Three-Word Recall in Normal Aging

Melanie J. Chandler, Laura H. Lacritz, Antionette R. Cicerello, Sandra B. Chapman, Lawrence S. Honig, Myron F. Weiner & C. Munro Cullum

a The University of Texas Southwestern Medical Center at Dallas, Dallas, TX
b Richardson, TX
c Center for Brain Health at the University of Texas at Dallas, Dallas, TX
d Columbia University Department of Neurology, New York, NY


To cite this article: Melanie J. Chandler, Laura H. Lacritz, Antionette R. Cicerello, Sandra B. Chapman, Lawrence S. Honig, Myron F. Weiner & C. Munro Cullum (2004): Three-Word Recall in Normal Aging, Journal of Clinical and Experimental Neuropsychology, 26:8, 1128-1133

To link to this article: http://dx.doi.org/10.1080/13803390490515540

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Three-Word Recall in Normal Aging

Melanie J. Chandler1, Laura H. Lacritz1, Antionette R. Cicerello2, Sandra B. Chapman3, Lawrence S. Honig4, Myron F. Weiner1, and C. Munro Cullum1

1The University of Texas Southwestern Medical Center at Dallas, Dallas, TX, 2Private Practice, Richardson, TX, 3Center for Brain Health at the University of Texas at Dallas, Dallas, TX, 4Columbia University Department of Neurology, New York, NY

ABSTRACT

Three-word recall tasks are widely used as brief measures of verbal memory function, although interpretation of performance is complicated by variations in test instructions and procedures. The purpose of this study was to examine 3-word recall performance in samples of healthy subjects aged 52–75 (M age = 70) and age 76–92 (M age = 82) compared to patients with Alzheimer’s Disease (AD) when explicit prompts to remember the words were given. Those in the younger aging group remembered significantly more words than those in the older sample after a brief delay (M = 2.8 and 2.3, respectively). However, the majority of control subjects recalled 2 or 3 words after the delay, with only 3% of the 50–75 year old group and 17% of the 76+ year old group recalling 0 or 1 word on delayed recall. This is in stark contrast to the 87% of individuals with AD who recalled 0 or 1 word. Even though 3-word recall performance decreases with age, good recall (2 or 3 words) can be expected in most cases of normal aging.

INTRODUCTION

Three-word recall tasks are popular, brief memory assessment tools utilized by a variety of professionals in emergency, neurology, psychiatry/psychology, nursing, and primary care settings. The use of “screening” measures such as three-word recall are useful in these settings because they are quick and easily administered. Despite its facility and popularity, however, questions remain about the accuracy and reliability of the three-word recall task.

Standardization and Normative Data

Interpretation of three-word recall performance is complicated by varying methods of item presentation, heterogeneity of words used, and the limited availability of normative data. The most commonly used version of three-word recall comes from the Folstein Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), a popular screening measure for cognitive impairment. Originally, there were no guidelines as to which words to use as part of the MMSE, but in 1979 the words “apple, table, and penny” were adopted when the measure became part of the Diagnostic Interview Schedule for the National Institute of Mental Health (Tombaugh & McIntyre, 1992). In the original 1975 instructions for the MMSE, three-word recall was an incidental memory task, as examinees were not instructed that they needed to remember the words for later recall. Current instructions [e.g., the Consortium to Establish a Registry for Alzheimer’s Disease]
THREE-WORD RECALL

(CERAD), Morris et al., 1989] include prompting from the examiner that the individual will be asked to remember the words after a delay. Three-word recall has been shown to account for more errors than any other item on the MMSE, and three-word recall is one of the most sensitive items on the MMSE to aging (Tombaugh & McIntyre). Crum, Anthony, Bassett, and Folstein (1993) established normative data for MMSE total scores based on age and education; however, no such norms exist for the three-word recall task on its own.

Given the lack of normative data for three-word recall performance, clinicians are left to rely primarily on their experience to make judgments about the significance of the number of words remembered. It has been suggested (Feher et al., 1992) that using a cut-off score of less than three words provides optimal sensitivity to memory impairment, and a cut-off of less than two words balances specificity and limits false positives. Several studies have suggested three-word recall produces an unacceptable number of false positives in differentiating patients with and without memory deficits. Kuslansky, Buschke, Katz, Sliwinski, and Lipton (2002) found that three-word recall was highly sensitive in predicting Alzheimer’s Disease (AD) in a sample of community dwelling individuals over age 65, although 14% of their non-demented elderly sample recalled zero out of three words when using the words “ball, flag, and tree.” Similar results were reported earlier by Cullum, Thompson, and Smernoff (1993), who found that when the words “rose, ball, and key” were administered without prompts for recall to a sample of healthy, elderly individuals with no history of memory impairment and normal performance on detailed memory testing, 14% recalled none and 19% remembered only one of the words. These authors also noted that the specific words administered drastically affected three-word recall performance. When the words “brown, tulip, and honesty” were administered to a subset of the subjects above, the percentage of the sample recalling none of the three words increased to 60%. Professionals commonly choose their own words when giving three-word recall in daily practice, sometimes including verbs, adverbs, or even the name of an individual or street address as a “word.” Leaving word choice to the discretion of the examiner changes the standardization of the test and reduces comparability of the results.

Another difficulty with standardization of the three-word recall task is variability in administration as to whether or not subjects are prompted to remember the words for later recall. In most studies involving three-word recall, it is unclear if such a prompt was given or not, although as noted, the original MMSE instructions did not include any prompting. Without a standard method of administration and standard test items, the comparability of findings in the existent literature on three-word recall is questionable.

Three-Word Recall in the Elderly

Research into three-word recall performance among elderly samples is sparse, but the existing literature suggests three-word recall is more sensitive to memory declines in the elderly than in younger adults (Guilmette, Tsok, & Malcolm, 1995), and recall decreases as age increases, even among younger and older elderly subgroups (Cullum et al., 1993). When Hassing, Wahlin, and Bäckman (1998) studied healthy individuals over 90 years of age (N = 80), they found mean recall of three words after a brief delay to be as low as 1.15 words (SD = .92). This is not surprising, as differences in performance between younger and older elderly individuals on more lengthy verbal memory tests is commonly accepted (e.g., Cregger & Rogers, 1998; Small, Dixon, Hultsch, & Hertzog, 1999).

When looking at elderly individuals with AD, Beardsall and Huppert (1991) noted that individuals with mild/moderate AD recalled significantly fewer words (M = 1.0; SD = 1.1) than those in their normal elderly sample (M = 2.0; SD = 1.0), but that three-word recall alone was inadequate at accurately discriminating individuals with AD from normal subjects.

Current Study

Reliable norms and standard administration procedures in research studies are limited for the three-word recall task, and results from the administration with the provision of prompts are lacking. The current study examined the performance of younger and older healthy elderly subjects on the three-word recall task using current standardized
procedures from the MMSE (Morris et al., 1989). Differences in three-word recall between normal aging controls and individuals with AD also were examined.

METHOD

Participants

Normal Aging Sample
One-hundred-and-forty-three healthy subjects over age 50 were recruited from a local retirement home and surrounding community for a study on cognition and normal aging. Subjects had to obtain a score of 24 or above on the MMSE and could not have a history of neurological impairment, psychiatric disorder, substance abuse, or learning disability, nor could they be taking any medication that may have affected their cognitive functioning. Subjects were divided into younger and older normal aging groups (age 52–75, n = 66; age 76–92, n = 77, respectively).

AD Sample
Seventy-six patients with autopsy confirmed AD who had previously been evaluated at the Clinic for Alzheimer’s and Related Diseases (ADC) at the University of Texas Southwestern Medical Center were included.

Measures
Three-word recall was administered according to standard procedures as part of the MMSE. The words “apple,” “table,” and “penny” were presented in that order, and subjects were prompted that they would be asked to later recall the words. Delayed recall was assessed as part of the MMSE, which entails a delay period of approximately two to three minutes following a distracter task (spelling of WORLD forward and backward).

RESULTS

Demographics
Demographic information for the participants can be found in Table 1.

Age was negatively correlated with MMSE (r = -.21, p = .01) and three-word recall performance (r = -.37, p < .001) among the control sample as a whole, with older age relating to poorer performance. Age was significantly lower in the AD sample than the controls as a whole, F(1, 217) = 16.80, p < .001, with the AD group being more similar in age to the younger aging group and significantly different from the older aging group (p < .001). Education did not correlate with either MMSE total score or three-word recall in the control group. No other significant differences with respect to demographic variables were noted.

Normal Aging and AD Comparisons
One-way ANOVA revealed significant differences for three-word recall performance between the younger aging (M = 2.79, SD = .48) and the older aging (M = 2.30, SD = .84) groups, F(1, 141) = 17.34, p < .001, with the younger group recalling more words. Total MMSE scores were also significantly different between the groups, F(1, 141) = 10.38, p = .002, despite similar means, with the younger group (M = 28.97, SD = 1.31) slightly

Table 1. Demographic and Test Results.

<table>
<thead>
<tr>
<th>Age 52 – 75</th>
<th>Age 76 – 92</th>
<th>Alzheimer’s Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 66</td>
<td>n = 77</td>
<td>n = 76</td>
</tr>
<tr>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Age</td>
<td>70.12a,e</td>
<td>81.52c</td>
</tr>
<tr>
<td>Education</td>
<td>14.33c</td>
<td>15.26e</td>
</tr>
<tr>
<td>MMSE</td>
<td>28.97ab,c</td>
<td>28.25bc</td>
</tr>
<tr>
<td>3 Word Recall</td>
<td>2.79abc</td>
<td>2.30bc</td>
</tr>
<tr>
<td>Percent Male</td>
<td>36.4</td>
<td>41.6</td>
</tr>
</tbody>
</table>

a Value significantly different from those in the 76–92 year old group, p < .001.
b Value significantly different from those in the 52–75 year old group p < .001.
c Value significantly different from the AD group, p < .05.
outperforming the older group \((M = 28.25, SD = 1.36)\). Figure 1 presents the proportion of each sample that recalled 0–3 words, and Table 2 presents normative data for three-word recall performance.

Overlapping age ranges were utilized in Table 2 across cells based on the rationale provided by Pauker (1988). This method provides larger sample sizes in each group for more reliable data and allows clinicians to choose the age group that corresponds best with the age of a patient [e.g., a patient aged 87 may be better represented in the age range in which he/she falls towards the midpoint (82–92) than at the extreme (77–87)].

The majority of normal subjects (82% of the younger and 51% of the older groups) were able to recall all three words after the brief delay. However, significantly more of the younger subjects recalled all three words compared to those in the older aging group, \(\chi^2 (1, N = 143) = 15.18, p < .001\). AD subjects recalled significantly fewer words \((M = .57, SD = .88)\) than those in either of the normal aging groups using ANCOVA with age as a covariate, \(F (2, 215) = 177.85, p < .001\). These recall findings remained significant when a subgroup of mild AD (MMSE >18, \(n = 36, M = 1.03\)) was compared to a random selection of 36 subjects from the original sample of normal controls while controlling for age and education, \(F (1, 68) = 53.27, p < .001\). The proportion of all AD subjects recalling less than 2 or greater than or equal to 2 words significantly differed from the normal aging groups, \(\chi^2 (1, N = 219) = 124.13, p < .001\). The percentage of AD subjects recalling 0–3 words is presented in Figure 1.

**DISCUSSION**

In keeping with the findings of Cullum et al. (1993), age was significantly correlated with the number of words recalled among the normal aging sample, with those in the 50–75 age group outperforming those in the 76–92 age group, although the difference between the groups was small owing to the limited range of possible scores on this measure. Similar age effects have been observed with other, more lengthy measures

![Fig. 1. Percentage of younger aging, older aging, and Alzheimer’s Disease (AD) subjects recalling 0, 1, 2, or 3 words. All groups significantly differed in the proportion of words recalled, \(p < .001\).](image)

**Table 2. T Scores for Three-Word Recall Using “Apple, Table, and Penny” in Normal Aging.**

<table>
<thead>
<tr>
<th>Age</th>
<th>62–72</th>
<th>68–77</th>
<th>72–82</th>
<th>77–87</th>
<th>82–92</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>42</td>
<td>65</td>
<td>83</td>
<td>63</td>
<td>35</td>
</tr>
<tr>
<td>Words Recalled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>&lt;1</td>
<td>4</td>
<td>12</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>21</td>
<td>27</td>
<td>33</td>
<td>38</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>38</td>
<td>42</td>
<td>46</td>
<td>48</td>
</tr>
<tr>
<td>3</td>
<td>54</td>
<td>55</td>
<td>57</td>
<td>59</td>
<td>59</td>
</tr>
<tr>
<td>(M)</td>
<td>2.79</td>
<td>2.69</td>
<td>2.55</td>
<td>2.33</td>
<td>2.17</td>
</tr>
<tr>
<td>((SD))</td>
<td>0.47</td>
<td>0.58</td>
<td>0.67</td>
<td>0.78</td>
<td>0.95</td>
</tr>
</tbody>
</table>

*Note.* Use age range in which midpoint best corresponds to age of patient.
of verbal memory and learning (Cregger & Rogers, 1998; Small et al., 1999). Older subjects also performed more poorly on the MMSE as a whole than the younger group, although this difference (< 1 point) was of arguably no clinical significance. Mitrushina and Satz (1994) reported that three-word recall correlated more highly with total MMSE score than any other item on the test in a sample of normal, elderly individuals. Therefore, it is not surprising that poorer three-word recall with increasing age also relates to lower overall MMSE scores.

The healthy aging subjects in this study performed quite well on the three-word recall task, with approximately 97% of the younger aging group and 83% of the older group recalling 2/3 or 3/3 words. Therefore, most intact individuals should be able to recall 2 or 3 words after a brief delay. Cullum et al. (1993) reported that only 66% of their healthy sample age 51–95 years recalled 2/3 or 3/3 words using the words rose, ball, and key. This apparent discrepancy seems to highlight the impact of explicit prompts on later recall. Cullum and colleagues did not prompt subjects that they would later be asked to recall the three words again (as per original MMSE instructions), while all subjects in the present study were given such a prompt. Current three-word recall performance is similar to results reported by Mitrushina and Satz (1994; \( M = 2.1; \ SD = 0.9; \) age range = 57–85) and Beardsall and Huppert (1991; \( M = 2.0; \ SD = 1.0; \) age range = 76–93) in their healthy elderly samples, although the current subjects performed slightly better (see Table 1). It is unclear if explicit prompts were given for recall in those studies, however.

Highly discrepant recall scores were obtained by the normal aging and AD samples, again demonstrating that three-word recall is effective in differentiating groups of healthy and demented subjects. The current results also support the proposal of Feher et al. (1992) for a cut-off score of less than two out of three words in screening for dementia. Using this cut-off, 3% of the younger aging group and 16.9% of the older group would have been misclassified as demented (false positives), and 13.2% of the AD sample would have been misclassified as normal (false negatives). The presence of false positives in the normal aging groups highlights both normal variation in the population of normal aging and the limits to reliability due to the restricted range of possible scores on the task. Cut-off scores should not be implemented on three-word recall in isolation for diagnostic purposes, as other authors have noted (Guilmette et al., 1995; Kuslansky et al., 2002). Further, base rates of the condition being examined can impact the sensitivity and specificity estimates from cut-off scores (Rosenfeld, Sands, and Van Gorp, 2000). Therefore, the base rate of AD for the population in which the clinician is giving three word recall (e.g., low in a community setting or higher in a clinical setting) must be considered.

**Education Level**

One of the major difficulties with many studies of memory functioning in normal, elderly individuals is higher than average education levels among research samples. The current sample is no exception, as the average education was approximately three years of college. This is slightly higher than the mean age reported by other studies (e.g., Cullum et al., 1993, \( M \) age = 14.4; Mitrushina & Satz, 1994, \( M \) age = 14.1). Education was not correlated with MMSE total score or three-word recall in the present study. However, the high level of education may reduce the generalizability of these results to lower education groups.

**CONCLUSION**

The majority of normal aging individuals in this study remembered two or three words on three-word recall after a delay when explicit prompts to remember the words “apple, table, and penny” were given, providing practitioners with guidelines of expected performance in these age groups. The younger aging group recalled more words than the older group, and, not surprisingly, both normal aging groups outperformed the AD subjects as a whole. These results further support the use of explicit prompts when giving three-word recall as part of a memory/cognitive screening evaluation, as even when prompted, individuals with significant memory impairment (e.g., persons with AD) forget most or all of the words, while
healthy subjects generally performed well on this task. Given its wide use, further efforts to validate the use of three-word recall as a screening measure are needed across different diagnostic and demographic groups, particularly in establishing standardized methodology (e.g., sets of items) and additional normative references.

ACKNOWLEDGEMENT

Supported in part by NIA grant AG10843.

REFERENCES


