th>
<th>Median</th>
<th>75th</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>70–79-years-old (N = 120)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>15.4</td>
<td>4.7</td>
<td>13</td>
<td>14</td>
<td>18</td>
</tr>
<tr>
<td>Confrontational naming</td>
<td>13.3</td>
<td>1.5</td>
<td>12</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Word list acquisition</td>
<td>17.6</td>
<td>4.2</td>
<td>15</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>Word list recall</td>
<td>6.0</td>
<td>2.2</td>
<td>4</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Word recall savings (%)</td>
<td>80.6</td>
<td>22.7</td>
<td>66</td>
<td>85</td>
<td>100</td>
</tr>
<tr>
<td>Constructional praxis</td>
<td>10.2</td>
<td>0.9</td>
<td>10</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Construction recall</td>
<td>8.1</td>
<td>2.4</td>
<td>7</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>Construction savings (%)</td>
<td>80.7</td>
<td>25.4</td>
<td>63</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Trail making, Part A (s)</td>
<td>45.3</td>
<td>16.3</td>
<td>48</td>
<td>44</td>
<td>34</td>
</tr>
<tr>
<td>Digit span (raw score)</td>
<td>11.8</td>
<td>3.5</td>
<td>9</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Digit span (scaled score)</td>
<td>9.1</td>
<td>2.6</td>
<td>7</td>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>Digit symbol (raw score)</td>
<td>41.2</td>
<td>9.9</td>
<td>33</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Digit symbol (scaled score)</td>
<td>12.4</td>
<td>2.5</td>
<td>10</td>
<td>12</td>
<td>15</td>
</tr>
<tr>
<td>Pegboard (dominant hand)</td>
<td>13.7</td>
<td>2.0</td>
<td>13</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Pegboard (non-dominant hand)</td>
<td>12.7</td>
<td>2.3</td>
<td>11</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Finger tapping (dominant hand)</td>
<td>42.9</td>
<td>8.1</td>
<td>38</td>
<td>44</td>
<td>49</td>
</tr>
<tr>
<td>Finger tapping (non-dominant hand)</td>
<td>39.4</td>
<td>6.1</td>
<td>36</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td><strong>80 + years-old (N = 81)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal fluency</td>
<td>12.3</td>
<td>3.7</td>
<td>10</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>Boston Naming Test</td>
<td>12.2</td>
<td>2.6</td>
<td>11</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Word list acquisition</td>
<td>14.1</td>
<td>3.9</td>
<td>12</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>Word list recall</td>
<td>4.6</td>
<td>3.0</td>
<td>2</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Word recall savings (%)</td>
<td>72.3</td>
<td>45.0</td>
<td>33</td>
<td>80</td>
<td>100</td>
</tr>
<tr>
<td>Constructional praxis</td>
<td>9.7</td>
<td>1.3</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Construction recall</td>
<td>4.9</td>
<td>2.9</td>
<td>2</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Construction savings (%)</td>
<td>50.4</td>
<td>30.3</td>
<td>20</td>
<td>54</td>
<td>72</td>
</tr>
<tr>
<td>Trail making, Part A (s)</td>
<td>79.8</td>
<td>49.2</td>
<td>93</td>
<td>67</td>
<td>44</td>
</tr>
<tr>
<td>Digit span (raw score)</td>
<td>10.4</td>
<td>2.3</td>
<td>9</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Digit span (total scaled score)</td>
<td>8.1</td>
<td>1.9</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Digit symbol (raw score)</td>
<td>32.0</td>
<td>12.6</td>
<td>24</td>
<td>31</td>
<td>37</td>
</tr>
<tr>
<td>Digit symbol (scaled score)</td>
<td>10.2</td>
<td>3.0</td>
<td>8</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>Pegboard (dominant hand)</td>
<td>12.5</td>
<td>2.3</td>
<td>11</td>
<td>12</td>
<td>14</td>
</tr>
<tr>
<td>Pegboard (non-dominant hand)</td>
<td>11.3</td>
<td>1.9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Finger tapping (dominant hand)</td>
<td>36.4</td>
<td>9.2</td>
<td>30</td>
<td>36</td>
<td>45</td>
</tr>
<tr>
<td>Finger tapping (non-dominant hand)</td>
<td>34.2</td>
<td>7.7</td>
<td>28</td>
<td>36</td>
<td>41</td>
</tr>
</tbody>
</table>
effects (age, education, gender, and language) were regressed on the test scores and any significant factors were retained. Finally, interactions between main effects with $p < 0.10$ were assessed.

3. Results

The neuropsychological test weighted means, standard deviations, and quartile distributions for each of the two age groups (70–79 years, or “younger-old”; and 80 + years, or “older-old” subjects) are shown in Table 2. The mean scores for all tests were significantly different ($p < 0.05$) between the two age groups: as expected, older age was associated with poorer performance on all tests, as well as greater variability in score distributions. For example, the inter-quartile range for delayed word list recall, when adjusted for amount of material acquired (percent savings) was 66 to 100 for the younger-old subjects, and 33 to 100 for the older-old subjects. Age differences were greatest for Construction Recall and the Trail Making Test; the older-old subjects on average remembered almost half as many drawings (mean = 4.9 vs. 8.1 for older-old and younger-old, respectively), and were almost 50% slower on the timed Trails test than the younger-old participants (mean = 72.3 vs. 45.3 s, respectively).

In addition to age, the effects of education, gender, and primary spoken language on mean test scores were examined using multivariate regression procedures (Table 3). Controlling for all other statistically significant factors, less than high school education was associated with

Table 3
Independent effects of age, education, gender, and language on mean neuropsychological test scores, including percentage of variance in test scores (cumulative $AR^2$)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Significant effects$^a$</th>
<th>Cumulative $AR^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal fluency</td>
<td>Age, education</td>
<td>11</td>
</tr>
<tr>
<td>Confrontational naming</td>
<td>Age, education</td>
<td>28</td>
</tr>
<tr>
<td>Word list acquisition</td>
<td>Age, education, gender, language</td>
<td>42</td>
</tr>
<tr>
<td>Word list recall</td>
<td>Age, education, gender</td>
<td>26</td>
</tr>
<tr>
<td>Word recall savings (%)</td>
<td>Age, education</td>
<td>2</td>
</tr>
<tr>
<td>Constructional praxis</td>
<td>Age, education, language</td>
<td>n/a$^b$</td>
</tr>
<tr>
<td>Construction recall</td>
<td>Age, education, gender</td>
<td>22</td>
</tr>
<tr>
<td>Construction savings (%)</td>
<td>Age, education, gender</td>
<td>24</td>
</tr>
<tr>
<td>Trail making, Part A (s)</td>
<td>Age, education, gender</td>
<td>29</td>
</tr>
<tr>
<td>Digit span (raw scores)</td>
<td>Age, gender</td>
<td>7</td>
</tr>
<tr>
<td>Digit span (scaled scores)</td>
<td>Age, gender</td>
<td>6</td>
</tr>
<tr>
<td>Digit symbol (raw scores)</td>
<td>Age, education, gender, language</td>
<td>42</td>
</tr>
<tr>
<td>Digit symbol (scaled scores)</td>
<td>Age, education, gender, language</td>
<td>41</td>
</tr>
<tr>
<td>Pegboard (dominant hand)</td>
<td>Age, gender</td>
<td>13</td>
</tr>
<tr>
<td>Pegboard (non-dominant hand)</td>
<td>Age, gender</td>
<td>9</td>
</tr>
<tr>
<td>Finger tapping (dominant hand)</td>
<td>Age</td>
<td>9</td>
</tr>
<tr>
<td>Finger tapping (non-dominant hand)</td>
<td>Age</td>
<td>8</td>
</tr>
</tbody>
</table>

$^a$ For all reported effects, significantly lower (or slower) test scores were associated with older age, less than high school education, male gender, and Japanese-speaking language.

$^b$ $AR^2$ not applicable to logistic regression.
significantly ($p < 0.05$) lower mean scores for all cognitive tests except for tests of motor speed and dexterity (Purdue Pegboard and Finger Tapping) and auditory attention (WAIS-R Digit Span subtest). Gender effects were observed for several tests, with female subjects scoring significantly higher than males on mean word list acquisition and recall, construction recall and percent savings, Trail Making, WAIS-R Digit Span and Digit Symbol, and the Purdue Pegboard. Finally, subjects tested in English scored significantly higher on word list acquisition, constructional praxis, and the WAIS-R Digit Symbol subtest than did Japanese-speaking subjects. The relative combined contributions of age, education, gender, and language preference to test performance varied widely, ranging from an adjusted $R^2$ of only 0.02 for the word list percent savings score, up to 0.42 for word list acquisition and the Digit Symbol subtest (Table 3).

4. Discussion

Research describing the use of cognitive testing with the very old and with ethnic minority individuals has been identified as a top priority for the diagnosis and treatment of neuropsychiatric disorders in older adults (La Rue & Markee, 1995; Heaton et al., 1996). The purpose of this paper is to provide clinicians with needed tools and preliminary standards for the cognitive evaluation of older Japanese American adults in the US. To our knowledge, this is the first report that provides neuropsychological test performance data stratified by age for this understudied segment of the US geriatric population, using a moderately large sample of 201 subjects age 70 or older.

As expected, results confirm that neuropsychological tests in the expanded CERAD battery used for this study are age-sensitive. The age effects noted here were similar, although not identical, to those that have been reported among majority culture older adults (Heaton et al., 1986; Spreen & Strauss, 1991; Elias et al., 1997; Moses & Pritchard, 1999). Subjects over age 80, a group for whom little normative data are available in any education or cultural group, had significantly lower scores on all tests than subjects in the 70–79 age range, with the greatest differences observed for visual-motor tests of memory and attention (construction recall and percent savings, Trails A). Similar age-related declines in visuospatial abilities have been widely described, particularly for timed tests that measure complex attention and motor speed (Salthouse, 1985; Heaton et al., 1991; Rasmussen et al., 1998). However, in this sample, there were also observed age effects in areas of cognitive function that are often considered to be relatively preserved with aging, such as simple attention (digit) span or constructional praxis (La Rue, 1992). One explanation might be that many studies of cognitive function with elderly, community-dwelling participants have been based on smaller sample sizes or subjects with higher average education (Howieson et al., 1993; Welsh et al., 1994; Corey-Bloom et al., 1996). The current findings emphasize the need for availability of normative test data on a wide distribution of older persons. This is particularly apropos given the rapidly growing numbers of older-old individuals from diverse sociocultural and educational backgrounds in the general population.

Tests in the expanded CERAD battery also were differentially affected by education, gender, and preferred language of administration. Our sample distributions were too small to
present reliable norms on subgroups other than age (for example, only 23 of the 81 oldest-old subjects were men). Nevertheless, it is of interest to consider which of these key covariates were shown in regression analyses to impact test performance, and how their effects compare to those in other populations. It is well known that formal education influences performance on most cognitive tests, and that normal older, less educated individuals often score below standard cutoffs for impairment on neuropsychological screening measures (Leckliter & Matarazzo, 1989; Mortimer & Graves, 1993; Osterweil et al., 1994; Marcopulos et al., 1997).

In the current sample, persons with a high school education or beyond had significantly higher mean scores on all cognitive tests with the exception of the WAIS-R Digit Span subtest, and the tests measuring motor speed and dexterity (Purdue Pegboard, Finger Tapping). These results are consistent with studies which show that Purdue Pegboard performance tends to be unrelated to educational level (Spreen & Strauss, 1991). Although Digit Span and Finger Tapping speed (which were not significant in our sample) have been shown elsewhere in the literature to correlate with increased education (Warner et al. 1987), the amount of variance in Digit Span and Tapping Test scores attributable to education effects for these tests is relatively low (Heaton et al., 1991).

Women scored significantly higher than men on mean word list acquisition and recall, construction recall and percent savings, the Trail Making Test, WAIS-R Digit Symbol and Digit Span, and the Purdue Pegboard. The superior performance of women in this study for word list acquisition and recall is consistent with research which has found that women tend to perform better than men on some tests of verbal ability, including verbal recall (Inouye et al., 1993). Other studies have found that men tend to do better on tests involving manipulating spatial relationships (Heaton et al., 1986; Leckliter & Matarazzo, 1989; Mazaux et al., 1995; Ylikoski et al., 1998), whereas we found that women had superior scores on several visuomotor tasks. Subjects tested in English in this study scored significantly higher on word list acquisition, constructional praxis, and the WAIS-R Digit Span subtest than did Japanese-speaking subjects. The tendency for non-English speakers to perform more poorly on Digit Span has been observed in other studies (Loewenstein et al., 1993), and may be due to language-specific differences in the number of syllables that must be held in working memory and repeated in digit repetition tasks. However, the poorer performance on word list acquisition and constructional recall is surprising in light of the fact that the Japanese language is based on both pictographs and ideographs as well as phonetic characters, which might be expected to enhance visual encoding skills for both verbal and nonverbal material. Additional studies are needed to examine the effects of spoken language on attention and memory test performance in this population, and to demonstrate the psychometric properties of the Japanese version of the CERAD battery used in this study.

In addition to actual linguistic factors, there may be unknown acculturation factors that also contributed to lower test scores in this sample of elderly Japanese speakers. Recent research has suggested that maintenance of a more traditional Japanese lifestyle may be associated with a decreased longitudinal risk for cognitive decline in older Japanese American adults (Graves et al., 1999). In the current study, choice of Japanese language for the testing interview was associated with endorsement of a number of variables indicative of a more traditional Asian lifestyle, such as preference for Japanese friends, foods, and movies, as well as poorer overall performance on various tests of memory, attention, and visuomotor ability. It may be that
Japanese speakers who have maintained a more traditional lifestyle have had less opportunity to become familiar with certain problem-solving strategies or cognitive skills demanded by tests such as the CERAD battery which were developed on majority culture subjects. The present sample was not large enough to test for cultural effects on test performance directly. However, we did compare test scores for Kame Project subjects between the ages of 70 and 89 who had less than 12 years education (N = 67; data not presented), with published scores for persons in the same age and education range reported in the original CERAD study (Welsh et al., 1994). Scores for Kame participants were consistently lower than those reported in the CERAD paper, and at least 1 standard deviation lower for confrontational naming, word list recall, and word list percent savings. Although circumstantial, these findings suggest that factors other than age and education, including the effects of primary spoken language and cultural variables, must be considered when evaluating cognitive test performance in clinical or research samples. Further research detailing the effects of cultural factors on test performance in minority culture populations is needed.

The following limitations and caveats should be noted when using these data. First, the neuropsychological test performance of persons sampled in the Kame Project may not be representative of all Japanese American elders because of differences in a variety of geographic, socioeconomic, and acculturation factors. Almost half (47%) of subjects received all or part of their grade school education in Japan, and a quarter (27%) attended high school outside of the US. Although observed education effects were largely consistent with those reported for majority culture samples, it is unknown to what extent the level of education reported for persons schooled in Japan is comparable to the US and can be represented with a simple numerical value. In addition, the subject sample of Japanese Americans in the present investigation is bicultural in nature rather than cross-cultural. That is to say, most of the subjects have lived on the West coast of the US for many years and therefore, have had the opportunity to acculturate to Western norms. It remains an empirical question as to what extent forces of acculturation influence cognitive styles, strengths, and weaknesses.

Second, the neuropsychological test scores reported here for non-demented subjects were obtained as part of a dementia screening evaluation. The decisions to classify these subjects as non-demented were based, in part, on these test scores. Thus, they should not be used as cut-off scores for identification of dementia without further empirical verification in an independent sample. Nevertheless, they do provide a reference distribution against which clinicians or researchers can compare their own findings with the CERAD battery and other tests here described.

Finally, subjects in this study were not excluded on the basis of common medical diseases that affect many older adults (e.g., diabetes, hypertension); however, they were carefully worked up for exclusion of significant cognitive impairment and as such, can be considered representative of the full spectrum of older adults with cognitive abilities that are considered to be functioning within normal limits. The subjects sampled for this phase of the Kame study were non-demented as defined by enrollment criteria and consensus review, but only at a single point in time. Given their advanced age, some of the non-demented subjects included in this sample will go on to become demented; we do not know the extent to which such subjects’ neuropsychological test scores were “preclinical” at the time of sampling. Nevertheless, given the current lack of data for older minority populations, results from the present
study should provide a useful comparison for clinicians or researchers when evaluating the
cognitive function of demographically similar clients.

In summary, our findings confirm that in addition to age and education, at the minimum,
gender and primary spoken language should be considered when evaluating cognitive test
scores in older Japanese Americans. This is particularly important in the case of screening for
suspected dementia in very old persons from different cultural backgrounds, since cutoff
scores commonly used for younger majority culture individuals may lead to elevated false-
positive misclassification rates. The average scores obtained in this study for the younger,
more educated, English-speaking groups were frequently below the cutoff scores that have
been proposed for many tests as indicative of cognitive impairment (Bornstein, 1986).
Additional future studies are needed to evaluate whether the pattern of results found among
Kame subjects is representative of older Japanese American adults in the US. Nevertheless,
these findings provide a reminder of the importance of developing demographically
appropriate norms when using assessment measures originally developed and standardized
among younger, well-educated Caucasian populations. In particular, clinical diagnostic
interpretations of cognitive test scores in older adults suspected of having an underlying
dementia should be made cautiously and in the context of other supporting cognitive and
functional data.

Acknowledgments

This study was supported by a grant from the National Institute on Aging #AG09769. The
authors wish to acknowledge Kame Project coordinator Nancy Zee, as well as participating
medical personnel, interviewers, and psychometrists: Dr. Brian Ito, Masaru Aoyama, Patti
Tsubota Boorkman, Yoshiko Ogasawara Heather, Lynne Salter, Jan St. John, and Manami
Wakuta. Data and technical assistance were provided by Duane Beekly, Ross Barnhart,
Malkah DuPrix, Mary Jacka, and Madeline Murguia Rice. Finally, we wish to thank Greta
Hoshibata, Joyce Nakamura, Megumi Sumitani, and the Japanese American community of
Seattle-King County, without whom this study would not have been possible.

References

Psychological Association (Brochure Working Group on the Older Adult).
Bornstein, R. A. (1986). Classification rates obtained with “standard” cut-off scores on selected neuro-
clinical and neuropsychological instruments into French: the CERAD experience. Age Ageing 23(6),
and gender on cognitive performance in the Framingham heart study: community-based norms. Exp
Aging Res 23(3), 201–235.


Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., vanBelle, G., Fillenbaum, G., Mellits, E. D., & Clark,


