Original Article

The Rey Auditory Verbal Learning Test: Alternate Forms Equivalency and Reliability for the Iranian Adult Population (Persian Version)

Mehrnaz Rezvanfard MD[1], Hamed Ekhtiari MD•[1,4], Maryam Noroozian MD[2], Alireza Rezvanifar MSc[1], Reza Nilipour PhD[3], Gelavizh Karimi Javan MSc[3]

Abstract

Background: Increasing demand for memory assessment in clinical settings in Iran, as well as the absence of a comprehensive and standardized task based upon the Persian culture and language, requires an appropriate culture- and language-specific version of the commonly used neuropsychological measure of verbal learning and memory, the Rey Auditory Verbal Learning Test (RAVLT).

Methods: The Persian adapted version of the original RAVLT and two other alternate word lists were generated based upon criteria previously set for developing new word lists. A total of 90 subjects (three groups of 30 persons), aged 29.75±7.10 years, volunteered to participate in our study and were tested using the original word list. The practice effect was assessed by retesting the first and second groups using the same word list after 30 and 60 days, respectively. The test-retest reliability was evaluated by retesting the third group of participants twice using two new alternate word lists with an interval of 30 days.

Results: The re-administration of the same list after one or even two months led to significant practice effects. However, the use of alternate forms after a one-month delay yielded no significant difference across the forms. The first and second trials, as well as the total, immediate, and delayed recall scores showed the best reliability in retesting by the alternate list.

Conclusion: The difference between the generated forms was minor, and it seems that the Persian version of the RAVLT is a reliable instrument for repeated neuropsychological testing as long as alternate forms are used and scores are carefully chosen.

Keywords: alternative lists, Iranian population, Persian version, practice effect, Rey auditory verbal learning test reliability

Introduction

Memory complaints in outpatient settings, particularly in the elderly populations, appear to be the most frequent reason for neuropsychological referral[1] and should be borne in mind in tandem with the fact that deficits in memory processes are begotten by a variety of common neurological and psychiatric disorders. [1–6]

According to the latest national population estimates of 2007 (Statistical Center of Iran), Iran is one of the youngest populations in the world (77.41% of the population is less than 40 years old). The anticipated increase in the elderly population and concomitant rise in demand for memory assessment prompted us to address the current lack of a comprehensive and standardized task based upon the Persian culture and language by developing an appropriate culture- and language-specific version of commonly used neuropsychological measures of verbal learning and memory, e.g., the Rey auditory verbal learning test (RAVLT). The RAVLT is generally utilized for memory evaluation in scientific research and in clinical practice[1] in that it is easily administrable and confers the encoding, consolidating, storing, and retrieving of verbal information via a five-trial presentation of a 15-word list (list A), a single presentation of an interference list (list B), two post-interference recall trials (one immediate and one delayed), and recognition of the target words presented with distracters. Measures essential for an understanding of the kind and severity of a patient's memory deficits can, consequently, be straightforwardly obtained.[1] The original French words[7] and their order have been translated into English without modification. Other language versions such as Flemish,[8] Spanish,[9] Hebrew,[10] German,[11] Chinese,[12] and Greek[13] are also provided.

The following linguistic criteria were established for the generation of word lists and additional forms[14, 15]:

- The chance of occurrence and frequency of the selected words based upon a large scale contemporary database in common usage with Thorndike-Lorge Tables as

Authors' affiliations: [1]Neurocognitive Laboratory, Iranian National Center for Addiction Studies (INCAS), Tehran University of Medical Sciences, Tehran, Iran[2]. Memory & Behavioral Neurology Department (MBND), Psychiatry Department, Roozbeh Hospital, Tehran University of Medical Sciences, [3]Department of Speech Therapy, University of Welfare & Rehabilitation Sciences-Evin, Tehran, Iran, [4]Translational Neuroscience Program, Institute for Cognitive Sciences Studies (ICSS), Tehran, Iran.
•Corresponding author and reprints: Hamed Ekhtiari MD, Neurocognitive Laboratory, Iranian National Center for Addiction Studies, Tehran University of Medical Sciences, Tehran, Iran. Tel: +98-215-542-1177
Fax: +98-215-542-1166, E-mail: h_ekhtiari@razi.tums.ac.ir
Accepted for publication: 21 July 2010

guidelines.[16] Only the most frequent words were used because of word usage and the likelihood of recall effect.[17]

• The imagery value using Pavio's Table.[18] Words with high rates of imagery value were selected.

• Word length measured by the number of letters and syllables. Only one- or two-syllable nouns were used to generate the lists.

• Control of any obvious semantic ambiguity and/or phonetic similarities or associations between the words on the same list.

The assessment of change is often critical in the neuropsychological evaluation, whether in research or in clinical settings. The repeated use of instruments confounds this assessment inasmuch as practice effects (gains in performance by prior experiences with the test) have been demonstrated for many measures.[19,20] Practice effects are particularly expected with memory testing on account of the fact that the learning gained during the preliminary evaluations tends to be transferred to the following ones.[14] A practice effect may be related to the fact that patients explicitly remember test items or the testing format (test sophisticated effects) previously presented.[21] It is possible to minimize the practice effect by using alternate forms of a test instead of retesting by the same form.[9,20] Several alternative forms have been developed for the RAVLT such as those created by: Lezak; Shapiro and Harrison; Crawford et al.; Majdan et al.; and Geffen et al.[1,14,15,22] They have been proven useful, particularly in longitudinal or sequential research designs where baseline performance, experimental manipulation, and recovery time are expected to feature within clinical populations.

Measurement error is another effect of significance in so far as it can baffle a patient's actual score in a test-retest design even when alternative forms are employed. Measurement error is often construed as the concept of reliability, which incorporates test-retest reliability (TRR). A score with a good TRR is free of measurement error unrelated to chance variance, and the second session of testing is likely to result in subjects having scored the highest remaining amongst the best.[20] It is also worthy of note that highly significant reliability coefficients were reported while studying different alternative lists of the RAVLT.[14, 20, 22]

The present study aimed at developing a Persian version of the original word list of the RAVLT and two other alternate word lists from Lezak and Shapiro & Harisson, taking into account the linguistic criteria specific to the Persian culture and language settings. Practice effect was measured using different word lists at different retest intervals and alternate form reliability coefficients were evaluated for each trial.

Materials and Methods

Participants

Ninety volunteers ages 18 to 50 (29.75±7.10 years) participated in the study after providing written informed consent. The study population was comprised of 44 men and 46 women, with educational levels ranging from 2 to 20 years (11.56±4.91). The exclusion criteria were history of brain surgery or psychiatric symptoms and any type of medical condition or using a medication that might affect cognitive functioning.

Word selection procedures

The participants were assessed using the Persian adapted list of the RAVLT and two other alternate forms, (Lezak and Shappiro & Harisson), based on the original English word lists reviewed by Hawkins et al. (2004).[14] The original English words, presented in Table 1, were initially translated into Persian without any change. Of the 90 words from the six original English word lists, 80 which were in accordance with desired linguistic criteria were retained from both lists A and B. Ten new words adapted to the Persian language were chosen for the development of the Persian lists. Finally, the word items on these newly formed Persian lists were compared with the original lists for consistency in terms of word length (as measured by number of phoneme counts in Persian), and all were one- or two-syllable concrete nouns. There were no obvious semantic or phonetic associations or similarities between the words on the same list and they were chosen from amongst frequently occurring words in the Persian language. The probability of the occurrence of the word in common usage in the Persian language was ascertained using the World Wide Web (WWW) as a database (from six billion contemporary Persian words on four million websites), Google as the search engine, and a Google-based applet, which provided word counts on the Web pages. This method was also employed by Cilibrasi (2007)[23] to automatically compute the similarity distance between words and phrases, and its performance was in agreement with that of the Word Net database. Recognition lists were constructed using target words and adding 20 new semantically associated or phonetically similar words as distracters. The above criteria helped establish form equivalence between the three new lists.

Procedure

All tests were administered by qualified examiners (psychology experts) trained in AVLT administration at quiet locations in accordance with Lezak's instructions.[1] The participants were screened at the beginning of the testing session for exclusion criteria such as health problems via a standardized interview before being categorized into three groups of 30 persons and matched based upon demographic variables that might affect performance on the auditory learning test such as age, gender, and education level.[6,24]

The first group was tested based upon the Persian adapted version of Rey's (word list 1) in two test sessions with intersession intervals of 30 days (±3); the second group received the Rey's 60 days (±5) after the initial testing; and

Table 1. Item characteristics of the original AVLT and new adapted Persian word lists.

	Original list 1 (Rey)		Adapted Persian version		Original list 2 (Lezak)		Adapted Persian version		Original list 3 (Shapiro & Harisson)		Adapted Persian list 3	
	A	B	A	B	A	B	A	B	A	B	A	B
	Drum	Desk	طبل	میز	Book	Bowel	کتاب	کاسه	Street	Baby	خیابان	بچه
	Curtain	Ranger	پرده	سرباز	Flower	Down	گل	سحر	Grass	Ocean	علف	اقیانوس
	Bell	Bird	زنگ	پرده	Train	Judge	قطار	قاضی	Door	Palace	درب	قصر
	Coffee	Shoe	چای	کفش	Rug	Grant	فرش	هدیه	Arm	Lip	دست	لب
	School	Stove	مدرسه	بخاری	Meadow	Insect	جنگل	حشره	Star	Bar	ستاره	میله
	Parent	Mountain	پدر	کوه	Harp	Plane	آواز	هواپیما	Wife	Dress	همسر	پیراهن
	Moon	Glasses	ماه	عینک	Salt	County	نمک	کشور	Window	Steam	پنجره	بخار
	Garden	Towel	باغ	حوله	Finger	Pool	انگشت	استخر	City	Coin	شهر	سکه
	Hat	Cloud	کلاه	ابر	Apple	Seed	سیب	دانه	Pupil	Rock	شاگرد	سنگ
	Farmer	Boat	دهقان	قایق	Chimney	Sheep	بخاری	گوسفند	Cabin	Army	اتاق	ارتش
	Turkey	Gum	گنجشک	دندان	Log	Coat	درخت	کت	Pipe	Friend	لوله	دوست
	Color	Pencil	رنگ	مداد	Key	Bottle	کلید	بطری	Skin	Storm	پوست	طوفان
	House	Church	خانه	مسجد	Rattle	Peach	شتر	ساحل	Fire	Village	آتش	روستا
	River	Fish	رود	تپه	Gold	Chair	طلا	صندلی	Clock	Cell	ساعت	زندان
	Nose	Lamb	بینی	بره	Button	Meal	دکمه	غذا	Lake	Building	دریاچه	ساختمان
Word Frequency	9 AA 4A	7AA 2 A	10 AA 5 A	13 AA 2 A	6 AA 3 A	4 AA 8 A	13 AA 2 A	11 AA 4 A	11 AA 4 A	10 AA 5 A	14 AA 1 A	13 AA 2 A
Word length (mean)	5.1	5.0	5.0	5.1	4.9	4.8	4.9	5.7	4.5	4.9	5.4	5.3
A: < 50 – 100 in 10 ^6 words; AA: >100 in 10 ^6 words												

the participants in the third group were evaluated in three test sessions via the Persian adapted version of the Rey's (word list 1), Lezak's (word list 2), and Shapiro & Harisson's (word list 3), respectively, with intersession intervals of 30 days (±3). All tests and retests were performed in a fixed sequence (non-counterbalanced design), and the delay intervals were chosen on the basis of previous literature.[25]

Description of the auditory verbal learning test and related measures

The test was administered in accordance with the procedure described by Lezak (2004).[1] In brief, subjects are asked to recall as many words on an initial oral 15-word list (list A) as possible, in any order. This procedure was repeated five times consecutively (trials 1 – 5 or acquisition trials), with the score for each trial being the number of words correctly recollected.

Participants were subsequently presented with a second list (list B); which was a measure of proactive interference, or the degree to which old learning can meddle with new learning. In the next step, participants were asked to remember as many words from list A as possible without reading it again; or immediate recall (ImmRec) and is believed to stimulate a retroactive interference situation in which new learning tangles with the recollection of old information. Delayed free recall (DelRec) was thereafter administered to the participants in the same manner as ImmRec (e.g., no access to list A) but following a 20 minute time interval. Subjects were then presented with a yes/no recognition trial of 50 words comprised of 15 target words from list A, 15 words from list B, and 20 words similar to word lists A and B in phonological or semantic terms.

The immediate memory span is usually measured based on the number of words recollected on the first trial (Trial 1). The learning curve is the variation in performance over the course of the five free-recall trials, with its slope providing a measure of the verbal learning test.[26] This slope is generally characterized as the difference between recall on the fifth trial and that on the first trial (Trial 5-Trial 1). Serial positions are obtained by adding up the number of words recalled across the five trials in blocks of three; in other words they can make a total amount of fifteen. Serial Position 1 (SP1) to Serial Position 5 (SP5) are utilized by studying the serial positions of the words recalled across the trials. Normal subjects are likely to recall the initial words (Primacy) but also the latest ones (Recency).[1]

The following is a list of other scores that are calculated: the total learning (TL=sum of words recalled in the five

Table 2. Descriptive statistics for the RAVLT scores and reliability coefficients in different test-retest administrations.

RAVLT scores	Interval:1 month (*n*=30)		Interval: 2 months (*n*=30)		Interval: 1 month (*n*=30)			r_{1-2}	r_{1-3}	r_{2-3}	ICC
	List 1	List 1	List 1	List 1	List 1	List 2	List 3				
Trial 1	7.03(2.43)	9.40(2.55)***	7.03(1.68)	9.20(1.94)***	6.80(1.85)	7.27(2.24)	7.23(2.14)	.69**	.55**	.56**	.59**
Trial 2	9.83(2.32)	11.73(2.20)***	9.33(2.15)	11.27(2.67)***	9.87(2.10)	9.47(2.25)	9.93(2.16)	.52**	.64**	.56**	.58**
Trial 3	11.23(2.30)	12.77(1.81)***	10.90(2.20)	12.50(1.60)***	11.27(1.82)	11.23(2.25)	11.80(1.94)	.48**	.56**	.61**	.54**
Trial 4	12.23(2.24)	13.00(1.58)*	11.80(1.89)	12.77(1.45)**	11.57(2.42)	11.77(2.33)	12.03(1.99)	.51**	.31	.59**	.47**
Trial 5	12.57(1.98)	13.23(1.55)*	12.03(2.24)	13.33(1.50)***	12.77(1.48)	11.93(2.08)	12.57(1.96)	.42*	.38*	.57**	.44**
Total	52.57(9.74)	58.07(13.03)**	51.07(8.00)	59.03(6.24)***	52.27(8.63)	51.63(9.54)	53.63(8.84)	.55**	.64**	.77**	.65**
List B	6.07(2.38)	6.50(2.62)	5.77(2.03)	6.73(2.27)	6.33(2.38)	6.20(2.04)	5.67(2.04)	.56**	.48**	.62**	.54**
Imm rec	11.73(2.27)	13.10(1.83)**	10.80(2.41)	11.90(2.29)**	11.23(2.58)	10.50(2.67)	11.63(2.34)	.64**	.51**	.60**	.56**
Del rec	11.27(2.57)	12.97(1.96)***	10.90(3.15)	11.70(2.53)*	11.30(2.73)	11.20(2.87)	11.43(2.45)	.66**	.72**	.71**	.70**
HIT	14.3(1.22)	14.50(0.86)	14.23(1.25)	14.37(0.96)	13.80(1.77)	13.77(1.60)	13.90(1.21)	.44*	47**	.345	.42**
MISS	0.87(1.22)	0.50(0.86)	0.77(1.25)	0.63(0.96)	1.20(1.77)	1.23(1.60)	1.10(1.21)	.44*	.47**	.345	.42**
CR	33.77(1.80)	34.27(1.20)	33.93(1.64)	34.13(1.07)	33.60(2.17)	33.90(2.17)	33.50(2.94)	.23	.22	.64**	.37**
FA	1.23(1.80)	0.73(1.20)	1.07(1.63)	0.87(1.07)	1.40(2.17)	1.10(2.17)	1.33(2.20)	.23	.31	.64**	.40**
Derived scores											
FAL	12.97(1.71)	13.53(1.46)*	12.40(2.39)	13.73(1.14)**	12.97(1.54)	12.60(1.85)	1.69))13.10	.37*	.35	.62**	.45**
Slop	5.93(1.88)	4.10(1.89)**	5.60(1.83)	4.53(1.72)	6.16(1.41)	5.37(1.92)	1.89))5.87	.55**	.02	.17	.19*
LOT	17.60(6.73)	12.87(6.26)*	15.57(7.74)	13.17(6.54)	18.27(5.04)	17.73(6.47)	5.57))17.13	.40*	.12	.28	.19*
FR	0.47(1.31)	0.27(0.87)	-0.03(2.20)	0.20(1.69)	-0.07(1.53)	-0.50(1.50)	1.05))0.30	-.13	.35	.06	.08
PI	-.97(2.50)	-2.90(1.90)**	-1.27(2.23)	-2.47(2.30)*	-6.43(2.25)	-5.73(1.94)	-6.90(2.06)	.18	.14	.33	.20*
RI	-.83(1.49)	-1.33(0.94)	-1.23(1.57)	-1.40(1.60)	-1.53(1.81)	-1.43(1.55)	-.93(1.51)	-.33	-.05	-.002	-.11
Age	31.63(8.48)		28.07(6.81)		31.23(5.20)						
Edu (yr)	10.50(6.21)		12.12(4.47)		12.00(3.47)						
Sex M/F%	50/50		50/50		46.7/53.3						

*P<0.05, **P<0.01, ***P<0.001; List 1:Rey's adapted Persian version; List 2: Lezak's adapted Persian version; List 3: Shapiro & Harisson's adapted Persian version; ImmRec: immediate recall; DelRec: delayed recall; CR: correctly rejected; FA: false accepted; FAL: final acquisition level; LOT: learning over trials; FR: forgetting rate; PI: proactive interference; RI: retroactive interference; ICC: interclass coefficient; RAVLT: Rey Auditory Verbal Learning Test.

free-recall trials), the learning over trials (LOT=TL-5* Trial 1), the proactive interference score (PIS=List B-Trial 1), the retroactive interference score (RIS=ImmRec-Trial 5), and the forgetting score (FR=DelRec-ImmRec). The recognition task generates the customary hits (Hit=number of words recognized from list A), misses (Miss= number of words not recognized from the list), correct rejection (CR=number of words not included on list A which were correctly identified as not being on the list), and false alarms (FA=numbers of words specified as being on list A which were not present).

Statistical analysis

With respect to the normal distribution of the scores, the effects of having administered the same list at different intervals were analyzed with the paired *t*-test. The practice effect and alternate forms equivalency were evaluated using the analysis of variance (ANOVA) for repeated measures. Taking place in pairs, comparisons were made via the Post Hoc Tukey test. The reliability of the tests was estimated by using Pearson's correlation coefficients across the three test sessions, and with interclass coefficients (ICC) using the absolute agreement definition. No values were excluded from the analysis.

Results

The new and original lists did not statistically differ with regards to any word item matching variables (Table 1). Table 2 demonstrates that administration of the same list after an interval of one or even two months yielded a significant practice effect on the word recall trials and some of the derived scores. The use of alternate forms after a one month delay, however, bore no significant difference across the forms. Figures 1a, 1b, and 2a, 2b show learning and serial positions curves utilizing the same list after one- or two-

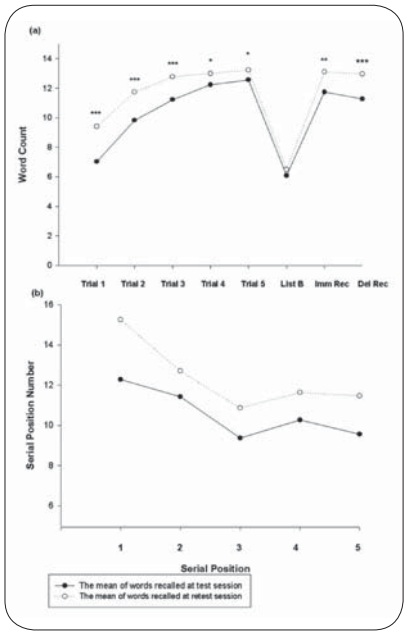

Figure 1. AVLT learning curves (a) and recalled words serial positions (b) of test-retest administration of the same word list (Rey) after a 30-day (±3) delay (*P<0.05, **P<0.01, ***P<0.001).

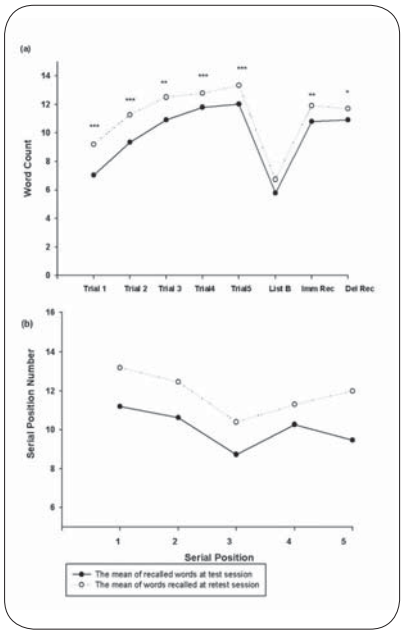

Figure 2. AVLT learning curves (a) and recalled words serial positions (b) of test-retest administration of the same word list (Rey) after a 60-day (±5) delay (*P<0.05, **P<0.01, ***P<0.001).

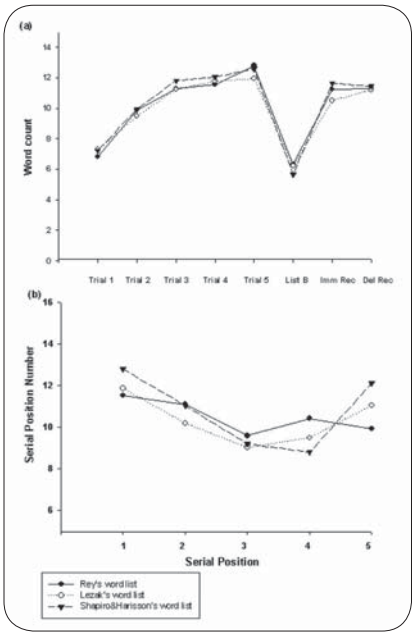

Figure 3. AVLT learning curves (a) and recalled words serial positions (b) in test-retest administration of alternate word lists (Rey, Lezak, and Shapiro & Harisson) at 30-day (±3) intersession intervals.

month intersession intervals. Figures 3a and 3b illustrate the equivalency of AVLT learning curves employing various alternate lists after a one-month intersession interval.

According to Table 2, the first and second trials, as well as the total, immediate, and delayed recall scores are the most reliable, whereas trial 3 and list B recall have modest reliability and the acquisition trials (trials 4 and 5) and recognition scores display poor ICCs.

Discussion

Our results showed that utilization of the same list resulted in the production of a significant practice effect in re-evaluation even after a two-month delay; repeated administration of alternate lists after a one-month intersession interval removed this undesirable effect. Insignificant difference between the mean scores of the trials across the three Persian adapted AVLT forms denotes equivalency and similarity in terms of difficulty level between these word lists.

In accordance with the results of the present study, previous direct comparisons of the two administration modes have indicated negligible improvement on alternate adaptations of verbal memory tasks, while a considerable practice effect is observed with the use of the same version.[25,27,28] Lezak also asserted that some patients were likely to exhibit an outstandingly good recall of word lists even after more than a year.[1] Shapiro and Harrison (1990) [15] recommended that subjects not be retested twice in a row within days of their initial evaluations even with alternative lists on account of the type of practice effect attributable to the meta-memoric influence, which may be the consequence of an earlier experience with the testing format. Our results are also in line with those reported by Crawford, et al.,[25] who found no practice effects when retesting normal subjects four weeks after the initial examination with an alternate word list. Furthermore, similar to previous studies,[1] our subjects showed normal primacy and recency patterns in serial position curves (Figures 1a, 2a, and 3a).

In the present study, Trial 1, Trial 2, Total, ImmRec, and DelRec scores showed the best reliability correlation (range: 0.56 to 0.70) whereas other acquisition trials were less reliable and recognition scores displayed poor ICCs. Test reliability across the forms was similar to that by Lemay, et al.,[20] who reported the best reliability for Total, ImmRec, and DelRec and modest reliability for acquisition trials and poor ICCs for recognition and derived scores. Some studies have yielded better correlation coefficients in retesting by alternate lists,[22,29] which could be partly attributed to different methods of design and analysis used by the researchers. We did not exclude any out of range scores, thus this difference is not unexpected. Duration intervals should also be regarded as one of the most important factors, which could influence the reliability coefficients; bearing in mind that exceeded correlations as reported by Ryan et al. (1986)[22] may be attributed to a relatively brief test-retest interval (140 min). Dikmen (1999)[30] suggested that poor reliability in memory tests might not necessarily be related to a measurement error but rather to the variable nature of memory. Low reliability

for verbal memory scores has also been observed for most tests used in clinical practice.[30–32]

To sum up, there is enough data to suggest that the difference between forms is sufficiently minor to be able to ignore it in clinical application. It seems that the Persian version of the RAVLT is a reliable instrument for repeated neuropsychological testing as long as alternate forms are used and scores chosen carefully. The results of the present study are useful for both clinical trials and clinical purposes such as pre- and post-assessments. Be that as it may, prior to general clinical application, normative data is required on larger samples of both clinical and normal subjects and for different age groups and education levels. Multiple studies with counter-balance design are needed to determine reliability, equivalency, and relative difficulty levels more definitively among different clinical populations.

Acknowledgment

This work was supported by a grant from Tehran University of Medical Sciences to Maryam Noroozian. We thank the staff of INCAS neurocognitive laboratory for their help in participant recruitment and management. We also thank Gholamreza Esmaeeli Djavid for his consultations for statistical analysis. We declare no conflicts of interest.

References

1. Lezak MD, Howieson DB, Loring DW. *Neuropsychological Assessment*. New York: Oxford University Press; 2004.
2. Bozikas VP, Kosmidis MH, Kiosseoglou G, Karavatos A. Neuropsychological profile of cognitively impaired patients with schizophrenia. *Compr Psychiatry*. 2006; **47:** 136 – 143.
3. Schoenberg MR, Dawson KA, Duff K, Patton D, Scott JG, Adams RL. Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch Clin Neuropsychol*. 2006; **21:** 693 – 703.
4. Mungas D. Differential clinical sensitivity of specific parameters of the Rey Auditory-Verbal Learning Test. *J Consult Clin Psychol.* 1983; **51:** 848 – 855.
5. Rosenberg SJ, Ryan JJ, Prifitera A. Rey Auditory-Verbal Learning Test performance of patients with and without memory impairment. *J Clin Psychol.* 1984; **40:** 785 – 787.
6. Messinis L, Tsakona I, Malefaki S, Papathanasopoulos P. Normative data and discriminant validity of Rey's Verbal Learning Test for the Greek adult population. *Arch Clin Neuropsychol.* 2007; **22:** 739 – 752.
7. Rey A. *L 'examen clinique en psychologie* [Clinical tests in psychology]. Paris: Presses Universitaires de France; 1964.
8. Lannoo E, Vingerhhoets G. Flemish normative data on common neuropsychological tests: influence of age, education, and gender. *Psychologica Belgica.* 1997; **37:** 141 – 155.
9. Miranda JP, Valencia RR. English and Spanish versions of a memory test: word length effects versus spoken duration effects. *Hisp Jl Behav Sci*. 1997; **19:** 171.
10. Vakil E, Blachstein H. Rey Auditory-Verbal Learning Test: structure analysis. *J Clin Psychol*. 1993; **49:** 883 – 890.
11. Muller H, Hasse-Sander I, Horn R, Helmstaedter C, Elger CE. Rey Auditory-Verbal Learning Test: structure of a modified German version. *J Clin Psychol.* 1997; **53:** 663 – 671.
12. Lee TMC. *Normative Data: Neuropsychological Measures for Hong Kong Chinese*. Hong Kong: Neuropsychology Laboratory, The University of Hong-Kong; 2003.
13. Messinis L, Tsakona I, Papathanasopoulos P. Rey Auditory Verbal Learning Test (RAVLT): Normative data of commonly used neuropsychological tests. Poster presented at the International Neuropsychological Society, 34th mid-year meeting. 2006.
14. Hawkins KA, Dean D, Pearlson GD. Alternative forms of the Rey Auditory Verbal Learning Test: a review. *Behav Neurol.* 2004; **15:** 99 – 107.
15. Shapiro DM, Harrison DW. Alternate forms of the AVLT: a procedure and test of form equivalency. *Arch Clin Neuropsychol.* 1990; **5:** 405 – 410.
16. Thorndike EL, Lorge I. *The Teacher's Word Book of 30,000 Words*. New York: Teacher's College, Columbia University Bureau of Publications; 1944.
17. Anderson JR, Bower GH. Recognition and retrieval processes in free recall. *Psychologica Rev.* 1972; **79:** 97 – 123.
18. Paivio A, Yuille JC, Madigan SA. Concreteness, imagery, and meaningfulness values for 925 nouns. *J Exp Psychol.* 1968; **76(suppl):** 1 – 25.
19. Hawkins KA, Sayward H. Examiner judgment and actual stability of psychiatric inpatient intelligence quotients. *Clin Neuropsychol*. 1994; **8:** 394 – 404.
20. Lemay S, Bedard MA, Rouleau I, Tremblay PL. Practice effect and test-retest reliability of intentional and executive tests in middle-aged to elderly subjects. *Clin Neuropsychol.* 2004; **18:** 284 – 302.
21. Wilson BA, Watson PC, Baddeley AD, Emslie H, Evans JJ. Improvement or simply practice? The effects of twenty repeated assessments on people with and without brain injury. *J Int Neuropsychol Soc*. 2000; **6:** 469 – 479.
22. Ryan JJ, Geisser ME, Randall DM, Georgemiller RJ. Alternate form reliability and equivalency of the Rey Auditory Verbal Learning Test. *J Clin Exp Neuropsychol*. 1986; **8:** 611 – 616.
23. Cilibrasi RL. The Google similarity distance. *IEEE Transactions on Knowledge and Data Engineering* (TKDE). 2007; **19:** 370 – 383.
24. Poreh A. Analysis of mean learning of normal participants on the Rey Auditory-Verbal Learning Test. *Psychol Assess*. 2005; **17:** 191 – 199.
25. Crawford JR, Stewart LE, Moore JW. Demonstration of savings on the AVLT and development of a parallel form. *J Clin Exp Neuropsychol*. 1989; **1:** 975 – 981.
26. Mitrushina M, Boone KB, D'Elia LF. *Handbook of Normative Data for Neuropsychological Assessment*. New York: Oxford University Press; 2006.
27. Benedict RH, Zgaljardic DJ. Practice effects during repeated administrations of memory tests with and without alternate forms. *J Clin Exp Neuropsychol*. 1998; **20:** 339 – 352.
28. Fastenau PS, Hankins WT, McGinnis CS, Moy T, Richard M. Effects of alternate forms on retest effects in clinical testing. *J Int Neuropsychol Soc.* 2002; **8:**151 – 161.
29. Delaney RC, Prevey ML, Cramer J, Mattson RH. Test-retest comparability and control subject data for the Rey-Auditory Verbal Learning Test and Rey-Osterrieth/Taylor Complex Figures. *Arch Clin Neuropsychol*. 1992; **7:** 523 – 528.
30. Dikmen SS, Heaton RK, Grant I, Temkin NR. Test-retest reliability and practice effects of expanded Halstead-Reitan Neuropsychological Test Battery. *J Int Neuropsychol Soc.* 1999; **5:** 346 – 356.
31. Mitrushina M, Satz P. Effect of repeated administration of a neuropsychological battery in the elderly. *J Clin Psychol*. 1991; **47:** 790 – 801.
32. Theisen ME, Rapport LJ, Axelrod BN, Brines DB. Effects of practice in repeated administrations of the Wechsler Memory Scale Revised in normal adults. *Assessment.* 1998; **5:** 85 – 92.