

Measuring Memory in Large Group Settings Using a Continuous Recognition Test

J. Wesson Ashford, Emily Gere and Peter J. Bayley*

Stanford/VA Aging Clinical Research Center, VA Palo Alto Health Care System, Palo Alto, CA, USA

Accepted 9 August 2011

Abstract. Memory function generally deteriorates with age, and memory impairments are a common symptom of serious illness such as dementia. Although screening tests are widely used throughout Medicine, they are not yet commonly used to detect memory impairments. The objective of this study was to characterize an audience-based memory test suitable for administration to a large number of individuals simultaneously. A continuous recognition test was developed to assess memory function in audiences using a slide-show in which 50 images were presented, of which 25 were repeated. Audience members responded by recording if an image was a repetition. The test was administered to a total of 1018 participants at 25 sites with an average audience size of 41 individuals (range = 9–142). A total of 868 individuals aged 40–97 y completed the test appropriately and provided their age, education level, and gender. Recognition memory as measured by discriminability (d') showed a significant decline with age (40–49 y old, $d' = 3.51$; 90–99 y old, $d' = 1.95$, $p < 0.001$) together with a greater than three-fold increase in variability. Individuals with less than 13 y of education had lower scores than those with more education ($d' = 2.13$ vs. 2.88, respectively, $p < 0.001$). These results are consistent with the known effects of age and education on memory. There were no significant effects of gender on test performance. Such memory tests represent a practical and novel approach to screen for the signs of early dementia.

Keywords: Alzheimer's disease, episodic memory, memory screening, object recognition, recognition memory, visual memory

Supplementary data available online: <http://www.j-alz.com/issues/27/vol27-4.html#supplementarydata02>

INTRODUCTION

Among the many causes of memory impairment, dementia is one of the most common [1–4]. Epidemiological studies indicate that approximately 14% of adults aged 71 and older exhibit signs of dementia [5]. However, despite the broad understanding of the association between memory disorders and dementia, memory disorders often go undiagnosed [6–11].

Several factors interfere with the detection of memory impairment associated with dementia, including a failure to screen, avoidance of the problem by affected individuals, a lack of information available to

healthcare providers about screening, and under use of available testing methods. With changes in the delivery of healthcare, physicians must work under strict time constraints, leading many physicians to not routinely screen their patients for memory problems related to dementia [12].

One solution to the failure to detect memory impairments due to dementia is to implement large-scale community memory screening programs. Numerous instruments have been developed to screen for such memory problems [13, 14], yet few have been used in community screening programs. The few published studies illustrate some of the difficulties in screening for memory impairment in the community. For example, Lawrence et al. [15] used two cognitive screening instruments, the Mini-Mental State Exam (MMSE) and the Seven-Minute Screen (7MS) to screen 59

*Correspondence to: Peter J. Bayley, VA Palo Alto Health Care System, 3801 Miranda Avenue, Mailcode 151Y, Palo Alto, CA 94304-1290, USA. Tel.: (650) 493 5000 × 68653; Fax: (650) 852 3297; E-mail: peter.bayley@va.gov.

participants in a community screening program. The MMSE [16], though showing poor sensitivity and specificity in dementia detection, has become a benchmark against which other dementia screening instruments are often compared. The MMSE is composed of items measuring orientation, memory, attention, language, and praxis. The 7MS is a brief neuropsychological test measuring temporal orientation, cued recall, category fluency, and clock drawing. Screening results [15] suggest that 29% of the 59 participants had signs of impairment on the 7MS, yet 47% of these participants had relatively normal MMSE scores (>24), illustrating that the 7MS and the MMSE have different sensitivities and specificities for the detection of early dementia in a community setting. In a larger study, the 7MS was used to screen 497 individuals [12]. Results support the conclusion that community screening programs can detect individuals who are previously unknown to have cognitive problems. However, it was concluded that the screening program was costly in terms of volunteer staff time and resources. Furthermore, only a small number of previously undiagnosed individuals were diagnosed with dementia indicating that this form of screening may be inefficient. These two community studies illustrate some of the difficulties of performing memory screening programs. They are logistically difficult, some currently available screening tests have poor sensitivity and specificity for screening purposes, and are not specific to memory assessment, and trained psychometricians are needed to administer tests in a one-to-one interaction.

Another approach to screening for memory disorders is to use computerized neuropsychological tests. Computerized tests have sought to overcome some of the difficulties inherent in large-scale screening programs for dementia, including offering objectivity, sensitivity and efficiency (for review see [17]). However, most programs still require an administrator who is costly and may bias the analysis. For example, one technician-administered office-based computer assessment has been shown to have good validity in assessing cognitive function in early dementia [18–20], but a trained clinician is still required to interact with the subject and monitor the use of the computer. The issue of administration by a trained professional has been addressed with the use of self-administered computerized screening tests [21]. Such tests have shown good sensitivity in detecting memory problems. Nonetheless, one such battery takes up to 45 minutes and necessitates comparatively complex computer interaction which may preclude its use in cognitively-impaired individuals.

To our knowledge audience-based memory assessment methods for screening large numbers of individuals simultaneously have not been utilized as cognitive screening tools. However, such tests could be used to screen large numbers of people for memory problems in order to identify high-risk individuals for further evaluation. Because memory impairment associated with dementia is commonly undiagnosed, a simple audience-based test designed to detect patients with early dementia would be valuable. Any population of older adults will contain individuals with diverse memory impairments and a significant issue in designing an audience-based screening test is what kind of memory problem should be screened. One answer is to design a screening test sensitive to detecting Alzheimer's disease (AD) which is the most common form of dementia, accounting for approximately two thirds of all dementia cases [22]. The initial symptom of AD is typically a prominent memory dysfunction in which the core problem is difficulty in encoding new information [23–27].

The process of memory encoding can be tested in several different ways. However, recognition memory tests are especially suitable for detecting encoding difficulties as they provide the target stimuli within the test framework. Thus, poor performance on a test of recognition memory provides evidence for an underlying encoding impairment, raising the possibility of an emerging dementia process [28]. In contrast to individuals with early dementia, healthy adults can quickly and accurately encode massive amounts of new information. For example, landmark studies from the 1960s, 1970s, and 1980s demonstrate that healthy individuals perform well above chance on tests of recognition memory after viewing thousands of images for a few seconds each [29–31].

Even though the presence of dementia has a profound influence on memory, it is well known that demographic variables including age, education, and gender also have a significant interaction with recognition memory [32, 33]. Indeed, numerous studies show that recognition memory performance declines with age even in healthy adults [34, 35]. The association between education, gender, and recognition memory may be less robust. For example, Diesfeldt [36] reported that among normal elderly subjects, neither gender nor education correlates significantly with the Recognition Memory Test [37]. However, in standardization studies [37], age, gender, and education are associated with recognition test performance.

The current study aimed to assess the suitability of an audience-based memory test as a screening

Table 1
Demographic characteristics of participants

Age group	40–49		50–59		60–69		70–79		80–89		90–99	
<i>n</i>	29		68		135		239		359		38	
Gender (% F)	76		82		76		65		66		63	
	<i>M</i>	(<i>SD</i>)										
Age (y)	44.8	3.1	55.3	2.8	65.8	2.9	75.8	2.9	84.4	2.7	92.1	1.6
Education (y)	16.9	2.0	15.9	2.3	16.4	2.3	15.9	2.7	16.1	2.5	15.6	2.6

test for early dementia. Although audience testing is widely used in educational assessment, audience testing procedures are unusual in cognitive neuroscience or clinical research. Accordingly, the primary aim of the study was to demonstrate that an audience based recognition memory test can be successfully administered to a large number of individuals from a relevant population. A recognition format was chosen due to the robust ability of healthy adults to perform this kind of task, and because it is sensitive to the encoding deficits found in early dementia. We chose a continuous recognition test (CRT) format in which participants were asked to detect item repetitions. The study sought to characterize the task and to evaluate task performance according to age, education, and gender. CRT performance was expected to decrease with age, whereas education and gender were not expected to influence test performance.

METHODS

Participants

The CRT was administered to 1018 subjects between July 2007 and June 2008 at 25 sites (community events, senior citizen centers, retirement living communities, etc., in the San Francisco Bay Area). Audience sizes were: 9, 10, 11, 14, 15, 16, 20, 22, 24, 27, 28, 29, 33, 36, 37, 38, 48, 88, 127, and 142 individuals ($M=41$; $SD=34$; $Mdn=29$). Of these participants, 868 individuals provided all three demographic items of age, education, and gender (age: $M=75.9$ y old; $SD=11.4$; range 40.0–97.6; education: $M=16.1$ y; $SD=2.52$; range 6–21; gender: 68.7% female). We therefore included data from these 868 participants in our analyses, 86.6% of whom reported being “white”. Limited information was available regarding the 150 participants who were excluded from the analysis due to incomplete data, however, mean age, education, and gender were statistically similar to the group of 868

who were retained for the full analysis (independent samples *t*-tests, all p 's > 0.05); (age: $M=78.0$ y old; $SD=9.9$; range 55.3–97.4, $n=72$; education: $M=15.6$ y; $SD=2.72$; range 6–20, $n=33$; gender: 61.0% female, $n=33$). [The d' scores available from the 150 participants with incomplete data ($M=2.50$, $SD=1.03$, $Mdn=2.56$, skewness = -1.03) were compared to those of the 868 individuals who were retained ($M=2.82$, $SD=0.81$, $Mdn=2.98$, skewness = -1.69)]. An independent samples *t*-test showed that the d' scores of the group of participants who were excluded from the analysis were significantly lower than the group who were retained ($t(938)=3.18$, $p<0.002$) suggesting that this group may have an increased frequency of age-related memory impairment, possibly as a result of dementia, which could have been related to their failure to complete the test materials. As part of the statistical analyses (see Data Analysis), participants were divided into six sub-groups according to age (Table 1). Education level declined by 1.3 y from 16.9 to 15.6 from the youngest to the oldest age group, though the variation did not reach statistical significance ($F(5, 867)=1.93$, $p>0.05$). All age groups contained more females than males (see Table 1) and the groups varied significantly in the proportion of males to females ($\chi^2(n=868)=12.9$, $p=0.02$).

Materials

The audience-based CRT was developed using numerous complex visual stimuli. The images were of discrete objects taken from five categories (4 tractors, 5 kitchen utensils, 4 hats, 5 barnyard items, and 7 household decorative items; see Supplemental Material for the actual pictures used and the order of presentation, available online: <http://www.j-alz.com/issues/27/vol27-4.html#supplementarydata02>). Items were not systematically chosen but were selected to have a range of difficulties. Visually similar objects and difficult to name objects were used

to avoid strict reliance on verbal cues, to provide a challenge and to maintain the interest of the subjects.

Twenty-five color images of manmade items taken with a digital camera were selected from a range of pictures. From these 25 items, a 50-item test was constructed in the following way. The 25 items were first arranged in a random sequence. Eleven items were presented only once. Fourteen of the items were repeated once (average inter-repetition-interval between the first and second presentations was 7.93 items; range = 2 to 25 intervening items). Eleven of these items were shown again for a third time. Multiple repetitions were integrated into the test design to make recognition easier for subjects with impaired memories by providing more repetitions, and allowing for a comparison of recognition memory performance between the second and third presentations (average inter-repetition-interval between the second and third presentations was 21.1 items; range = 10 to 36 items).

The 50 items were transferred to a PowerPoint presentation and numbered in sequence (1–50) with a large numeral in the top left hand corner. Another series of ten images was constructed and was used as a practice test before the full test was given (5 images, 3 repeated once, 2 repeated a second time). The need for such a practice test had become obvious during pilot work, which indicated that about 10% of audience members could not follow the verbal instructions on the first try.

Each participant was provided with two sheets of paper. One sheet of paper contained only an ID number and was used by the participant to retrieve their test scores after the test administration without the need to provide personally identifying information. On the other sheet, demographic information was collected on one side of the page (age, education, and race) and the other side was used as an answer sheet for the recognition memory testing. The answer sheet had columns of numbers corresponding to the 10 slides of the practice-test and the 50 slides of the full test. A single circle was adjacent to each number on which the participant could indicate their response by filling in the circle. The sheet was organized so that it could be scanned for data entry. To enable efficient scoring of large numbers of answer sheets, a scoring template was made so it could be placed over the answer sheet by the scorer and the correct responses could be quickly identified.

Procedure

Testing at all sites adhered to a standard format, which began with a 20-minute introductory talk about

AD and the signs of dementia. As part of the talk, all participants were offered the memory test. Audiences were told that participating in the memory test was optional, and that individual test scores would be provided anonymously at the end of the presentation. A statement outlining the subjects' rights was provided to all audience members on a written page and reviewed on a slide. The Protocol was approved by Stanford University Institutional Review Board. No identifying information was collected, and therefore written consent was not required. The same 10-item practice test and 50-item memory test were used at all sites. Two individuals publicly acknowledged taking the test before, but were not identified.

The CRT was presented by projecting test items onto a screen using a laptop computer and projector. No effort was made to assess visual acuity of audience members or to assure adequate visibility from all parts of the room. However, the presenter ensured that the slides were generally easily seen from all vantage points of every room in which the test was administered. Participants were told that they would see a series of 50 pictures one at a time for 5 s per image with no inter-image interval. They were instructed to look at each picture carefully and any time they thought an image was repeated they should note the image number shown in the top left hand corner and immediately mark the circle corresponding to that number on their answer sheet. No response was required if they thought an image was not repeated (i.e., novel). The 10-item practice test was given first. The presenter then addressed any questions relating to the test procedure, and then the full 50-slide test was given (250 s). After the test, the participants handed their papers to the rater to be scored using the scoring template. A rater scored each participant's answer sheet and calculated the percentage of items correctly identified, and the percentage of items incorrectly identified. These scores and the corresponding participant ID were noted on a pre-prepared score sheet which was handed anonymously to the participant. The score sheets were given to the participants. The score sheets were given to the correct participant by matching the ID number on the score sheet with the ID number on the sheet of paper retained by the participant. The score sheet contained instructions on how to interpret the test scores. Specifically, if scores indicated a high probability of memory problems, a notation was made on the sheet encouraging the subject to visit their clinician for further evaluation. It has been reported by Boustani et al. [38] that about 50% of individuals receiving positive screens accept such a referral.

Data analysis

Results from the CRT were analyzed using the correct and incorrect response information. All statistical analyses were performed using IBM SPSS Statistics software (version 18). The correct recognition rate (hit rate) and the false positive rate were used to determine the signal detection parameter, discriminability score (d') [39]¹. The d' scores were calculated for each participant individually as: $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$ (note: a higher d' indicates that the signal can be more readily detected). The correct recognition scores were calculated using only the first repetition of the items ($n = 14$) since subsequent repetitions also represented learning trials and were not independent (the same pattern of results was obtained when data from all trials were analyzed). The false positive scores applied to all 25 items. A standard correction was necessary when calculating d' values if the hit rate or the false positive rate were 100% or 0%. Following MacMillan and Creelman [40], we converted 0% to $1/(2N)\%$ and 100% to $1-1/(2N)\%$ where N = the number of items.

CRT scores were analyzed in two ways. First, a linear regression analysis was performed on individual d' scores using Age, Gender, and Education as predictor variables. Age and Education are continuous variables and were centered by subtracting the mean of the variable (Age, $M = 75.9$ y; Education, $M = 16.1$ y) from each participants' score on that variable. Gender was coded as -0.5 and $+0.5$. Two-way interaction terms were constructed by multiplying the centered values of the three pairs of predictor variables (Age, Education, and Gender) and entering them into the regression model. The regression analysis of d' scores revealed statistically significant effects of Age and Education (see below). In order to examine the effects of Age and Education on d' more closely, participants were divided into groups based on Age and Education. Six age-groups were defined (see Table 1) and five educational groups were defined corresponding to major divisions of attainment in the U.S. educational system [i.e., ≤ 12 y (high school), 13–15 y (some college), 16 y (college completion), 17–19 y (masters degree), and 20–21 y (advanced degree)]. The effect of Age and Education on d' scores were investigated using a two-way analysis of variance (ANOVA). Significant effects were examined using the Dunnett's T3 posthoc test (IBM SPSS Statistics software). This test is designed

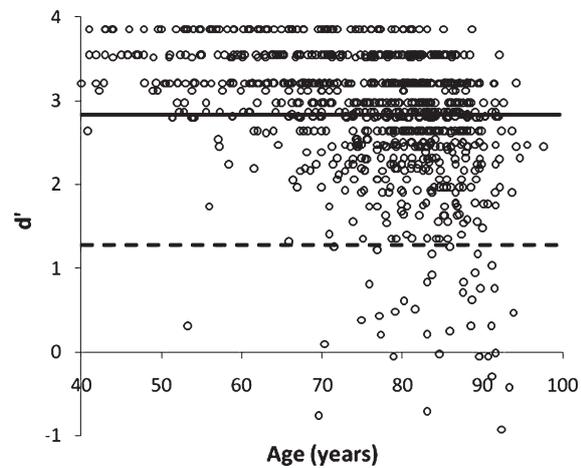


Fig. 1. The relationship between discriminability (d') and age on the audience-based continuous recognition test of memory. Each data point represents the recognition score of an individual participant expressed as a discriminability index d' . One individual whose score was unusually poor ($d' = -2.64$) was removed from the plot. The solid line represents the mean d' score and the dashed line represents a score 2 SD below the mean.

for comparing groups of unequal variance, it is conservative, and it automatically corrects for multiple comparisons. An alpha level of 0.05 was used for all statistical tests.

RESULTS

Multiple linear regression analysis revealed a significant effect of Age ($\beta = -0.026$, $SE = 0.029$, $p < 0.001$) such that greater age was associated with lower d' scores (Fig. 1). A significant effect of Education was also found ($\beta = 0.045$, $SE = 0.011$, $p < 0.001$), such that greater education was associated with higher d' scores. No significant effect was found for Gender ($\beta = -0.008$, $SE = 0.058$, $p = 0.89$) and no significant interactions were found (Age \times Gender, $\beta = -0.001$, $SE = 0.005$, $p = 0.92$; Age \times Education, $\beta = 0.002$, $SE = 0.001$, $p = 0.07$; Education \times Gender, $\beta = 0.042$, $SE = 0.023$, $p = 0.06$).

The effects of Age and Education were further investigated using a 2-way ANOVA which again showed significant main effects of Age ($F(5, 811) = 12.97$, $p < 0.001$) and Education ($F(4, 811) = 5.46$, $p < 0.001$) on d' scores. The interaction between Age and Education did not reach statistical significance, $F(19, 867) = 1.56$, $p = 0.06$. However, this interaction was explored further (see below).

ANOVA results indicated that the CRT was more difficult for older adults than younger adults. Test dis-

¹ Preliminary analyses indicated that the correlations between Age and test performance was best accounted for using d' ($r^2 = -0.371$, $p < 0.001$) rather than hit rate ($r^2 = -0.240$, $p < 0.001$) or false alarm rate ($r^2 = -0.254$, $p < 0.001$).

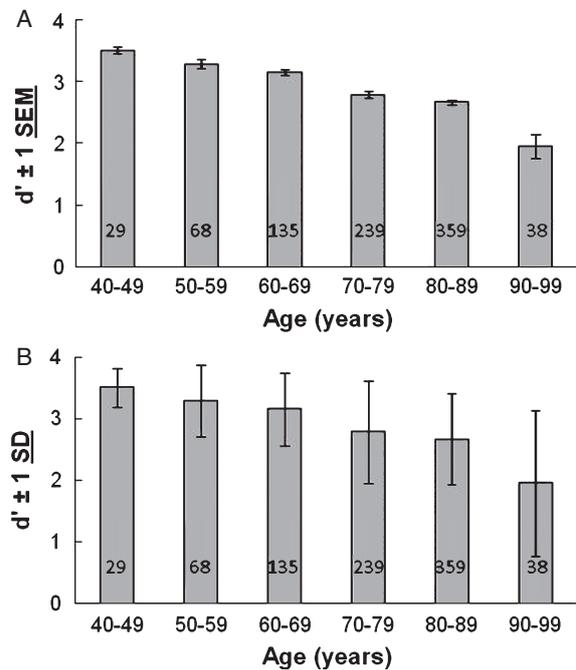


Fig. 2. A) The relationship between discriminability performance (d') and age in 868 individuals on the continuous recognition test of memory. Numbers inside the bars indicate the group n . The bars show the mean discriminability score for each age group and brackets show SEM. B) The same data is shown with brackets showing one standard deviation on either side of the mean.

criminability gradually but significantly declined with increasing age (Fig. 2A). (Similar results were found when the effect of Age on d' scores was analyzed using a nonparametric Kruskal-Wallis test, which showed a significant decline in d' with increasing age, $p < 0.001$). Of note, the standard deviation of d' generally increased with increasing age (Fig. 2B), indicating more variability in the population with increasing age, a common observation. Also, the post-hoc subsets consistently showed declines of performance with age. Note that the post-hoc Dunnett's T3 tests generally revealed similar performance across adjacent groups of participants aged 40–59 y, 50–69 y, and 70–89 y, but the group aged 90–99 y was significantly worse than all other groups.

Although the six age groups did not differ significantly in number of years of education (see Participant section), it was noted that the oldest group also had numerically the lowest average level of education. In order to verify that the effect of age on CRT performance was not confounded by educational level, a secondary analysis was performed in which participants with low levels of education (i.e., < 12 y of

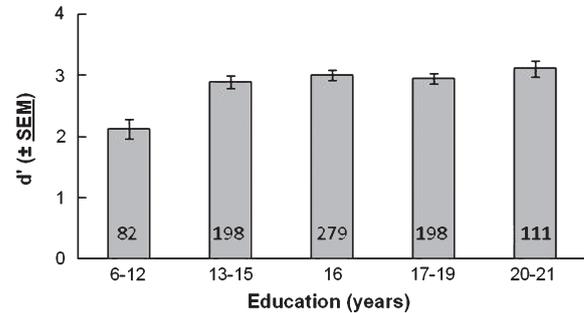


Fig. 3. The relationship between discriminability index (d') and education in 868 individuals on the continuous recognition test of memory. Numbers inside the bars indicate the group n . The bars show the mean discriminability score for each group and brackets show SEM.

education, $n = 82$) were excluded. Results were the same as when all participants were included, $F(5, 785) = 26.20$, $p < 0.001$. These results strongly indicated that test discriminability declined significantly with increasing age, and were not confounded by education.

The effect of education on CRT performance is shown in Fig. 3. Test performance was lower for those with education levels of 12 y or less relative to those with more education. However, performance reached a plateau after 12 y of education above which no significant improvement in performance was seen. Post-hoc tests showed that the test scores of the group with ≤ 12 y of education were significantly below all other groups (i.e., 13–21 y of education), and that the scores of the groups with more than 12 y of education did not differ significantly from one another. A one-way ANOVA confirmed that the mean age did not vary significantly across the five education groups, $F(4, 867) = 2.15$, $p > 0.05$. However, the lowest educational group (≤ 12 y) was also numerically the oldest (79.5 y vs. group mean of 76.4 y for those with over 12 y of education), suggesting that levels of education varied systematically with age, and the poorer performance of the lower education group may have actually been due to an age effect. In order to demonstrate that the effects of education on test score were associated with low levels of education rather than age, the analysis was repeated using only individuals having more than 12 y of education (i.e., excluding those with ≤ 12 y of education, $n = 82$). When individuals with low education were excluded, ANOVA showed no significant effects of education on test performance $F(3, 785) = 1.65$, $p > 0.05$. A subsequent analysis was also conducted in which the oldest participants (i.e., ≥ 90 y old, $n = 38$) were removed before the performance

across the five educational groups were compared. ANOVA showed the performance varied significantly across the five education groups, $F(4, 829) = 10.87$, $p < 0.001$, and post-hoc tests showed that the test scores of the group with ≤ 12 y of education were significantly below all other groups. Finally, an analysis was conducted in which participants ≥ 80 y old, ($n = 377$) were removed before the performance across the five educational groups were compared. ANOVA again showed the performance varied significantly across the five education groups, $F(4, 466) = 6.52$, $p < 0.001$. Post-hoc tests showed that the group with ≤ 12 y of education had a significantly lower score that the group with 16 y of education (which had numerically the highest mean score of all groups). Taken together, these results suggest that test discriminability was significantly worse for participants with low education, and was not confounded by age.

Due to the repeat-detection format of the CRT, participants were required to hold items in memory across a variable delay. The inter-repetition-interval ranged from 2 to 25 images. This delay could disrupt recognition performance in two ways. First, as the number of intervening items increased, the time delay between the first and subsequent presentations of the same item could reduce recognition. Second, as other test items were presented during the delay, interference could build up. To explore these effects, a linear regression analysis was performed between the number of intervening items and percent correct. No significant relationship was found between the number of intervening items and CRT performance, $\beta = -0.023$, $SE = 0.015$, $p > 0.05$, (Fig. 4). The inter-repetition-interval had little overall effect on recognition and average performance was maintained at a high level across repeated items (average % correct = 89%).

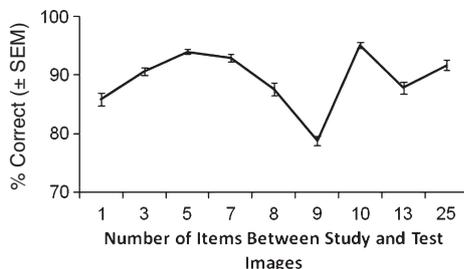


Fig. 4. The relationship between the number of intervening items (between initial and first repeat presentations) and percent correct on those items across 868 individuals on the continuous recognition test of memory. The brackets show SEM. Note that each item was shown for 5 seconds, so the correspondence with the temporal interval can be calculated.

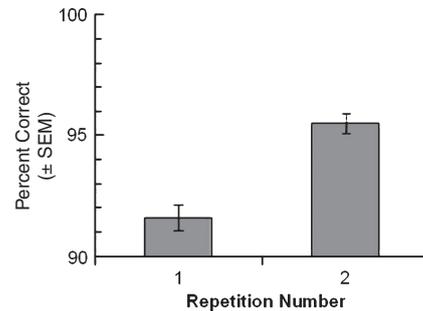


Fig. 5. The relationship between percent correct and item repetition in 868 individuals on the continuous recognition test of memory. Eleven items were shown three times during the test. Recognition performance increased between the second and third presentation (average number of intervening items = 21.1, range; 10–36 items). A paired t -test demonstrated that the difference was significant ($p < 0.005$).

Another issue related to a repeat-detection format is that when test items are repeated multiple times, each subsequent presentation serves as a retrieval cue to reactivate and strengthen the memory representation of the information stored during earlier presentations [41]. In the current test, eleven items were shown three times, and CRT performance did increase across repeated presentations as shown in Fig. 5. A paired t -test compared the mean percent correct between the first and second repetitions and showed that this difference (91.6% vs. 95.5% correct) was statistically significant, $t(867) = -10.30$, $p < 0.005$.

DISCUSSION

The results from this study of community-dwelling participants showed that memory can be measured using a continuous recognition format in a large group setting. The decreased performance with age in this study is consistent with the well-known pattern of age-related memory decline [35, 42–44]. The results encourage the development of this test format as a memory screen capable of testing large numbers of individuals.

The CRT may be a useful screening tool for the detection of memory problems related to a variety of disorders. The type of memory measured by the CRT is referred to as declarative memory, which is information that can be consciously recalled, such as facts and events [45]. Among the many causes of declarative memory failure, dementia due to AD is by far the most common [22]. If the CRT is to be useful in community screening programs, it should be sensitive to the memory failures that occur as a result of AD dementia.

Although no data were collected in the current study regarding the ability of the CRT to detect dementia, the most common symptom of early AD is usually a severe difficulty in encoding new declarative memory [13, 46, 47]. A recognition format such as that used by the CRT may be especially suitable for the detection of encoding impairments such as occur in AD because the target stimuli are given as part of the test materials which makes the test less dependent on retrieval processes than are other commonly used testing formats such as free- and cued-recall [28]. An impairment in recognition memory is therefore strong evidence for an impairment in encoding [48].

It should be acknowledged however that the present study did not assess the relative contributions of encoding and retrieval to memory performance. It should also be acknowledged that the test was given under relatively uncontrolled conditions, and test performance may have been affected by a variety of factors including all visual impairment at any level and heightened levels of stress due to taking a test. In addition, there is a possibility of hearing impairment interfering with performance, e.g., poor understanding of test instructions, especially in the older subjects and especially in men, in whom hearing impairment can be more common than in women [49]. Other possible test confounds include the integrity of attentional processes, as well as numerous psychiatric and neurological conditions such as aphasia, anomia, stroke, trauma, alcoholism, and depression. Future studies could include rating scales for how well the participant could perceive the images and understand test instructions. For recommendation of this test for clinical purposes, studies will first need to be done on well-characterized populations to test the reliability and validity of the CRT to detect memory impairments. In order to ascertain how many of those who do poorly at the task will go on to develop dementia, a longitudinal study design would be particularly informative.

Average performance on the CRT decreased markedly with increasing age. This pattern of results is consistent with the well-established effects of age on declarative memory which is known to decrease in accuracy and increase in reaction time during the retrieval of previously encoded information including word lists, faces, and complex scenes [35, 50–52] (For review see [53]). The CRT also showed a substantial increase in standard deviation with increasing age (SD, Fig. 2B). This pattern of variability with age stands in contrast to other measures of memory which show approximately equal variability across all ages at least in the healthy samples used to establish norms

for standardized tests [53, 54]. The increase in variability with age in our results is consistent with the increased frequency of dementia with increasing age [55]. Further, there is an increased frequency of factors that may introduce noise into the measurement of memory with increasing age, such as greater difficulty in hearing, seeing or attending to relevant stimuli [56]. These factors were not measured or controlled for in the current study. Alternatively, the decreased variability in younger individuals could be related to a ceiling effect, which could be explored further in future studies by using more difficult sets of stimuli. However, the general pattern of memory deterioration with age in our sample is consistent with the Gompertz Law, which states that the rate of system failures increases exponentially with age [57]. In this case, the failures of mechanisms subserving CRT performance were increasingly impaired with age. It should be noted that the Gompertz Law becomes increasingly inaccurate over 95 y of age, and this interaction warrants further research. For clinical application, further studies will be needed to establish the sources of variation in memory function across age in carefully characterized populations of healthy adults so that norms can be defined for individuals of specific ages.

It is noteworthy that recognition improved for the images that were repeated (Fig. 5). This finding is consistent with the notion that each repetition of an image acted as a new learning trial [41] during which the images were re-encoded. This observation suggests that participants were actively engaged in the task and were attending to the images. This observation can be related to an fMRI finding of decreased neural response to repetition of identical stimuli (repetition suppression) associated with an increase in effectiveness of connectivity between distinct cortical systems corresponding to associative learning [58], which would presumably be related to processing selectively disrupted by AD. This feature of the test, increase of learning across repeated images, may reduce floor effects thereby allowing a broader assessment of memory function in highly impaired individuals. However, the observation does not address the issue of what strategies participants may have used to perform the task. Recognition memory is widely viewed as consisting of two components; an episodic component which supports the ability to remember the episode in which an item was encountered and a familiarity component that supports the ability to know that an item was presented, but without providing contextual details of the episode itself [59, 60]. There is evidence that some older adults show a decrease in recollection together

with an increase in familiarity [61, 62]. It would be interesting to address issues of episodic memory function and familiarity in future studies.

When considering a test for clinical screening purposes (e.g., for dementia related to AD, or other pathology affecting memory), one issue to consider is what constitutes poor performance. One approach is to define poor performance as a score 2 SDs below the mean. This approach has drawbacks. For example, a younger individual (e.g., between 40–50 y) with a performance level 2 SDs below the mean for this age group may be considered to be of clinical concern. The problem is that the same low score obtained by older individuals may lie within 2 SDs of the mean of the older age group, and thus would not be “abnormal”. To avoid this problem, norms are usually generated for specific age groups and have no absolute cut-off score across all ages.

When developing a screen for memory problems, it is also necessary to consider cost-effectiveness, including consideration of the disease targeted for screening. The decision about whether to screen an individual and the critical level for clinical concern depend on an analysis of many factors. The factors to consider for such an analysis include: incidence of disease; the benefit of a true-positive screen; the cost of a false-positive screen; the incidence of the target problems in the population; and the cost of the test. Although no formal cost analysis was performed in the current study, the cost of the test was estimated at approximately \$10 per individual (\$400 to manage and present at a venue, for a mean audience size of 40 participants the cost is \$10 per individual, range 9 to 142 provides a cost range of \$3 to \$45 per individual), which appears to compare favorably with the potential cost of administering the test in a face-to-face interaction with a professional clinician (unlikely to be less than \$45 per individual). An important factor in screening memory in older adults is the likelihood of disease and specifically the exponential increase of dementia incidence with age [55]. The value of using a particular level of test performance as a positive screen for an individual is approximated by a cost-worthiness analysis [14]. This approach is more difficult than a simple cut-off value for screening as described above, but is more cost effective for addressing clinical needs.

A test such as the CRT could be useful in longitudinal studies in which memory performance is monitored over time to detect progressive memory disorders. An important observation in this regard is that the CRT is well tolerated so that individuals may be willing to take the test on repeated occasions. Although no

formal data were collected on this aspect of the test, audiences were asked by the experimenter at the end of each testing session whether they enjoyed the test and would be prepared to take the test again in the future. This question was usually answered by a unanimous show of hands, suggesting that the test format is acceptable to audiences and suitable for repeated testing. Of course, changes over time should be assessed with respect to age-cohorts since normal performance levels and changes over time do vary according to age [44]. The observation that the test was well-tolerated suggests that test anxiety was relatively low, at least after the test began. This is somewhat surprising for a test in which poor performance may indicate serious medical problems. However, casual observation suggested that test anxiety was reduced by administration of the practice test. Future versions of the test may include the use of audience voting technology which would allow responses to be monitored by the test administrators. This would allow for the identification of individuals who performed poorly on the test who could be approached individually and offered specific advice. Stress may also be reduced in participants obtaining a low test score by showing a slide at the end of the test session explaining the many ways, other than dementia, in which a low test score could be achieved. Future versions of the test could assess the stress of taking the test by including stress rating scales before and after taking the test.

The CRT was not affected by education level beyond high-school, probably because the memory processes targeted by this test require relatively simple object-recognition, similar to what laboratory animals can be trained to do [31, 63, 64]. Accordingly, education appears to have a minimal effect on performance. However, participants with a high-school education or less performed significantly less well. This pattern of results is consistent with standardization studies which report an association between education and recognition memory test performance [e.g., 37]. However, there needs to be further study of individuals with low education on this test. In conclusion, while screening tests are widely used throughout Medicine, they are not yet commonly used to detect memory problems that accompany serious illness such as dementia [38]. The lack of memory screening is based in part on the lack of easily administered and validated screening tools [6, 13]. The methodology of the CRT presented here could be developed to serve this important need. The CRT has many features that make it potentially suitable as a memory screen; it can be administered to a large number of individuals simultaneously in an audi-

ence setting, it is brief and easy to administer, and it is well tolerated by elderly audiences. These preliminary results suggest that the test can detect memory problems, and the pattern of results is consistent with what is known about the effects of aging and education on memory. However, further studies are needed to establish the reliability, validity, and utility of this test in well-defined clinical populations. These results suggest that development of this type of testing is warranted.

ACKNOWLEDGMENTS

Drs. Samina Gul and Raj Kumar helped with data entry. Ms. Deryl Wicks assisted with coordinating most of the presentations. Pictures used were courtesy of Mr. Henry Bowles of Bowles-Langley Technology. We thank Dr. Booil Jo for statistical assistance. Several members of the Stanford/VA Aging Clinical Research Center, directed by Dr. Jerome Yesavage, assisted with the presentations and scoring of the tests.

This work was supported by the War Related Illness and Injury Study Center (WRIISC) at the Veterans Affairs Palo Alto Health Care System, the Medical Research Service of the Veterans Affairs Palo Alto Health Care System, and the Department of Veterans Affairs Sierra-Pacific Mental Illness Research, Education, and Clinical Center (MIRECC).

Authors' disclosures available online (<http://www.j-alz.com/disclosures/view.php?id=967>).

REFERENCES

- [1] Budson AE, Price BH (2005) Memory dysfunction. *N Engl J Med* **352**, 692-699.
- [2] Hy LX, Keller DM (2000) Prevalence of AD among whites: A summary by levels of severity. *Neurology* **55**, 198-204.
- [3] Mesulam MM (2000) *Principles of behavioral and cognitive neurology*, Oxford University Press, New York.
- [4] Newman MF, Kirchner JL, Phillips-Bute B, Gaver V, Grocott H, Jones RH, Mark DB, Reves JG, Blumenthal JA (2001) Longitudinal assessment of neurocognitive function after coronary-artery bypass surgery. *N Engl J Med* **344**, 395-402.
- [5] Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, Burke JR, Hurd MD, Potter GG, Rodgers WL, Steffens DC, McArdle JJ, Willis RJ, Wallace RB (2008) Prevalence of cognitive impairment without dementia in the United States. *Ann Intern Med* **148**, 427-434.
- [6] Ashford J, Borson S, O'Hara R, Dash P, Frank L, Robert P, Shankle W, Tierney M, Brodaty H, Schmitt F, Kraemer H, Buschke H (2007) Should older adults be screened for dementia? It is important to screen for evidence of dementia. *Alzheimers Dement* **3**, 75-80.
- [7] Callahan CM, Hendrie HC, Tierney WM (1995) Documentation and evaluation of cognitive impairment in elderly primary care patients. *Ann Intern Med* **122**, 422-429.
- [8] Finkel SI (2003) Cognitive screening in the primary care setting. The role of physicians at the first point of entry. *Geriatrics* **58**, 43-44.
- [9] Ross GW, Abbott RD, Petrovitch H, Masaki KH, Murdaugh C, Trockman C, Curb JD, White LR (1997) Frequency and characteristics of silent dementia among elderly Japanese-American men. The Honolulu-Asia Aging Study. *JAMA* **277**, 800-805.
- [10] Sternberg SA, Wolfson C, Baumgarten M (2000) Undetected dementia in community-dwelling older people: The Canadian Study of Health and Aging. *J Am Geriatr Soc* **48**, 1430-1434.
- [11] Valcour VG, Masaki KH, Curb JD, Blanchette PL (2000) The detection of dementia in the primary care setting. *Arch Intern Med* **160**, 2964-2968.
- [12] Lawrence, Davidoff DA, Katt-Lloyd D, Connell A, Berlow YA, Savoie JA (2003) Is large-scale community memory screening feasible? Experience from a regional memory-screening day. *J Am Geriatr Soc* **51**, 1072-1078.
- [13] Ashford JW, Borson S (2008) Primary care screening for dementia and mild cognitive impairment. *JAMA* **299**, 1132-1133; author reply 1133-1134.
- [14] Ashford JW (2008) Screening for memory disorder, dementia, and Alzheimer's disease. *Aging Health* **4**, 399-432.
- [15] Lawrence, Davidoff D, Katt-Lloyd D, Auerbach M, Hennen J (2001) A pilot program of improved methods for community-based screening for dementia. *Am J Geriatr Psychiatry* **9**, 205-211.
- [16] Folstein MF, Folstein SE, McHugh PR (1975) "Mini-mental state" A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.
- [17] Wild K, Howieson D, Webbe F, Seelye A, Kaye J (2008) Status of computerized cognitive testing in aging: A systematic review. *Alzheimers Dement* **4**, 428-437.
- [18] Fillit HM, Simon ES, Doniger GM, Cummings JL (2008) Practicality of a computerized system for cognitive assessment in the elderly. *Alzheimers Dement* **4**, 14-21.
- [19] Dwolatzky T, Whitehead V, Doniger GM, Simon ES, Schweiger A, Jaffe D, Chertkow H (2004) Validity of the Mindstreams computerized cognitive battery for mild cognitive impairment. *J Mol Neurosci* **24**, 33-44.
- [20] Doniger GM, Zucker DM, Schweiger A, Dwolatzky T, Chertkow H, Crystal H, Simon ES (2005) Towards practical cognitive assessment for detection of early dementia: A 30-minute computerized battery discriminates as well as longer testing. *Curr Alzheimer Res* **2**, 117-124.
- [21] Korczyn AD, Aharonson V (2007) Computerized methods in the assessment and prediction of dementia. *Curr Alzheimer Res* **4**, 364-369.
- [22] Alzheimer's A (2010) Alzheimer's disease facts and figures. *Alzheimers Dement* **6**, 158-194.
- [23] Ashford JW, Jarvik L (1985) Alzheimer's disease: Does neuron plasticity predispose to axonal neurofibrillary degeneration? *N Engl J Med* **313**, 388-389.
- [24] Ashford JW, Kolm P, Colliver JA, Bekian C, Hsu LN (1989) Alzheimer patient evaluation and the mini-mental state: Item characteristic curve analysis. *J Gerontol* **44**, P139-P146.
- [25] Ashford JW, Schmitt FA (2001) Modeling the time-course of Alzheimer dementia. *Curr Psychiatry Rep* **3**, 20-28.
- [26] Elias MF, Beiser A, Wolf PA, Au R, White RF, D'Agostino RB (2000) The preclinical phase of Alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch Neurol* **57**, 808-813.

- [27] Teter B, Ashford JW (2002) Neuroplasticity in Alzheimer's disease. *J Neurosci Res* **70**, 402-437.
- [28] Lowndes G, Savage G (2007) Early detection of memory impairment in Alzheimer's disease: A neurocognitive perspective on assessment. *Neuropsychol Rev* **17**, 193-202.
- [29] Shepard RN (1967) Recognition memory for words, sentences, and pictures. *J Verb Learn Verb Behav* **6**, 156-163.
- [30] Standing L (1973) Learning 10,000 pictures. *Q J Exp Psychol* **25**, 207-222.
- [31] Wright AA, Santiago H, Sands SF, Kendrick DF, Cook RG (1985) Memory processing of serial lists by pigeons, monkeys, and people. *Science* **229**, 287-289.
- [32] Heaton RK, Taylor MJ, Manly J (2003) In *Clinical Interpretation of the WAIS-III and WMS-III*, ed. Tulskey DSea (Academic Press, San Diego), pp. 181-210.
- [33] Lange RT, Chelune GJ, Taylor MJ, Woodward TS, Heaton RK (2006) Development of demographic norms for four new WAIS-III/WMS-III indexes. *Psychol Assess* **18**, 174-181.
- [34] Grady CL, McIntosh AR, Horwitz B, Maisog JM, Ungerleider LG, Mentis MJ, Pietrini P, Schapiro MB, Haxby JV (1995) Age-related reductions in human recognition memory due to impaired encoding. *Science* **269**, 218-221.
- [35] Schacter DL, Cooper LA, Valdiserri M (1992) Implicit and explicit memory for novel visual objects in older and younger adults. *Psychol Aging* **7**, 299-308.
- [36] Diesfeldt HF (1990) Recognition memory for words and faces in primary degenerative dementia of the Alzheimer type and normal old age. *J Clin Exp Neuropsychol* **12**, 931-945.
- [37] Warrington EK (1984) (Western Psychological Services, Los Angeles).
- [38] Boustani M, Callahan CM, Unverzagt FW, Austrom MG, Perkins AJ, Fultz BA, Hui SL, Hendrie HC (2005) Implementing a screening and diagnosis program for dementia in primary care. *J Gen Intern Med* **20**, 572-577.
- [39] Green DM, Swets JA (1966) *Signal detection theory and psychophysics*, Wiley, New York.
- [40] MacMillan NA, Creelman CD (1991) *Detection theory: A users guide*, Cambridge University Press, Cambridge.
- [41] Thios S, D'Agostino RB (1976) Effects of repetition as a function of study-phase retrieval. *J Verb Learn Verb Behav* **15**, 529-536.
- [42] Crook TH, 3rd, Larrabee GJ, Youngjohn JR (1993) Age and incidental recall for a simulated everyday memory task. *J Gerontol* **48**, P45-47.
- [43] Salthouse TA (2009) When does age-related cognitive decline begin? *Neurobiol Aging* **30**, 507-514.
- [44] Schaie KW (2009) "When does age-related cognitive decline begin?" Salthouse again reifies the "cross-sectional fallacy". *Neurobiol Aging* **30**, 528-529; discussion 530-533.
- [45] Squire LR, Stark CEL, Clark RE (2004) The medial temporal lobe. *Annu Rev Neurosci* **27**, 279-306.
- [46] Mickes L, Wixted JT, Fennema-Notestine C, Galasko D, Bondi MW, Thal LJ, Salmon DP (2007) Progressive impairment on neuropsychological tasks in a longitudinal study of preclinical Alzheimer's disease. *Neuropsychology* **21**, 696-705.
- [47] Salmon DP, Bondi MW (1999) Neuropsychology of Alzheimer's disease In *Alzheimer Disease*, Terry RD, Katzman R, Bick KL, Sisodia SS, eds. Lippincott Williams & Wilkens, Philadelphia, pp. 39-56.
- [48] Wisdom NM, Callahan JL, Hawkins KA (2011) The effects of apolipoprotein E on non-impaired cognitive functioning: A meta-analysis. *Neurobiol Aging* **32**, 63-74.
- [49] Dillon CF, Gu Q, Hoffman HJ, Ko CW (2010) Vision, hearing, balance, and sensory impairment in Americans aged 70 years and over: United States, 1999-2006. *NCHS Data Brief* 1-8.
- [50] Bartlett JC, Leslie JE, Tubbs A, Fulton A (1989) Aging and memory for pictures of faces. *Psychol Aging* **4**, 276-283.
- [51] Park DC, Smith AD, Lautenschlager G, Earles JL, Frieske D, Zwahr M, Gaines CL (1996) Mediators of long-term memory performance across the life span. *Psychol Aging* **11**, 621-637.
- [52] Parkin AJ, Walter BM (1992) Recollective experience, normal aging, and frontal dysfunction. *Psychol Aging* **7**, 290-298.
- [53] Salthouse TA (2003) Memory aging from 18 to 80. *Alzheimer Dis Assoc Disord* **17**, 162-167.
- [54] Wechsler, D (1997) *Wechsler Adult Intelligence Scale, 3rd ed. (WAIS III): Administration and Scoring Manual*, Psychological Corporation, San Antonio, TX.
- [55] Jorm AF, Jolley D (1998) The incidence of dementia: A meta-analysis. *Neurology* **51**, 728-733.
- [56] Drag LL, Bieliauskas LA (2010) Contemporary Review 2009: Cognitive aging. *J Geriatr Psychiatry Neurol* **23**, 75-93.
- [57] Ashford JW, Atwood CS, Blass JP, Bowen RL, Finch CE, Iqbal K, Joseph JA, Perry G (2005) What is aging? What is its role in Alzheimer's disease? What can we do about it? *J Alzheimers Dis* **7**, 247-253; discussion 255-262.
- [58] Buchel C, Coull JT, Friston KJ (1999) The predictive value of changes in effective connectivity for human learning. *Science* **283**, 1538-1541.
- [59] Tulving E (1985) Memory and consciousness. *Canad Psychol* **26**, 1-12.
- [60] Mandler G (1980) Recognizing: The judgment of previous occurrence. *Psychol Rev* **87**, 252-271.
- [61] Bastin C, Van der Linden M (2003) The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology* **17**, 14-24.
- [62] Davidson PS, Glisky EL (2002) Neuropsychological correlates of recollection and familiarity in normal aging. *Cogn Affect Behav Neurosci* **2**, 174-186.
- [63] Ashford JW, Coburn KL, Fuster JM (1998) Functional Cognitive Networks in Primates In *Fundamentals of Neural Networks: Neuropsychology and Cognitive Neuroscience*, Parks RW, Levine DS, eds. The MIT Press, Cambridge, Mass.
- [64] Mishkin M, Appenzeller T (1987) The anatomy of memory. *Scientific American* **256**, 80-89.