The Rey Auditory Verbal Learning Test: Normative data for the Arabic-speaking population and analysis of the differential influence of demographic variables

Amir Poreh[1,2], Alya Sultan[1] and Jennifer Levin[2]

1- Cleveland State University, Cleveland, OH, USA

2- Case Western Reserve University, Cleveland, OH, USA

Abstract

The Rey Auditory Verbal Learning Test (RAVLT) is the one of the most widely used neuropsychological tests of verbal memory. It has been translated into numerous languages including Spanish, German, Arabic, Hebrew, Czech, Portuguese, and English. The present study examined the hypothesis that the learning of word lists forms an algorithmic pattern across all cultures. To this end, a sample of 200 Arabic-speaking Omani adults between the ages of 18 and 50 years was collected. The resulting norms were then compared with existing American and Brazilian samples. The study confirmed that the first trial on the RAVLT correlates with demographic variables, whereas the learning slope on subsequent trials is almost identical across all cultures. Based on the above finding, the slope of the verbal learning test is hypothesized to measure a psychophysiological process linked with the hippocampal formation and allows for the laying down of new memories. In contrast, the first trial of the test is amenable to more cultural, demographic, and environmental factors. **Keywords:** memory; neuropsychological tests; neuropsychology.

Introduction

The Rey Auditory Verbal Learning Test (RAVLT) is one of the first commercially used measures of verbal memory. It was originally published in 1907 by Edouard Claparède and later translated to German, Italian, and Spanish (1919; see Boake, 2000). In 1958, Andre Rey, Claparède's doctoral student, expanded the test and had patients repeatedly listen to and recall the same list of words five times. Following the presentation of a new list, Rey had his patient recall the previously presented words (Rey, 1958). That same year, Edith Taylor also added a 30-min delay trial and recognition trial (Taylor, 1959). Today, the RAVLT is considered the most widely used measure of verbal memory in the psychological literature (Mitrushina, Boone, Rzani, & D'Elia, 2005; Strauss, Sherman, & Spreen, 2006).

To properly evaluate the performance of an examinee on this test, one usually calculates the deviation of the participant's recall in each session of the test from the mean recall in that session for normal participants with similar demographic characteristics. The means and standard deviations for different age groups are typically presented in a tabular form in numerous normative datasets (e.g., Lezak, Howieson, & Loring, 2004; Mitrushina et al., 2005; Strauss et al., 2006). Ideally, normative data should be based on tests of groups with a large number of participants matched on demographic variables that may affect performance on auditory learning tests such as age, gender, and intelligence quotient. Additionally, to be useful, the clinician must consider cultural differences as well as having adequate normative data with regard to the performance on such tests in the target population.

The role of cultural differences in neuropsychological test performance is further complicated by the influence of education on one's performance on such tests. For example, Manly, Schupf, Tang, & Stern (2005) noted that literacy level and not years of education appeared to be an important mediator of the interactions between biological and environmental factors. Thus, most recent normative studies, particularly those that used regression-based norms, use education level as a mediating factor when calculating within-subject test performance.

A review of the literature shows that only a few studies attempted to translate neuropsychological measures into Arabic. Stanczak, Stanczak, & Awadalla (2001) translated and adopted the Expanded Trail

Amir Poreh and Alya Sultan, Neuropsychology Laboratory, Department of Psychology, Cleveland State University. Amir Poreh and Jennifer Levin, Department of Psychiatry, Case Western Reserve School of Medicine. Correspondence regarding this article should be directed to: Amir Poreh, Department of Psychology, Cleveland State University, 2121 Euclid Ave. CB 178, Cleveland, OH 44115. Phone: 216-687-3718, Fax: 216 916-3704. E-mail: a.poreh@csuohio.edu.

Making Test to Arabic and examined the performance of a large sample of Sudanese subjects on this measure. The results showed that Americans exhibited markedly lower completion times than their Sudanese counterparts. Stanczaket al. (2001) attributed these differences to age, education, and sociocultural factors. Al-Ghatani, Obonsawin, & Al Moutaery (2009) and more recently Khalil (2010) examined the performance of healthy Arabic-speaking adults in Saudi Arabia on verbal fluency tests. Both studies used the Arabic version of the semantic (i.e., naming as many animals as possible) and phonemic (i.e., reciting as many words that start with a certain sound) versions of this test. In both studies, the average number of words generated in 60 sec was lower compared with English-speaking adults. To our knowledge, no study has yet attempted to develop normative data for the RAVLT for Arabic-speaking adults.

When developing normative data for tests such as the RAVLT, distinguishing between the various components of the test is important. Many studies have shown that performance on the first presentation of the word list may be conceptualized as a measure of working memory or retention, and the slope that produces the five learning trials should be regarded as a measure of verbal learning.To clarify the strategies that underlie word list learning, Dunlosky & Salthouse (1996) decomposed the five-trial word list learning using lost and gained access indices. *Lost access* refers to the number of words recalled in a given trial that were not recalled during the immediately subsequent trial. *Gained access* reflects the ability to profit from the repeated presentation of the word list and refers to the number of new words recalled in a given trial that were not recalled in the immediately preceding trial. Dunlosky and Salthouse and other investigators showed that patients who suffer from dementia have low gained access and high lost access across trials.

Poreh (2005) decomposed the learning trials using the best-fit linear logarithmic equation method. Using multiple North American normative samples (meta norms), Poreh demonstrated that the learning curve $R^{(t)}$ can be expressed by Equation 1:

$$M(t) = R^{(t)} = R^1 + S \, ln \, (t) \qquad \text{Equation (1)}$$

R^1 (trial 1) is the performance of a subject on the first trial. The coefficient S describes the mean rate of learning and henceforth is referred to as the logarithmic learning slope, learning slope, or slope.

To test the above model, the normative data of North American English-speaking subjects were examined across the 58 published normative samples. The study showed an extremely robust curve fitting (mean r^2 = .985, SD = .018). Additional analyses also showed that the immediate memory span on the first trial (R^1) varies considerably with age and is affected by other background factors, whereas the mean learning slope (S) has a natural logarithmic curve (*ln*) and is practically

independent of the mean age and other demographic factors. A weighted, linear, stepwise, multiple regression analysis of all of the adult data yielded Equation 2:

$$M(t) = R^{(t)} = 4.015 - 0.0649 \, Age - 0.0126 \, p.m. + 0.30 \, Edu + 0.0224 \, IQ + S \, ln \, (t) \qquad \text{Equation (2)}$$

Poreh (2005) suggested that one may be able to use the values of R^1 and S to evaluate patient performance instead of using their recall in each of the trials. Poreh also argued that the logarithmic learning curve reflects a fundamental psychophysiological process and would therefore emerge in every sample, regardless of the subject's age, education, or gender.

Support for Poreh's (2005) distinction between the initial recall and learning of word lists was also separately reported by Jones et al. (2005). Using a single sample of North American older adults, they showed that the first trial on the RAVLT correlated with demographic variables. Additionally, they also demonstrated that the algorithmic curve fitting model provides the best-fit model of verbal list learning. However, Jones et al. argued that the rate of learning was significantly associated with ethnicity, in addition to general cognitive status measured by the Mini-Mental State Examination.

The first aim of the present study was to publish norms for the RAVLT for Arabic-speaking adults. To this end, a large sample of Omani adults was administered the Arabic version of the RAVLT (A-RAVLT). The second goal of the study was to assess the hypothesis that the logarithmic learning curve on the RAVLT transcends ethnic groups and cultures. Much like their English-speaking counterparts, learning of the A-RAVLT list among Omani adults was predicted to follow the same natural logarithmic curve pattern, whereas the first trial (R1) would be affected by demographic factors. Additionally, for cross-cultural comparison, the study examined curve fitting of the RAVLT learning curve using recently published RAVLT Brazilian normative data (Magalhaes & Hamdan, 2010).

Methods

Participants

Two hundred adults were randomly selected from a local university and the surrounding community. A small sample was collected in a small village located somewhat inland of Oman where inhabitants are less exposed to modern standardized testing. The age range of the subjects was 18 to 49 years (*M* = 25.46 years, *SD* = 6.2 years). The education level of the subjects ranged from 0 to 18 years (*M* = 10.48 years, *SD* = 21.29 years).The majority of the subjects were female (66%) and right-handed (93.5%). Educational distribution was bimodal with 29.5% of the population reporting 11 years of education. Hence, resulting data were subdivided into six subgroups. Table 1 shows the general characteristics of the groups. One sees the phenomenon of hand-shifting

Table 1. Demographic characteristics of the four age groups in the sample

Age group	Educational group	Average age (years)	Female (%)	Right-handed (%)	Average education (years)
18–20 years	Low (n = 21)	20.86 (SD = 4.29)	61.9	90.5	5.71 (SD = 4.41)
	High (n = 28)	19.02 (SD = 1.05)	71.4	96.4	18.86 (SD = 1.17)
21–30 years	Low (n = 42)	24.95 (SD = 1.96)	57.1	95.2	6.00 (SD = 2.41)
	High (n = 76)	24.59 (SD = 1.97)	59.2	95.2	13.82 (SD = 1.81)
31–49 years	Low (n = 18)	38.11 (SD = 4.49)	94.4	100	6.06 (SD = 3.11)
	High (n = 15)	34.67 (SD = 3.96)	86.7	100	13.60 (SD = 2.19)

SD, standard deviation.

in the older age group, which has been recorded in most other cultures (Porac & Friesen, 2000).

Procedure

The A-RAVLT was translated by a professional Arabic-speaking translator and then reverse translated and adapted to the Omani dialect by the second author. The words selected for the final version of the test were almost identical to the original French version of the test. Before administering the tests to the subjects, three examiners were trained to administer the test using A-RAVLT software (Poreh, 2005). After the examiners demonstrated proficiency in the administration, they proceeded to recruit subjects. Each subject read and signed a consent form prior to the test administration. The consent form was approved by the Institutional Review and Ethics Board at Midwestern University and then translated to Arabic. The study was approved by Oman's Ministry of Health. Subjects who had a history of psychiatric illness, head injury, and history of psychoactive drug use were excluded from the study. After completing a demographics form, the examiner opened the computer laptop and set it up so the examinee could not see the screen. The computer screen provided the examiner with the instructions for administering the test and inserted a time stamp when the software was administered. In this fashion, standardization and quality control of the test administration were maintained.

Statistical analysis

Initial statistical analysis was performed using SPSS version 16 software. The best-fit linear logarithmic equation for the raw mean recall data in each age group was initially determined using the Trend function in Microsoft Excel without any preprocessing or preconditioning. This analysis was then repeated using the SPSS curve fitting module. In the original article, to correct for a possible ceiling effect produced by participants who recalled the entire list of 15 words by the fourth trial, the researcher fitted the logarithmic equation to the mean recalls of the first four trials. However, this procedure was not used in the present study because none of the samples achieved the ceiling (15/15 words) during the first four trials. Multiple paired t-tests

using the mean T scores derived from Equation 2 of the Omani sample results were computed using GraphPad InStat software.

Results

Verbal learning and memory

The demographic background of the Arabic-speaking Omani sample is summarized in Table 1. Multiple analysis of variance (MANOVA) revealed only a main effect of gender on A-RAVLT performance in the first five trials (F[31,94] = 1.760, p = .02). A two-way age × education interaction was also found (F[31,94] = 1.590, p = .046). Fisher's Least Significant Difference *post hoc* test showed that the group of younger participants with higher education performed significantly better than the group of older and less-educated participants on the A-RAVLT (F[3,1] = 4.694, p = .03). As in the study by Poreh (2005), performance in Trial 1 correlated significantly with age (r = -.58, p < .001) and to a lesser degree with education (r = .26, p < .01).

To examine the differences between North American and Omani subjects, the T score for the slope based on Equation 2 was examined using the normative data adapted from the North American samples (Table 2). Notice that the level of intellectual functioning was not available and was therefore estimated as being in the average range. The resulting repeated two-tailed t-tests with the American sample (T = 50, SD = 10) showed a single significant difference for the Omani 21- to 30-year-old low-education group (t = 2.16, p = .032). This difference became statistically nonsignificant after correcting for the multiple comparisons using the Bonferroni method.

To examine Poreh's best-fit linear logarithmic equation for the raw mean data, recall data in each age group were examined using the trend function procedure in SPSS version 11. Table 3 shows that r^2 ranged from .87 to .99, with the low education groups producing a curve fit that was similar to the high education group. As previously mentioned; for comparison purposes we also included in the present study the mean scores for the RAVLT recently published by Magalhaes & Hamdan (2010) for the Brazilian population. Table 3 shows that the r^2 for the Brazilian sample ranged from

Table 2. Normative data for the Arabic version of the RAVLT

Age group	Educational group	I	II	III	IV	V	PI	30-min delay	Recognition	Mean slope T-score	T score[‡]
18–20 years	Low (n = 21)	8.08 (2.51)	11.50 (2.01)	12.04 (1.89)	11.80 (1.96)	12.80 (1.70)	10.96 (3.90)	10.96 (3.92)	13.40 (4.01)	45.26 (9.80)	$t = 2.16$, $p = .032$
	High (n = 28)	7.32 (1.76)	11.47 (1.81)	12.4 (2.56)	12.47 (1.96)	12.89 (1.47)	11.00 (3.76)	12.53 (2.20)	13.42 (1.10)	49.37 (9.99)	$t = .33$, $p = .741$
21–30 years	Low (n = 42)	7.39 (2.83)	9.39 (2.52)	11.15 (2.19)	11.68 (2.11)	12.36 (1.97)	10.60 (3.12)	10.76 (2.89)	14.2 (2.3)	47.86 (10.39)	$t = 1.35$, $p = .175$
	High (n = 76)	6.61 (1.92)	9.95 (2.10)	11.37 (2.21)	12.34 (1.89)	12.60 (1.67)	11.12 (2.87)	11.16 (2.87)	13.6 (2.8)	52.20 (8.93)	$t = 1.86$, $p = 0.062$
31–49 years	Low (n = 18)	7.35 (2.39)	9.58 (2.24)	10.69 (2.53)	11.88 (2.44)	12.34 (2.29)	10.73 (3.77)	10.77 (2.99)	14.4 (.69)	48.25 (11.68)	$t = 0.639$, $p = 0.531$
	High (n = 15)	6.66 (1.23)	10.00 (2.28)	12.22 (2.39)	12.81 (2.54)	13.21 (2.18)	10.10 (5.15)	9.20 (5.36)	14.1 (1.1)	55.13 (11.54)	$t = 1.71$, $p = .109$

[‡]Comparison with the U.S. standardized norms generated using the logarithmic equation with corrections for age, education, and level of intelligence.

Table 3. Logarithmic analyses of the learning curve data for the Oman and Brazilian mean scores

Omani normative data			Brazilian normative data		
Age group	Educational group	r^2	Age group	Educational group	r^2
18–20 years	Low (n = 21)	.91	17–34 years	≤11 years	.98
	High (n = 28)	.89		≥12 years	.93
21–30 years	Low (n = 42)	.99	34–49 years	≤11 years	.99
	High (n = 76)	.98		≥12 years	.99
31–49 years	Low (n = 18)	.99	50–56 years	≤11 years	.99
	High (n = 15)	.98		≥12 years	.98
			65–85 years	≤11 years	.99
				≥12 years	.98

.93 to .99, with the low education groups producing a curve fit that was similar to the high education group. A two-way ANOVA conducted by Magalhaes & Hamdan (2010) showed that performance in the first trial was significantly affected by age ($F[3,294] = 6.173$, $p < .001$) and education ($F[1,294] = 9.611$, $p < .01$), with no age × education interaction. An ANOVA that compared the mean scores of the Omani sample with the mean scores of the Brazilian sample published by Magalhaes & Hamdan (2010) showed that the two groups did not differ in Trial 1 for the very young ($F = .1.73$, $df = 3$, $p = .323$)and young ($F = .99$, $df = 3$, $p = 0.39$) groups but did differ for the middle-aged groups ($F = 5.01$, $df = 3$, $p < .002$). This finding suggests that performance in the first trial does not significantly differ across cultures when one controls for age and education.

Discussion

Aside from providing preliminary normative data for Omani Arabic-speaking adults aged 18 to 50 years,

the present study is the first reported in the literature to show that the initial trial of word list learning is strongly affected by various demographic and cultural variables, whereas other features, particularly the learning slope, are relatively immune to such effects and produce a logarithmic-shaped curve across cultural and educational groups.

One question is why the learning slope is relatively immune to environmental and demographic variables. One of the possibilities is that the learning curve reflects innate human neurocognitive function. The logarithmic characteristic of the RAVLT learning slope, which conforms to Weber's law, appears to reflect a psychophysical phenomenon much like the perception of brightness or pure tone. In other words, the learning curve appears to be a putative marker for an innate neurological substrate that controls one's ability to lay down new memories. The area in the brain that is typically associated with the formation of new memories is the hippocampal formation (Squire & Schacter, 2002).

Studies have repeatedly shown that direct or indirect damage to this area, particularly in the left hemisphere, produces severe verbal declarative memory deficits (Di Gennaro et al., 2006; Poreh et al., 2006).

Moscovitch (2008) proposed that the hippocampal formation is a "dumb" module. Following Fodor's (1983) conceptualization of modular cognitive functions, Moscovitch claimed that hippocampal functions are encapsulated and automatic. Specifically, they are unaffected by other cognitive processing including one's educational level or age unless the hippocampus suffers a direct insult. Other parts of the memory system according to Moscovitch (2008), such as working memory and executive function, are hierarchical, nonmodular, effortful, and therefore nonencapsulated. Performance of these portions of the memory system would be highly affected by education and age. Indeed, the present study confirmed this finding, and the first trial on the RAVLT fits this description.

What are the implications for the universality of the learning curve to the cross-cultural assessment of neurological disease? Given that performance in the first trial of the RAVLT is highly correlated with background variables that have often been associated with what researchers have termed *cognitive reserve capacity* (Stern, 2002, 2003), use of this portion of the test to identify neurological disorders is highly suspect. Clinicians, particularly those who work in culturally diverse settings, should place greater emphasis on the analysis of the RAVLT slope. The first trial of the test, on the other hand, can be used to determine the extent of the subject's cognitive reserve capacity and to some extent determine whether his or her presentation may be accounted for by various demographic variables. Given this conclusion, both poorly and highly educated patients with Alzheimer's disease are predicted to show a relatively flat learning curve, with the last portion of the list (i.e., recency effect) being higher for the group with the greater "cognitive reserve capacity." This recency effect will also remain consistent across all five trials.

Several limitations of the present study should be noted. First, it compared only the mean scores of the two ethnic/cultural groups. Second, our interpretation regarding the link between the hippocampal formation and learning slope of the RAVLT is tenuous. Ultimately, to test this hypothesis, conducting a study in which normal adults are instructed to learn a list of words while they undergo functional imaging would be desirable. As the individuals learn the list, activation of the hippocampal formation is predicted to show the same logarithmic pattern in which the activation will progressively increase until the system reaches an asymptote.

In summary, the present study provides new insights into the complex role of cultural and educational factors in performance on neuropsychological measures. The study shows that some aspects of certain tests require the use of solid normative data, whereas other portions of the same measure assess more innate (i.e., modular) neurological processes and are more robust, thus, not requiring much adaptation.

References

Al-Ghatani, A. M., Obonsawin, M., & Al Moutaery, K. R. (2009). Normative data for the two equivalent forms of the Arabic verbal fluency test. *Pan Arab Journal of Neurosurgery, 13*(2), 57-65.

Boake, C. (2000). Edouard Claparède and the origin of the auditory verbal learningtest. *Journal of Clinical and Experimental Neuropsychology, 22,* 286-292.

Di Gennaro, G., Grammaldo, L. G., Quarato, P. P., Esposito, V., Mascia, A., Sparano, A., Meldolesi, G. N., & Picardi, A. (2006). Severe amnesia following bilateral medial temporal lobe damage occurring on two distinct occasions. *Neurological Science, 27*, 129-133.

Dunlosky, J., & Salthouse, T. A. (1996). A decomposition of age-related differences in multitrial free recall. *Aging, Neuropsychology, and Cognition, 3*, 2-14.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology.* Cambridge, M. A: MIT Press.

Jones, R. N., Rosenberg, A. L., Morris, J. N., Allaire, J. C., McCoy, K. J. M., Marsiske, M., Kleinman, K. P., Rebok, G. W., & Malloy, P. F. (2005). A growth curve model of learning acquisition among cognitively normal older adults. *Experimental Aging Research, 31*, 291-312.

Khalil, M. S. (2010). Preliminary Arabic normative data of neuropsychological tests: The verbal and design fluency. *Journal of Clinical and Experimental Neuropsychology, 32*, 1028-1035.

Lezak, M., Howieson, D. B., & Loring D. W. (2004). *Neuropsychological assessment*, 4th edition. Oxford: Oxford University Press.

Magalhaes, S., & Hamdan, A. C. (2010). The Rey Auditory Verbal Learning Test: Normative data for the Brazilian population and analysis of the influence of demographic variables. *Psychology & Neuroscience, 3*, 85-91.

Manly, J. J., Schupf, N., Tang, M. X., & Stern, Y. (2005). Cognitive decline and literacy among ethnically diverse elders. *Journal of Geriatric Psychiatry and Neurology, 18*, 213-217

Mitrushina, M., Boone, K. B., Rzani, J., & D'Elia, L. F. (2005). *Handbook of normative data for neuropsychological assessment,* 2nd edition. New York: Oxford University Press.

Moscovitch, M. (2008). The hippocampus as a "stupid," domain-specific module: Implications for theories of recent and remote memory, and of imagination. *Canadian Journal of Experimental Psychology, 62,* 62-79.

Porac, C., & Friesen, I. C. (2000). Hand preference side and its relation to hand preference switch history among old and oldest-old adults. *Developmental Neuropsychology, 17*, 225-239.

Poreh, A. (2005). Analysis of mean learning of normal subjects on the Rey Auditory-Verbal Learning Test. *Psychological Assessment, 17*, 191-199.

Poreh, A., Winocur, G., Moscovitch, M., Backon, M., Goshen, E., Ram, Z., & Feldman, Z. (2006). Anterograde and retrograde amnesia in a person with bilateral fornix lesions following removal of a colloid cyst. *Neuropsychologia, 44*, 2241-2248.

Rey, A. (1958). *L'examenclinique en psychologie* [The psychological examination]. Paris: Presses Universitaires de France.

Squire, L. R., & Schacter, D. L. (2002). *The neuropsychology of memory.* New York: Guilford Press.

Stanczak, D. E., Stanczak, E. M., & Awadalla, A. W. (2001). Development and initial validation of an Arabic version of the Expanded Trail Making Test: Implications for cross-cultural assessment. *Archives of Clinical Neuropsychology, 16*, 141-149.

Strauss, E., Sherman, E. M. S, & Spreen, O. (2006). *A compendium of neuropsychological tests:Administration, norms and commentary*, 3rd edition. New York: Oxford University Press.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society, 8*, 448-460.

Stern, Y. (2003). The concept of cognitive reserve: A catalyst for research. *Journal of Clinical and Experimental Neuropsychology, 25*, 589-593.

Taylor, E. M. (1959). *Psychological appraisal of children with cerebral defects.* Cambridge, M. A.: Harvard University Press.