



Published in final edited form as:

Assessment. 2019 September ; 26(6): 1154–1161. doi:10.1177/1073191117723113.

The ModRey: An Episodic Memory Test for Nonclinical and Preclinical Populations

Christiane Hale¹, Briana S. Last¹, Irene B. Meier¹, Lok-Kin Yeung¹, Mariana Budge¹, Richard P. Sloan¹, Scott A. Small¹, Adam M. Brickman¹

¹Columbia University, New York, NY, USA

Abstract

List learning tests are used in practice for diagnosis and in research to characterize episodic memory, but often suffer from ceiling effects in unimpaired individuals. We developed the Modified Rey Auditory Verbal Learning Test, or ModRey, an episodic memory test for use in normal and preclinical populations. We administered the ModRey to 230 healthy adults and to 86 of the same individuals 102 days later and examined psychometric properties and effects of demographic factors. Primary measures were normally distributed without evidence of ceiling effect. Differences between alternate forms were of very small magnitude and not significant. Test–retest reliability was good. Higher participant age and lower participant education was associated with poorer performance across most outcome measures. We conclude that the ModRey is appropriate for episodic memory characterization in normal populations and could be used as an outcome measure in studies involving preclinical populations.

Keywords

memory; Alzheimer’s disease; Rey Auditory Verbal Learning Test; preclinical populations; neuropsychological assessment

Verbal list learning and memory tests are core components of neuropsychological batteries and are used in practice for diagnostic formulation and in research to characterize episodic memory. The typical list learning test comprises a series of semantically related or unrelated words that are read to the patient over a number of trials, each followed by attempts at immediate recall, followed by short delay (on the order of a few minutes) and long delay (on the order of 20 minutes or greater) recall trials. The tasks are designed to examine hippocampal/entorhinal cortex-dependent memory systems and are often used to diagnose or characterize neurological conditions that have a propensity to affect this region (Lillywhite et al., 2007; Tierney et al., 1996).

Reprints and permissions: sagepub.com/journalsPermissions.nav

Corresponding Author: Adam M. Brickman, Taub Institute for Research on Alzheimer’s Disease and the Aging Brain, PS Box 16, 630 West 168th Street, New York, NY 10032, USA. amb2139@columbia.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Over the years, neuropsychological investigations have expanded beyond a focus on diagnosis and characterization of clinical populations to analysis of developmentally normal cognitive functioning, identification of cognitive markers in preclinical populations, and consideration of cognition as a primary outcome measure in behavioral and pharmacological intervention studies. Existing clinical neuropsychological instruments, which were designed with diagnostic specificity and reliability among clinical populations in mind, may not be universally appropriate for this shift in focus for several reasons. Highly diagnostically specific neuropsychological instruments identify individuals with frank impairment with great accuracy but may not be sensitive to detect subtle abnormalities or characterize normal individual differences in function. Similarly, tests designed for clinical populations may not be sensitive to capture change in function over time among healthy and preclinical populations because the expected change in the targeted cognitive construct may be smaller than the amount of measurement error in the test itself. Finally, and perhaps most notably, many neuropsychological instruments designed for clinical populations are simply too easy for nonclinical populations and result in ceiling effects (i.e., invariant performance at or near the maximum possible score).

We created a list learning and memory test designed to address the limitations of standard neuropsychological instruments for use in normal and preclinical populations. We based our test on the Rey Auditory Verbal Learning Test (RAVLT; Lezak, 1983), one of the most widely used verbal list learning test that was originally created in 1941 (Taylor, 1959). The RAVLT has been well validated (Schmidt, 1996), has excellent test–retest reliability (Knight, McMahon, Skeaff, & Green, 2007), and has been used in several clinical populations known to have hippocampal-dependent memory abnormalities (Barzotti et al., 2004). However, among nonclinical populations, the RAVLT has well-documented ceiling effects (Davis et al., 2003; Drolet et al., 2014; Sullivan, Deffenti, & Keane, 2002; Uttl, 2005), which limits its utility among individuals without frank memory impairment. We modified the content and structure of the RAVLT to create the “Modified Rey Auditory Verbal Learning Test,” or *ModRey*, for use in normal and preclinical populations. Our modification included several notable features and was guided by literature focused on increasing the utility of memory tests in nonclinical populations (Uttl, 2005): First, we increased the number of learning list (“List A”) and distractor list (“List B”) items from 15 to 20. Second, we decreased the number of learning trials from 5 to 3. Third, after short (~5 minutes) and long (~30–60 minutes) free recall trials of List A, we added a free recall trial of List B. Fourth, after a forced-choice recognition trial, we added a source memory trial in which the participant indicates whether each presented word came from List A or List B. We initially created two alternate forms of the ModRey for use in repeated-measures observational research or in clinical trials.

Our focus on a list-learning test of declarative memory, which is conceptualized as the ability to explicitly store and retrieve information, was motivated by several factors. We have been interested in understanding the complexity of the hippocampal formation and the subfields that comprise its circuitry (Brickman, Small, & Fleisher, 2009; Small, Schobel, Buxton, Witter, & Barnes, 2011). The entorhinal cortex is the region that is affected earliest in Alzheimer’s disease (Khan et al., 2014), prior to the onset of frank dementia. Models of hippocampal circuitry function suggest that the entorhinal cortex is involved with the

retention of information of brief and delayed time periods (Small et al., 2011), which converges with the observation that a loss in memory retention of words is among the most sensitive neuropsychological indicators of Alzheimer's disease (Albert, 1996). Our primary motivation, therefore, was to develop a test that could capture individual differences in entorhinal cortex function in normal populations and in those at particular risk for the development of Alzheimer's disease. We previously demonstrated the validity of the retention score on the ModRey with cerebral blood volume functional magnetic resonance imaging by showing a double anatomical dissociation: increased performance on the ModRey was selectively associated with greater cerebral blood volume (CBV) in the entorhinal cortex but not with CBV in the dentate gyrus, whereas increased performance on an object recognition test correlated with increased CBV in the dentate gyrus but not in the entorhinal cortex (Brickman et al., 2014). The purpose of this study was to examine further the psychometric properties of the ModRey by considering performance distributions, test–retest reliability, systematic performance differences in alternative forms, and the association of basic demographic features with ModRey performance metrics.

Method

Test Construction

Task Design—Figure 1 illustrates the ModRey administration paradigm. During learning Trials 1 to 3, the participant is read 20 unique, semantically/phonemically unrelated words (List A) and is asked to free recall those words after each trial. Immediately following administration of the three learning trials, the participant is read a list of 20 unique, semantically/phonemically unrelated words (List B) and is asked to recall as many of those words as possible. This trial is followed by a short delay free recall (SDFR) trial in which the participant is asked to recall as many words as possible from List A. After a 1.5-hour period, during which other tasks that do not interfere with performance on this test are administered, the participant is first asked to long delay free recall (LDFR) words from List A and then from List B. These trials are followed by a 66-item forced recognition trial (Recog), in which the participant is read a list of 66 words and asked to distinguish List A words (targets) from semantic and phonemic distractors, including words from List B. Finally, a source memory trial is administered in which the examiner serially reads words from List A and List B and asks the participant to specify from which list each word came.

Outcome Variables—There are several outcome measures on the ModRey, including the number of words recalled for each of the three learning trials, the total number of scores across the three learning trials, total words recalled on the List B immediate trial, total SDFR words for Lists A and B, total LDFR words for Lists A and B, the number of target words correctly identified on the recognition trial (i.e., “hits”), the number of distractor words incorrectly endorsed on the recognition trial (i.e., “false positives”), recognition discrimination (i.e., hits—false positives), and the number of words correctly identified on the source memory trial. It is also possible to derive short and long delayed “retention” scores, which reflect the ratio of items recalled at SDFR and LDFR to the number of items learned during the learning trials or the number of words recalled on the last learning trial; we believe that both short and long retention values closely capture cognitive processes

mediated by the entorhinal cortex specifically (Brickman, Stern, & Small, 2011; Small et al., 2011).

Word Lists—ModRey Lists A and B comprise 20 semantically and phonemically unrelated words. To create these lists, we concatenated word lists from the original RAVLT (Taylor, 1959), words that had been selected and successfully tested for reliability by Crawford and colleagues (Crawford, Stewart, & Moore, 1989), and words evaluated by Majdan and colleagues (Majdan, Sziklas, & Jones-Got-man, 1996; Strauss, Sherman, & Spreen, 2006). All words were originally matched for frequency, syllabic count, and semantic association, and are considered equivalent. To create the recognition trials, we maintained the same ratio of target words to distractors from the original RAVLT; in addition to the 20 target words, the 20 words from List B, and 26 semantically and/or phonemically related words were selected from original recognition lists as additional distractors. Two versions of each list were generated to create alternative forms of the test for repeated assessment. Table 1 displays the word lists used for the two ModRey forms.

Participants

All 230 participants examined for the current study were drawn from three separate studies that examined the effect of aerobic exercise in individuals ranging from 20 to 75 years of age. All three studies shared common inclusion criteria, including (a) absence of major medical or psychiatric histories or active conditions, (b) sedentary fitness levels, and (c) lack of cognitive impairment among individuals above 50 years of age, defined by Dementia Rating Scale (Mattis, 1988) scores of 135 or less. Baseline data, prior to the implementation of any study intervention, were used for primary analyses. During the delay intervals, other neuropsychological measures, including the Stroop test, tests of verbal fluency, Digit Symbol Modalities Test, tests of object recognition, and a computerized battery that assessed attention and executive functioning, were administered; no other tests of verbal memory were included in the battery. For test–retest reliability analyses, we included data from the second visit, approximately 14 weeks after the baseline visit, of participants who were randomized to the nonintervention control group. Test form (Form 1 vs. Form 2) was randomly selected for the baseline visit and then counterbalanced at the follow-up visit. Table 2 displays participant characteristics.

Statistical Analysis

Three sets of analyses were conducted. First, to determine whether the two test forms can be considered equivalent, we compared performance on each of the outcome measures between individuals who received Form 1 and those who received Form 2 at baseline with a series of *t* tests. Although the two groups were statistically similar in age ($t = 1.62, p = .106$), those who received Form 2 were slightly younger than those receiving Form 1 (38.72 [14.44] vs. 41.93 [15.61]), so we repeated the analyses after controlling for age with analysis of variance to adjust for these subtle age differences. Second, once we established practical form equivalency, we examined test–retest reliability by examining the correlation between performance at baseline and follow-up for each outcome measure using Pearson bivariate correlation statistics; for these analyses, we “collapsed” across the two test forms. We also examined differences between baseline and follow-up on each outcome measure with paired

t test analyses. Following Woods, Delis, Scott, Kramer, and Holdnack (2006), we generated difference scores between follow-up and baseline (follow-up – baseline) and computed the standard deviations of the mean differences scores (M_{diff}), also referred to as standard error of differences or SE_{diff} . These values were used to generate 90% confidence intervals (CIs) so that users of the instrument can compute reliable change indices (RCIs) for individual participants or patients; the formula to compute the 90% CI was $M_{\text{diff}} \pm (SE_{\text{diff}} * 1.645$; Woods et al., 2006). Third, we examined the relationship between key demographic factors, including age, sex, and number of years of education, with each of the outcome measures with regression analysis; separate regression models for each demographic factor were run for each of the key outcome variables. We explored ethnicity- and race-related differences on three key outcome variables (Total Learning List A, SDFR List A Retention, LDFR List A Retention) by comparing performance between non-Hispanic ($n = 179$) and Hispanic participants ($n = 51$) and between White ($n = 91$) and non-White ($n = 139$) participants with *t* tests. We note that the study was not designed to examine ethnicity/race effects so results should be interpreted with caution.

Results

Form Equivalency

Forty-seven percent ($n = 107$) of the participants were evaluated with Form 1 at baseline, and 53% ($n = 123$) of the participants were evaluated with Form 2. Individuals in both groups were similar in age (41.93 [15.61] vs. 38.72 [14.44] years; $t = 1.62$, $p = .106$), sex distribution (64.5% women vs. 56.9% women; $\chi^2 = 1.37$, 0.24), and education (17.70 [2.85] vs. 17.22 [2.72] years; $t = 1.30$, $p = .194$). As displayed in Table 3, overall, we found no difference in performance on the two test forms. Significantly more false-positive errors were made on the recognition trial of Form 1, resulting in a significantly lower discrimination index. However, the differences in false-positive errors and discrimination were attenuated when controlling for the slight difference in age between those tested with Form 1 and those tested with Form 2 ($F[1, 229] = 3.79$, $p = .053$, and $F[1, 229] = 2.73$, $p = .10$, respectively), as was the trend-level difference in Trial 1 List A ($F[1, 229] = 2.56$, $p = .11$). It is important to note several test characteristics within the forms. On average, performance metrics did not evidence either “floor” or “ceiling” effects but did show a distribution of performance; apart from the recognition trials, which were slightly negatively skewed, the frequency distributions of the outcome measures were normal. Frequency histograms for key outcome variables are displayed in Figure 2 and interquartile range values are included in Table 3. Across the three learning trials, participants on average evidenced expected learning effects and subsequent forgetting at short delay and, more so, at long delay. Similarly, recognition discriminability and source memory indices showed overall good performance but with some variability across participants.

Test–Retest Reliability

Test–retest reliability statistics for measured variables are displayed in Table 4. Intervals between baseline and follow-up testing ranged from 83 to 216 days (mean = 102 ± 20.2). Overall, test–retest reliability was good across outcome measures. Apart from the first

learning trial, performance across the outcome measures between the visits was statistically similar.

Demographic Factors

The influence of demographic factors, including age, sex, and number of years of education, is displayed in Table 5. Increased age was associated with poorer performance on all metrics apart from the number of recognition hits. Similarly, lower number of years of education was associated with lower scores on all measures, apart from recognition false positive errors and recognition discrimination. However, there were no systematic differences in performance across men and women. Performance on the three key outcome variables explored did not differ between Hispanic and non-Hispanic participants (all p s > .55); similarly, there were no differences in performance between White and non-White participants (all p s > .32).

Discussion

We created a list learning and memory test, the ModRey, that was designed to capture declarative memory performance among nonclinical and preclinical populations. Based on the Rey Auditory Verbal Learning Test (Taylor, 1959), the ModRey requires participants to learn and recall word lists containing 20 semantically and phonetically unrelated items. Cognitive operations assessed by the ModRey include learning, recall and retention, interference, recognition memory, and source memory. We designed the ModRey to be used in nonclinical and preclinical populations. That is, our goal was to assess individual differences in memory functioning within a normal range of ability, while avoiding ceiling and floor effects. Previous work with the ModRey (Brickman et al., 2014) showed that memory retention performance correlates selectively with function of the entorhinal cortex, as assessed by high-resolution functional magnetic resonance imaging, and not with function in other hippocampus-related subfields. In that study, we did not collect data with other list learning and memory measures, so we were not able to establish whether performance on the ModRey is more strongly associated with entorhinal cortex function than other similar measures, but we do believe this double anatomical dissociation provides evidence for the validity of the ModRey test. Furthermore, the administration format of the test, in which participants are required to learn a list of words and recall them at shorter and longer delay intervals, mirrors a well-validated format that has been employed for more than 100 years to assess episodic memory.

In the current study, we examined other psychometric properties of the ModRey, including the general equivalency of alternative forms, test–retest reliability, and the influence of demographic factors on performance. Overall, we showed that the performance on the two alternate forms was statistically very similar. Slight differences in first trial learning and recognition memory across the forms were attenuated when we accounted for the small differences in age between the two groups of participants that were tested with each form. We conclude that the two constructed forms can be used interchangeably. The advantage of having multiple forms of the same instrument are that it can be employed in repeat-assessment studies, such as clinical trials, in which practice or carryover effects are a

potential source of error in memory. In terms of test–retest reliability, we showed adequate reliability across the ModRey measured metrics, which was comparable with or better than previous reports (Geffen, Butterworth, & Geffen, 1994; Ryan, Geisser, Randall, & Georgemiller, 1986) and other efforts with similar scales, although perhaps a little low for clinical standards. We also showed that, overall, mean performance did not change reliably between the two visits, apart from a slight improvement for the first learning trial. We provide 90% CIs to compute RCIs. Test–retest intervals did vary across participants and we speculate that reliability coefficients would have been higher had the interval been consistently shorter.

Our consideration of demographic variables showed strong effects of age and education, but not sex. Participants in this study were well screened and deemed neurologically and psychiatrically healthy. Thus, normative data, stratified by age and/or education, can be derived from these data to establish standardized measures of function or impairment. We did not observe significant differences between ethnicity or race groups on the variables considered, although the study was not designed to consider these demographic variables explicitly.

We believe that the ModRey is a valuable instrument for the assessment of memory in unimpaired populations. The ModRey will have utility in observational studies that seek to interrogate memory function and in clinical trials in which memory function is a key outcome. Importantly, with greater emphasis on studying Alzheimer’s disease in its earliest or preclinical states, the ModRey would likely be sensitive enough to detect subtle disease-related variability in memory function, change over time, and, potentially, response to pharmacological treatment or other intervention.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by grants from the U.S. National Institutes of Health Grants AG035015, AG008702, AG029949, the James S. McDonnell Foundation, and an unrestricted grant by Mars, Inc.

References

- Albert MS. 1996; Cognitive and neurobiologic markers of early Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America*. 93:13547–13551. [PubMed: 8942970]
- Barzotti T, Gargiulo A, Marotta MG, Tedeschi G, Zannino G, Guglielmi S, Marigliano V. 2004; Correlation between cognitive impairment and the Rey Auditory-Verbal Learning Test in a population with Alzheimer disease. *Archives of Gerontology and Geriatrics, Supplement*. (9):57–62. DOI: 10.1016/j.archger.2004.04.010 [PubMed: 15207397]
- Brickman AM, Khan UA, Provenzano FA, Yeung LK, Suzuki W, Schroeter H, Small SA. 2014; Enhancing dentate gyrus function with dietary flavanols improves cognition in older adults. *Nature Neuroscience*. 17:1798–1803. DOI: 10.1038/nn.3850 [PubMed: 25344629]
- Brickman AM, Small SA, Fleisher A. 2009; Pinpointing synaptic loss caused by Alzheimer’s disease with fMRI. *Behavioural Neurology*. 21:93–100. DOI: 10.3233/BEN-2009-0240 [PubMed: 19847048]

- Brickman AM, Stern Y, Small SA. 2011; Hippocampal subregions differentially associate with standardized memory tests. *Hippocampus*. 21:923–928. DOI: 10.1002/hipo.20840 [PubMed: 20824727]
- Crawford JR, Stewart LE, Moore JW. 1989; Demonstration of savings on the AVLT and development of a parallel form. *Journal of Clinical and Experimental Neuropsychology*. 11:975–981. DOI: 10.1080/01688638908400950 [PubMed: 2592534]
- Davis HP, Small SA, Stern Y, Mayeux R, Feldstein SN, Keller FR. 2003; Acquisition, recall, and forgetting of verbal information in long-term memory by young, middle-aged, and elderly individuals. *Cortex*. 39:1063–1091. [PubMed: 14584567]
- Drolet V, Vallet GT, Imbeault H, Lecomte S, Limoges F, Joubert S, Rouleau I. 2014; A comparison of the performances between healthy older adults and persons with Alzheimer's disease on the Rey auditory verbal learning test and the Test de rappel libre/rappel indice 16 items. *Gériatrie et Psychologie Neuropsychiatrie du Vieillissement*. 12:218–226. DOI: 10.1684/pnv.2014.0469 [PubMed: 24939409]
- Geffen GM, Butterworth P, Geffen LB. 1994; Test-retest reliability of a new form of the Auditory Verbal Learning Test (AVLT). *Archives of Clinical Neuropsychology*. 9:303–316. [PubMed: 14589623]
- Khan UA, Liu L, Provenzano FA, Berman DE, Profaci CP, Sloan R, Small SA. 2014; Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nature Neuroscience*. 17:304–311. DOI: 10.1038/nn.3606 [PubMed: 24362760]
- Knight RG, McMahon J, Skeaff CM, Green TJ. 2007; Reliable Change Index scores for persons over the age of 65 tested on alternate forms of the Rey AVLT. *Archives of Clinical Neuropsychology*. 22:513–518. DOI: 10.1016/j.acn.2007.03.005 [PubMed: 17452090]
- Lezak, MD. *Neuropsychological assessment*. 2nd. New York, NY: Oxford University Press; 1983.
- Lillywhite LM, Saling MM, Briellmann RS, Weintrob DL, Pell GS, Jackson GD. 2007; Differential contributions of the hippocampus and rhinal cortices to verbal memory in epilepsy. *Epilepsy & Behavior*. 10:553–559. DOI: 10.1016/j.yebeh.2007.03.002 [PubMed: 17452129]
- Majdan A, Sziklas V, Jones-Gotman M. 1996; Performance of healthy subjects and patients with resection from the anterior temporal lobe on matched tests of verbal and visuo-perceptual learning. *Journal of Clinical and Experimental Neuropsychology*. 18:416–430. DOI: 10.1080/01688639608408998 [PubMed: 8877625]
- Mattis, S. *Dementia Rating Scale (DRS)*. Odessa, FL: Psychological Assessment Resources; 1988.
- Ryan JJ, Geisser ME, Randall DM, Georgemiller RJ. 1986; Alternate form reliability and equivalency of the Rey Auditory Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology*. 8:611–616. DOI: 10.1080/01688638608405179 [PubMed: 3805255]
- Schmidt, M. *Rey Auditory and Verbal Learning test: A handbook*. Los Angeles, CA: Western Psychological Services; 1996.
- Small SA, Schobel SA, Buxton RB, Witter MP, Barnes CA. 2011; A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nature Reviews Neuroscience*. 12:585–601. DOI: 10.1038/nrn3085 [PubMed: 21897434]
- Strauss, E, Sherman, EMS, Spreen, O. *A compendium of neuropsychological tests: Administration, norms, and commentary*. 3rd. New York, NY: Oxford University Press; 2006.
- Sullivan K, Deffenti C, Keane B. 2002; Malinger on the RAVLT: Part II. Detection strategies. *Archives of Clinical Neuropsychology*. 17:223–233. DOI: 10.1016/S0887-6177(01)00110-X [PubMed: 14589725]
- Taylor, EM. *The appraisal of children with cerebral deficits*. Cambridge, MA: Harvard University Press; 1959.
- Tierney MC, Szalai JP, Snow WG, Fisher RH, Nores A, Nadon G, St George-Hyslop PH. 1996; Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study. *Neurology*. 46:661–665. DOI: 10.1212/wnl.46.3.661 [PubMed: 8618663]
- Uttl B. 2005; Measurement of individual differences: Lessons from memory assessment in research and clinical practice. *Psychological Science*. 16:460–467. DOI: 10.1111/j.0956-7976.2005.01557.x [PubMed: 15943672]

Woods SP, Delis DC, Scott JC, Kramer JH, Holdnack JA. 2006; The California Verbal Learning Test–Second Edition: Test-retest reliability, practice effects, and reliable change indices for the standard and alternate forms. *Archives of Clinical Neuropsychology*. 21:413–420. DOI: 10.1016/j.acn.2006.06.002 [PubMed: 16843636]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Figure 1.

ModRey design schematic.

Note. A = List A; B = List B; SDFR = short delay free recall; LDFR = long delay free recall; Recog = recognition trial; Source = source memory trial.

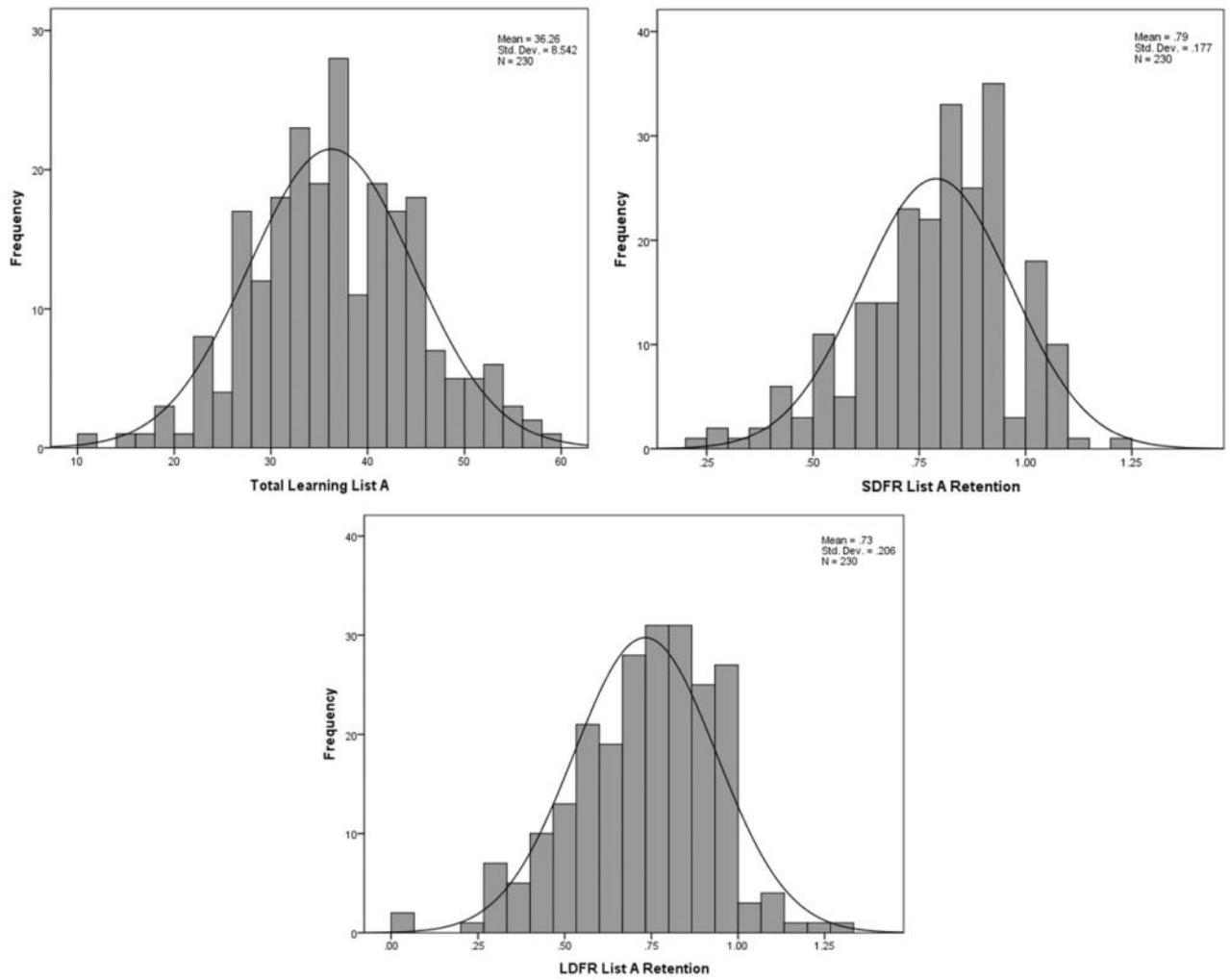


Figure 2. Frequency histograms illustrating lack of ceiling effects for three key ModRey outcome variables.

Table 1

Words Lists Used for the Two Versions of the ModRey.

	List A		List B		Recog					
	Form 1	Form 2	Form 1	Form 2	Form 1	Form 2				
Drum	Doll	Mirror	Desk	Dish	Bell	Window	Water	Nail	Forest	Pear
Curtain	Mirror	Mirror	Ranger	Jester	Window	Hat	Farmer	Sand	Sailor	Street
Bell	Nail	Sailor	Bird	Hill	Hat	Violin	Rose	Bed	Dart	Machine
Coffee	Sailor	Sailor	Shoe	Coat	Violin	River	Gloves	Pony	Train	Head
School	Heart	Heart	Stove	Tool	Barn	Towel	Cloud	Jester	Road	Army
Parent	Desert	Desert	Mountain	Forest	Ranger	Frog	House	Milk	Ladder	Girl
Moon	Face	Face	Glasses	Water	Nose	Curtain	Stranger	Plate	Corn	Horse
Garden	Letter	Letter	Towel	Ladder	Cousin	Flower	Ham	Field	Mirror	Soot
Hat	Bed	Bed	Cloud	Girl	Weather	Color	Garden	Heart	Armchair	Letter
Farmer	Machine	Machine	Boat	Foot	School	Knife	Glasses	Jail	Screw	Peel
Nose	Milk	Milk	Lamb	Shield	Hand	Desk	Stocking	Hunter	Music	Water
Turkey	Helmet	Helmet	Gun	Pie	Grass	Gun	Piano	Insect	Dish	Joker
Color	Music	Music	Pencil	Insect	Pencil	Crayon	Shoe	Orange	Bucket	Coat
House	Horse	Horse	Church	Ball	Home	Stair	Teacher	Envelope	Pie	Captain
River	Road	Road	Fish	Car	Fish	Church	Stove	Car	Wood	Cork
Violin	Banana	Banana	Cousin	Orange	Earth	Turkey	Suitcase	Banana	Ball	Tool
Tree	Radio	Radio	Earth	Armchair	Moon	Fountain	Nest	Face	Bus	Fly
Scarf	Hunter	Hunter	Knife	Toad	Finger	Scarf	Children	Toad	Helmet	Song
Ham	Bucket	Bucket	Stair	Cork	Balloon	Boat	Drum	Blanket	Stool	Doll
Suitcase	Field	Field	Dog	Bus	Lunchbox	Hot	Chest	Silk	Foot	Doctor
					Bird	Parent	Toffee	Radio	Bread	Stall
					Mountain	Dog	Lamb	Hill	Desert	Shield

Table 2

Participant Characteristics at Baseline and Follow-Up Visits.

	Baseline	Follow-up
<i>N</i>	230	86
Age (mean \pm SD years, range)	40.21 \pm 15.05; range: 20–75	37.74 \pm 14.52; range: 20–70
Sex (<i>n</i> , % women)	139, 60.4%	53, 61.6%
Education (mean \pm SD years)	17.45 \pm 2.79	17.26 \pm 2.55
Ethnicity		
Non-Hispanic (<i>n</i> , %)	179, 78%	66, 77%
Hispanic (<i>n</i> , %)	51, 22%	20, 23%
Race		
White (<i>n</i> , %)	91, 39.6%	28, 33%
African American (<i>n</i> , %)	43, 19%	18, 21%
Asian (<i>n</i> , %)	44, 19%	21, 24%
Hawaiian/Pacific Islander (<i>n</i> , %)	2, 0.9%	0, 0%
Other (<i>n</i> , %)	50, 22%	19, 22%
Test form (<i>n</i> , % tested with Form 1)	107, 47%	42, 49%

Table 3

Performance Metrics on the Alternate Forms of the ModRey.

Outcome measure	Form 1	Form 2	Combined	IQR	Statistic
List A Trial 1	8.31 (2.68)	9.07 (3.12)	8.71 (2.94)	7–10	$t = -1.96, p = .051$
List A Trial 2	12.23 (3.23)	12.88 (3.12)	12.58 (3.18)	10–15	$t = -1.54, p = .13$
List A Trial 3	14.96 (3.05)	14.97 (3.33)	14.97 (3.19)	13–17	$t = -0.01, p = .99$
Total Learning List A (Trials 1 to 3)	35.51 (8.21)	36.91 (8.81)	36.26 (8.54)	30–42	$t = -1.25, p = .21$
List B	7.82 (2.74)	7.61 (3.25)	7.71 (3.02)	6–10	$t = 0.53, p = 0.60$
SDFR List A	12.00 (4.13)	12.11 (4.21)	12.06 (4.16)	9–15	$t = -0.21, p = 0.84$
SDFR List A Retention	0.78 (.18)	0.80 (0.18)	0.79 (0.18)	0.70–0.93	$t = -0.45, p = 0.65$
LDfR List A	11.28 (4.58)	11.28 (4.48)	11.28 (4.52)	8–15	$t = 0.00, p = 1.00$
LDfR List A Retention	0.73 (0.21)	0.73 (0.20)	0.73 (0.21)	0.60–0.88	$t = -0.03, p = 0.97$
LDfR List B	3.56 (3.17)	3.20 (3.21)	3.37 (3.20)	1–5	$t = 0.87, p = 0.39$
LDfR List B Retention	0.41 (0.31)	0.37 (0.27)	0.39 (0.29)	0.17–0.60	$t = 1.23, p = 0.22$
Recognition Hits	16.94 (2.93)	16.90 (2.58)	16.92 (2.74)	15–19	$t = 0.11, p = 0.91$
Recognition False Positives	4.62 (6.15)	3.04 (3.86)	3.77 (5.11)	0–5	$t = 2.36, p = 0.02$
Recognition Discriminability (Hits-False Positives)	12.15 (6.15)	13.86 (4.87)	13.07 (6.40)	10–18	$t = -2.04, p = 0.04$
Source—A	16.51 (3.02)	16.78 (2.78)	16.66 (2.89)	15–19	$t = -0.70, p = 0.49$
Source—B	17.32 (3.26)	17.48 (2.65)	17.40 (2.94)	16–20	$t = -0.42, p = 0.68$

Note. Interquartile range (IQR) shows the lower (25th percentile) and upper (75th percentile) bounds. A = List A; B = List B; SDFR = short delay free recall; LDfR Retention = long delay free recall retention score, calculated as the ratio of List A trial 3 to LDfR; Source = source memory trial.

Table 4

Test–Retest Reliability.

Outcome measure	Baseline	Follow-up	Agreement	Difference	M_{diff}	SD	90% CI
List A Trial 1	8.73 (2.86)	9.35 (3.11)	$r = .63, p < .001$	$t = 2.21, p = .03$	0.62	2.59	-3.64, 3.21
List A Trial 2	12.8 (3.05)	13.29 (3.47)	$r = .62, p < .001$	$t = 1.57, p = .12$	0.49	2.88	-4.25, 5.23
List A Trial 3	15.38 (2.93)	15.04 (3.13)	$r = .55, p < .001$	$t = 1.16, p = .25$	-0.35	2.79	-4.94, 4.24
Total Learning List A (Trials 1 to 3)	36.92 (8.05)	37.67 (8.87)	$r = .69, p < .001$	$t = 1.05, p = .30$	0.76	6.68	-10.23, 7.44
List B	8.02 (3.32)	8.12 (3.48)	$r = .60, p < .001$	$t = 0.28, p = .78$.093	3.04	-4.91, 5.09
SDFR List A	12.56 (3.76)	12.37 (3.59)	$r = .57, p < .001$	$t = 0.44, p = .66$	-0.19	3.95	-6.69, 6.31
LDFR List A	11.58 (4.14)	11.64 (4.69)	$r = .65, p < 0.001$	$t = 0.15, p = .89$	0.06	3.71	-6.04, 6.16
LDFR List B	3.59 (3.65)	3.48 (3.75)	$r = .60, p < .001$	$t = 0.32, p = .075$	-0.12	3.33	-5.60, 5.36
Recognition Hits	17.07 (2.73)	17.3 (2.48)	$r = .64, p < .001$	$t = 0.98, p = .33$	0.23	2.21	-3.41, 3.87
Recognition False Positives	2.77 (3.79)	2.74 (3.85)	$r = .56, p < .001$	$t = 0.06, p = .95$	-0.02	3.59	-5.93, 5.70
Source—A	16.93 (2.89)	17.07 (2.61)	$r = .53, p < .001$	$t = 0.48, p = .63$	0.14	2.69	-4.29, 4.57
Source—B	17.72 (2.97)	18.23 (2.4)	$r = .49, p < .001$	$t = 1.72, p = .89$	0.51	2.76	-4.03, 5.05

Note. Pearson correlations are used to show the degree of agreement across testing sessions and within-participant t tests are used to show degree of difference. The mean difference (M_{diff}), follow-up – baseline (performance) score and standard deviation (SD) of the mean difference scores are used to generate 90% confidence interval (CI) scores, which can be used to generate RCIs. Test–retest reliability statistics for baseline and follow-up. Forms 1 and 2 were combined, administered an equal number of times within each session, and counterbalanced.

Table 5

Influence of Demographic Factors on Performance.

Outcome measure	Age	Sex	Education
List A Trial 1	$\beta = -.050, p < .001$	$\beta = .366, p = .357$	$\beta = .257, p < .001$
List A Trial 2	$\beta = -.064, p < .001$	$\beta = -.193, p = .653$	$\beta = .307, p < .001$
List A Trial 3	$\beta = -.073, p < .001$	$\beta = .076, p = .861$	$\beta = .282, p < .001$
Total Learning List A (Trials 1 to 3)	$\beta = -.187, p < .001$	$\beta = .248, p = .830$	$\beta = .847, p < .001$
List B	$\beta = -.062, p < .001$	$\beta = .355, p = .385$	$\beta = .187, p = .008$
SDFR List A	$\beta = -.100, p < .001$	$\beta = .081, p = .886$	$\beta = .390, p < .001$
SDFR List A Retention	$\beta = -.003, p < .001$	$\beta < .001, p = .988$	$\beta = .011, p = .007$
LDFR List A	$\beta = -.119, p < .001$	$\beta = .249, p = .684$	$\beta = .494, p < .001$
LDFR List A Retention	$\beta = -.005, p < .001$	$\beta = .013, p = .646$	$\beta = .019, p < .001$
LDFR List B	$\beta = -.070, p < .001$	$\beta = .287, p = .507$	$\beta = .238, p = .002$
LDFR List A Retention	$\beta = -.006, p < .001$	$\beta = .020, p = .606$	$\beta = .018, p = .009$
Recognition Hits	$\beta = -.017, p = .160$	$\beta = -.234, p = .528$	$\beta = .183, p = .005$
Recognition False Positives	$\beta = .106, p < .001$	$\beta = -.317, p = .646$	$\beta = -.133, p = .273$
Recognition Discriminability (Hits-False Positives)	$\beta = -.118, p < .001$	$\beta = .383, p = .658$	$\beta = .288, p = .059$
Source—A	$\beta = -.050, p < .001$	$\beta = -.141, p = .719$	$\beta = .168, p = .014$
Source—B	$\beta = -.071, p < .001$	$\beta = .331, p = .405$	$\beta = .207, p = .003$