1988

Rey Auditory-Verbal Learning Test: Development of Norms for Healthy Young Adults

Arthur N. Wiens

Mark R. McMinn

John R. Crossen

Rey Auditory-Verbal Learning Test: Development of Norms for Healthy Young Adults

Arthur N. Wiens
Oregon Health Sciences University

Mark R. McMinn
George Fox College

John R. Crossen
Oregon Health Sciences University

ABSTRACT

The Rey Auditory-Verbal Learning Test (AVLT) has been widely used in clinical neuropsychology because of the usefulness of its multiple measures of learning and memory and its ease of administration. Normative data has been reported for some patient populations but little normative data exists for healthy individuals. This study reports AVLT data for 222 job applicants, presently employed in a variety of occupations, who had previously passed basic-academic-skills tests and physical examinations and were motivated to perform well on AVLT testing. AVLT normative data are presented by WAIS-R FSIQ, Age, Education, and by Age and FSIQ combined. We also present a proposal for sharing and pooling data to expand the data matrix we present.

The Rey Auditory-Verbal Learning Test (AVLT), developed by the Swiss psychologist Andre Rey (1941, 1964) and described in English by Taylor (1959) and Lezak (1976, 1983), has been widely used in clinical neuropsychology because of the usefulness of its multiple measures of learning and memory and its ease of administration. Originally included in a battery of tests derived from experimental psychology, the AVLT word list was initially used to evaluate contradictory memory performance that raised a question of functional vs. organic problems, e.g., when memory on an easier recognition task was inferior to memory on a more difficult recall task (Rey, 1941). Later, Rey (1964) modified

---

Address for reprint requests: Arthur N. Wiens, Ph.D., Department of Medical Psychology, OP336, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Rd., Portland, OR 97201-3098, USA.

the task to assess verbal learning and memory by requiring free-recall on five successive trials and by incorporating a written recognition task.

The AVLT consists of 15 words spoken by the examiner on five successive trials, with a free-recall task following each successive trial. In a contemporary adaptation of Rey's original procedure, a second group of words (distraction list) is read aloud with free-recall. After the distraction list recall, the subject is asked to recall as many of the original 15 words as possible (postdistraction recall) (Lezak, 1976, 1983). Written recognition tasks have also been employed to assess memory for the original words in a story format (Lezak, 1976, 1983; Rey, 1964) or in a matrix array (Lezak, 1983).

The AVLT has been used in comparisons of numerous clinical groups of patients with neurologic and psychiatric disorders. Different authors report that the AVLT is sensitive to the following differences: lateralized brain damage (Lezak, 1979; Miceli, Caltagirone, Gainotti, Masullo, & Silveri, 1981), memory-impaired vs. non-memory-impaired neurologic patients (Rosenberg, Ryan, & Prifitera, 1984), differences between (a) amnestics and head trauma patients vs. (b) schizophrenics, nonpsychotic psychiatric patients, and attention deficit patients (Mungas, 1983), depressed vs. alcoholic patients (Query & Megran, 1984), types of errors for depessed vs. medical patients (Chiulli, Haaland, Ellis, & Rhodes, 1985), bipolar patients with and without tardive dyskinesia (Wolf, Ryan, & Mosnaim, 1983), and younger vs. older normals (Cohen, Andres, & Smolen, 1986; Montgomery & Costa, 1983; Rey, 1964).

The AVLT has also been found by Query and Berger (1980) to be sensitive to certain age differences for acute head trauma/stroke patients, and to education and IQ. In a factor-analytic study with a diverse neurologic/psychiatric sample the AVLT was found to have a verbal learning and memory loading with the verbal subtests of the Wechsler Memory Scale (Ryan, Rosenberg, & Miltenberg, 1984). Two studies have reported changes in AVLT performance following pharmacological interventions in patients with dementia (Delwaide, Devoitille, & Ylieff, 1980; Miceli, Caltagirone, & Gainotti, 1977). Mixed results regarding organizational memory process in several diagnostic samples have been obtained (Chiulli et al., 1985; Mungas, 1983).

In spite of the clinical usefulness of the AVLT suggested by the above studies, no reasearch to date has yet established empirical guidelines for AVLT performance in a large sample of healthy, well-motivated individuals.

Methodological Considerations

Several procedural variations in the administration of the AVLT illustrate the test's flexibility, but have also lead to methodological problems when comparing results across studies. Most surprisingly, the list of words adapted from Rey's investigations contains three significant changes from his original French words: "bell" instead of "belt" for "ceinture", "moon" instead of "sun" for "soleil", and "nose" instead of "mustache" for "moustache" (Taylor, 1959). The effect of these changes has not been empirically investigated, but laboratory experiments

have shown that word imagery, concreteness, and meaningfulness influence verbal memory (Paivio, 1971). In addition, the potential effects of different word lists, educational and cultural differences between Rey's French subjects of three decades ago and other recently evaluated samples remain undefined. Until the potential effects of these factors can be ascertained, clinicians and researchers would be advised to use caution in applying Rey's norms (reported in Lezak, 1976, 1983) to other individuals who may not share his sample's characteristics.

Both Rey (1964) and Taylor (1959) recommend that the words be separated by 1-s intervals. Lezak's (1976, 1983) instructions defined a word presentation rate of one per second, which appears to be a more rapid presentation rate. The latter procedure could be expected to result in somewhat lower performance because less rehearsal time is available between words when they are presented at a faster pace. Further empirical research is necessary to clarify the possible effects of such differences in presentation rate.

There have been numerous variations in administration of the recognition trial. Rey (1964) described oral reading of a story with instructions for the subject to stop the examiner when a word was recognized. In the original version of the AVLT (Rey, 1964), the recognition paragraph contained twice as many distractor nouns and is unrecognizable compared to the one that has been used in subsequent studies (Lezak, 1976, 1983). In a subsequent revision (Lezak, 1976), the recognition trial has involved presentation of a written story quite different from Rey's, rather than an auditory recognition task. Although not all AVLT studies have employed a written recognition trial, most that have done so have adopted a brief written story containing the 15 English words. Other investigators (Chiulli et al., 1985; Lezak, 1983) have used a written matrix array of words containing the AVLT originals, phonemic foils, and semantic foils.

Alternate word lists for repeated AVLT administrations have been described (Rey, 1964; Lezak, 1983). The equivalence of these lists has been empirically demonstrated by Ryan, Geisser, Randall, and Georgemiller (1986) who reported alternate form reliabilities of .60 to .77.

Rey (1964) did not include a distraction trial prior to incidental free recall of the original list. This distraction trial variation was first developed by Taylor (1959) using a list of words Rey (1964) attributed to Claparède. Again, there are some changes from the original French words: "ranger" instead of "shepherd" for "berger," "towel" instead of "sponge" for "éponge," and "cloud" instead of "picture" for "image." According to Rey's (1964) and Taylor's (1959) instructions, the examiner was to give feedback each time the subject repeated a word within a trial. In Rey's version, the individuals were also given feedback on the number of words recalled at the end of each recall trial and on the fifth recall trial the examiner announced a last trial. These instructions have apparently not been part of the standard administration procedure in subsequent studies. Lezak has advised the examiner not to volunteer information about whether a word has been repeated unless asked by the subject, because of the potential for distraction. The changes in content and procedure would make comparisons

using Rey's (1964) norms reported in Lezak (1976, 1983) quite difficult to interpret.

Lezak (1982) investigated repeated AVLT administrations with a group of normal subjects at 6- and 12- month intervals. Statistically significant increases in the number of words recalled were noted for Trials I, V, and postdistraction recall. However, the mean magnitude of these improvements was small, attaining a maximum improvement of only 1.36 words on any trial. Thus, initial research indicates that small, but statistically significant improvements (i.e., practice effects) can be expected on successive serial AVLT administrations.

Performance Measures of the AVLT

The AVLT has yielded numerous parameters of interest for assessment of memory performance. In addition to recall on each of the five trials individually, total five-trial recall, postdistraction recall, and written recognition memory, several other variables have been investigated. Rey (1964) reported the number of correct responses, the number of false responses, the number of repetitions, and the number of repetitions where the subject questioned whether the word had already been recalled. He also recommended recording the number of words remembered in each 15-s block of time during the 60-s recall period. Chiulli et al. (1985) are the only other investigators to evaluate the nature of AVLT error performance. They found that depressed individuals are more likely to make phonemic recognition errors than are normals.

Lezak (1979, 1983) evaluated the number of words recalled on the postdistraction trial. She found that graduate students recalled an average of 1.52 fewer words than on Trial V and patients recalled 1.97 fewer words (ns). However, a significantly larger proportion of the mixed brain-damaged patients showed reduced recall of three words or more. Mungas (1983) investigated the amount of loss from Trial V to the postdistraction recall trial and also found it to be a useful index. In the same study, a measure of subjective organization was computed, using the number of trials on which a given word was recalled to assess consistency of memory sequences.

Query and Berger (1980) defined their measure of learning as the increase in the number of words recalled on Trial V over the number of words recalled on the first trial. Query and Megran (1983) later defined learning as the highest number of words recalled on any trial minus the recall score on Trial I. In a third study, data on words learned are reported but the method of computation was not clarified (Query & Megran, 1984).

Additional research is necessary to identify which parameters of AVLT performance are most useful for the assessment of memory in specific clinical populations. Among the measures that might be investigated are: (a) recall on individual AVLT trials, (b) total words recalled in five trials, (c) difference between recall on Trial I and Trial V (or maximum recall), (d) subjective organization, and (e) type of error.

Need for Normative AVLT Data

Very few studies of the AVLT have included a control group of normal individuals as comparisons for patients with clinical conditions. Some studies have used medical patients in control groups, but only Lezak (1982) and Cohen et al. (1986) have reported performance of nonmedical control subjects on the AVLT. Unfortunately, the sample size in the Lezak study was too small to establish solid normative parameters and the same was the case in the Cohen et al. study for male subjects.

Appropriate normative data for the AVLT based on large samples of healthy individuals have not yet been reported. The norms for Europeans reported in Lezak (1976, 1983) may be inadequate for at least three reasons previously discussed: (a) the English translations for 20% of the words are different from the original words; (b) the current administration is different from the original Rey (1964) procedures, in which errors and repetitions were identified as they occurred, feedback was provided on each trial about the correct number of words recalled, no distraction trial was administered, and a different recognition paragraph was used, and (c) educational and cultural differences may further invalidate comparison of current American samples to those collected by Rey 20-30 years ago.

Although the AVLT has been demonstrated to be sensitive to organic brain impairment in a variety of populations, the availability of norms based on a large sample of healthy adults would be advantageous to future clinical and research investigations of cognitive impairments associated with specific diagnostic conditions. The largest sample size ($N = 677$) reported to date on adult norms for the AVLT comes from a study of medically ill male inpatients at a VA hospital (Query & Megran, 1983). Those norms may be inappropriate for healthy individuals because of the unspecified medical conditions of the VA patients.

The present study presents the first normative AVLT data collected from a large sample of healthy young American job applicants with optimal motivation for maximum test performance. In addition to age, the present study evaluates the potential role of IQ, education, and gender as moderator variables on AVLT scores.

METHOD

Subjects.

Participants in the present study were 222 successive applicants for civil service positions in the Pacific Northwest. The current employment of the applicants included about 100 different specific jobs. Twenty percent of the jobs were in some aspect of law enforcement, e.g., community service officer, parking patrol, corrections, parole-probation, store detective, police officer. Other current jobs were in the areas of health care and education. The majority of the jobs (about 70%) were in some aspect of business, both white collar and blue collar. The applicants appear to be an occupational cross-section of the community.

The applicants had previously passed basic-academic-skills tests and physical and

agility examinations as part of the job application process. They were deemed to be free of physical illness or limitations. Subjects with identified alcohol or other substance abuse had been dropped from the application process. The subjects in this study are thus a healthier sample than one would obtain by randomly selecting subjects from a community at large. Ages of subjects ranged from 19 to 51 years, with a mean age of 29.1 ($SD = 6.0$). The sample was comprised of 193 males (87%) and 29 females (13%), and included a small representation of 12 (5.4%) racial minority subjects.

Materials and Procedure.

A battery of psychological tests was administered to each participant, including the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1981), the State-Trait Anxiety Inventory (Spielberger, Gorsuch, & Lushene, 1970), a modification of the Stroop Color Naming Test (Stroop, 1935), the Rey-Osterrieth Complex Figure Test (Rey, 1941; Lezak, 1976, 1983), and the Rey Auditory-Verbal Learning Test (Rey, 1964; Lezak, 1976, 1983). The Complex Figure Test (CFT) and the Auditory-Verbal Learning Test were given to assess spatial and verbal learning and memory capacity and the WAIS-R and the Stroop Color Naming Test were administered to assess overall and specific cognitive abilities. Other tests were included in the overall battery, but will not be discussed in this paper.

Directions for administration of the AVLT were those found in Lezak (1983), including the instructions for administering the five learning trials, the distractor trial, the recall trial, and the recognition trial. The rate of word presentation involved a 1-s interval between each word. The word list and recall instructions were repeated after each trial without revealing the number of correct responses. After the fifth learning trial, a second list of 15 words (Lezak, 1983, p. 423) was administered as a distraction task. Following the distractor trial, each participant was asked to remember words from the original list in a postdistraction recall trial. Upon completion of the recall test, each participant was asked to read a short paragraph reported by Lezak (1976, p. 355) and to circle any of the words that were in the first list in an incidental recognition trial. All 15 of the words in the first list are included in the paragraph.

No performance feedback was provided to the participants about the words recalled on the trials. For each participant, results of each learning trial were recorded, including repetitions of words already identified and false identification of words not on the list. Similar results were recorded for performance on the distraction recall, the post-distraction recall trial, and the recognition trial. No delayed recall trial was used in this study.

RESULTS

Performances on each of the learning, distraction, postdistraction recall, and recognition trials were evaluated separately by IQ, age, and education groupings. Percentage recall was computed for each subject by dividing the number of words correct on the postdistraction recall trial by the number correct on the final learning trial (Trial V). Also, a learning index was computed by subtracting the performance on the first learning trial from the performance on the final learning trial as in Query and Berger (1980). The resulting difference score is the increase in words learned as a result of repeated presentation and recall of the same list. The sum of the number of recalled words over all five trials was also computed.

Table 1. Rey Auditory-Verbal Learning Test Scores by WAIS-R Full Scale IQ.

WAIS-R		I		II		III		IV		V		Recall		Recognition	
FISQ	*n*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
80-89	5	8.0	2.5	10.4	1.7	10.8	2.2	11.0	2.1	11.0	3.0	10.6	2.4	14.0	0.7
90-99	29	7.1	1.6	9.7	1.8	11.4	2.1	12.2	1.8	13.0	2.0	11.2	2.2	14.0	1.1
100-109	81	7.2	1.8	9.9	2.5	11.8	2.0	12.4	1.9	12.9	1.8	11.6	2.3	14.2	0.9
110-119	55	7.5	1.7	10.4	2.2	11.9	1.9	13.1	1.4	13.2	1.6	12.1	2.3	14.0	1.2
120-129	38	7.7	1.8	10.7	2.2	12.7	1.7	13.3	1.5	13.7	1.7	12.6	1.9	14.4	0.8
130-139	3	10.0	2.6	12.3	2.5	13.7	1.5	15.0	0	14.7	0.6	14.3	1.2	15.0	0

WAIS-R		Distractor Trial (List B)		Words Learned (Trials V-I)		Percentage Recall		Errors		Repetitions		Total Words (Trials I ... V)	
FSIQ	*n*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
80- 89	5	6.6	2.6	3.0	1.7	99.5	22.0	2.2	1.8	2.6	3.2	51.2	10.9
90- 99	29	6.0	1.6	5.9	2.2	86.7	14.5	3.2	3.9	5.6	5.7	53.4	7.4
100-109	81	6.5	1.5	5.7	1.9	90.1	12.0	2.2	2.9	5.2	5.8	54.2	8.2
110-119	55	6.8	1.5	5.7	2.1	91.8	12.1	2.1	2.4	7.0	6.9	56.1	7.0
120-129	38	7.2	1.9	6.0	1.8	92.5	9.9	2.0	2.3	5.4	6.3	58.1	7.2
130-139	3	7.7	1.5	4.7	2.5	97.9	10.4	2.0	1.7	.7	1.2	65.7	6.7

AVLT and FISQ

Table 1 shows the normative data for each of the recall measures at various FSIQ classifications of the WAIS-R. The distribution of scores extends over a wide range of intelligence levels with a slightly greater representation in the "above average" range than the "below average" range. While the trends show slightly increasing performance on AVLT learning trials with increasing IQ scores, it is of interest to note that performance on the learning trials is quite consistent for all IQ groups except that the very highest group (IQ 130-139) shows greater recall ability on all five trials. This finding should be considered tentative because of the small sample size for that group. Recognition scores are high for all IQ groups and may reflect a ceiling effect.

It may be noted that the number of words recalled on the Distractor Trial (List B) is less than the number recalled on Trial I for all of the FSIQ groups, suggesting some proactive interference from List A. The number of words learned (Trial V—I) seems quite similar for FISQ ranges 90 to 129. Fewer words were learned by both FSIQ extreme groups. In the case of the 130-139 FSIQ group one sees a ceiling effect because the group already had a Trial-I score of 10.0 with a maximum possible score of only 15.0 on Trial V. The percentage of words recalled on the postdistraction recall trial increased linearly for the FSIQ groups 90 to 139. The 80-89 FSIQ group had a mean recall score of 99.5% which is higher than any of the other groups. The five subjects in this group had recall percentage scores of 66.7, 100, 112, and 93. That is, three of the subjects recalled more words on the postdistraction trial than on Trial V. Further research will be necessary to address the question of whether they compensate for fewer words learned by more words remembered. Errors and Repetitions on learning Trials I to V are shown in their respective columns. The last column in the table shows the average sum of the number of recalled words over all five trials. The total words recalled increased linearly for the six FSIQ groups.

Total Group.

Although not shown in Table 1, when the data for all of the FSIQ groups are combined they show that these normal job applicants had a mean score of 7.4 ($SD = 1.8$) words remembered on Trial I and 13.1 ($SD = 1.8$) words remembered on Trial V. Thus, they learned an average of 5.7 ($SD = 2.0$) new words from Trial I to Trial V. Their recall for the words on delayed recall, after an interpolated list had been presented, was 11.9 ($SD = 2.3$) words. That is, they recalled an average of 1.2 ($SD = 1.6$) fewer words on the Recall trial than on Trial V. The average percentage recall was 90.9 ($SD = 12.4$) percent. The average number of total words recalled across all five trials was 55.4 words ($SD = 7.9$).

The values across all trials are slightly lower than the values obtained by Rey (summarized in Lezak, 1983) for normal laborers, professionals, and students. The number of words lost from Trial V to the Recall trial for this group (1.2 words) is slightly less than the number of words lost by graduate students (1.52

Table 2. Rey Auditory-Verbal Learning Test Scores by Age.

		I		II		III		IV		V		Recall		Recognition	
Age	*n*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
20-29	126	7.4	1.7	10.4	2.2	12.2	1.9	13.0	1.7	13.4	1.7	12.1	2.2	14.2	1.0
30-39	71	7.4	1.9	9.9	2.5	11.7	2.0	12.4	1.8	12.7	1.8	11.7	2.2	14.2	1.1
40-49	12	7.3	2.2	9.8	2.7	11.4	2.6	12.3	1.8	12.5	2.5	11.2	3.1	13.8	.9

		Distractor Trial (List B)		Words Learned (Trials V-I)		Percentage Recall		Errors		Repetitions		Total Words (Trials I ... V)	
Age	*n*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
20-29	126	6.8	1.6	6.0	2.0	90.4	12.4	2.2	3.0	5.8	6.1	56.3	7.4
30-39	71	6.5	1.7	5.3	1.9	92.0	12.7	2.3	2.5	5.0	5.5	54.2	8.3
40-49	12	6.6	1.8	5.2	2.6	88.9	10.8	2.7	2.3	7.3	10.1	53.3	10.3

words) as reported in Lezak (1983). These differences in values again call attention to possible effects of differences in test administration or cultural variables.

AVLT and Age.

Table 2 shows the normative data broken down by three age groupings. The fact that this is a relatively young age sample is shown by the distribution of subjects, e.g., 126 in the age range of 20-29 years. There is also a sizeable group of subjects in the 30-39 year age and a small number age 40 or older. All of the learning trials show slight decreases in scores with increasing age (except for Trial I), but the differences are quite slight.

The Distractor Trial Scores, Errors , Repetitions and Percentage of Words Recalled do not consistently change across the three age groupings. The number of words learned (Trials V—I) does decrease slightly across the three age groups. The total number of words recalled across all five trials also decreases slightly across the three age groups.

Although not shown in Table 2, the mean FSIQ scores vary slightly across the three groups. The mean scores are 108.7, 110.5, and 112.9 respectively. Thus, there is the possibility that slight increases in the FSIQ scores may moderate the decrements that could otherwise appear with increasing age. On the other hand, it may not be possible to discern age effects in a sample of relatively young individuals.

AVLT and Education.

Table 3 shows the normative data presented for years of education. All subjects had completed at least 12 years of education. A number of subjects had completed secondary school by passing equivalency examinations (GED) but then had gone on to gain some college credits. As also noted in Table 3, several subjects ($n = 5$) had completed one or more years of graduate education. For the entire sample, the mean educational level was 14.5 years ($SD = 1.5$) with a range of 12 to 19 years.

Examination of the data presented in Table 3 suggests that there is little systematic variability in AVLT scores across years of education except at the extremes of this sample range. Only in Trial V do the mean scores show some incremental linearity with years of education, and significant difference between the means for subjects with 12 years of education and those with 16 years of education ($t = -2.07, p < .05$). The mean scores do not differ for 13 to 16 years of education on any of the learning trials.

None of the additional AVLT scores covary significantly with years of education. Although not shown in Table 3, the mean WAIS-R FSIQ scores for the six levels of education shown in these tables are, respectively, FSIQs of 108.2, 108.7, 108.4, 108.5, 111.2, and 110.2. For the entire sample, the correlation between years of education and WAIS-R FSIQ is .10 (ns).

Table 3. Rey Auditory-Verbal Learning Test Scores by Education.

Years of Education	n	I		II		III		IV		V		Recall		Recognition	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
12	34	7.0	1.6	9.9	3.5	11.7	2.0	11.9	3.5	12.4	2.3	11.4	2.4	13.9	1.2
13	25	7.5	1.2	10.1	2.4	11.9	2.4	12.7	1.7	13.2	1.6	12.1	2.1	13.9	1.2
14	50	7.2	1.9	9.9	2.3	11.8	2.1	13.0	1.9	13.2	2.0	12.3	2.2	14.4	.8
15	19	7.4	2.2	10.3.	2.9	12.4	2.0	12.6	2.1	13.2	1.9	11.4	2.7	14.3	.7
16	80	7.6	1.9	10.5	2.2	12.1	1.8	13.0	1.4	13.3	1.5	12.0	2.1	14.2	.9
17+	5	7.8	2.6	10.4	3.0	12.4	1.1	13.8	1.3	13.4	1.7	11.2	3.1	13.4	1.8

Years of Education	n	Distractor Trial (List B)		Words Learned (Trials V-I)		Percentage Recall		Errors		Repetitions		Total Words (Trials I ... V)	
		Mean	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
12	34	6.6	1.8	5.3	2.0	93.1	13.5	1.9	1.9	6.6	6.7	52.9	7.9
13	25	6.0	1.6	5.7	1.6	91.4	10.4	1.4	1.7	7.1	7.3	55.3	7.7
14	50	6.7	1.8	6.0	2.2	93.3	11.0	2.5	3.2	6.5	6.9	55.2	8.4
15	19	6.6	.9	5.7	2.1	86.0	15.7	2.3	3.5	2.6	2.7	55.8	9.5
16	80	6.8	1.7	5.7	2.0	90.0	11.8	2.6	3.0	5.0	5.6	56.5	7.2
17+	5	7.8	1.1	5.6	3.0	83.0	16.6	1.2	1.6	4.2	1.6	57.8	7.7

AVLT Correlations with Other Test Scores.

Intercorrelations of the five AVLT learning trials and the Recall and Recognition scores with other tests are shown in Table 4. Age and Education are only slightly correlated with the scores on the different learning trials for this subject population (age, —.04 to —.17 for various trials; education, .03 to .18 for various trials). It may be noted that these subjects had a minimum of 12 years of education; the correlation may be attenuated by a restricted range in the education variable. By contrast, completion times for the Stroop Color Naming Test consistently show significant mild negative correlations with AVLT learning trials and Recall and Recognition scores (—.16 to —.29). Somewhat surprisingly, none of the Spielberger State and Trait Anxiety Scale scores correlated with any of the AVLT scores. This observation is not a function of absence of anxiety in this subject population; some of them were noticeably anxious in this assessment situation.

Correlations of AVLT scores with WAIS-R subscale scores are mostly nonsignificant except for the Arithmetic (.13 to .19) and Similarities (.16 to .29) Verbal subscale scores and the Object Assembly (.07 to .24) and Digit Symbol (.11 to .32) Performance subscale scores. Verbal (.12 to .22), Performance (.07 to .28) and Full Scale (.12 to .29) IQ scores show significant correlations with AVLT scores. Since the normative differences (shown in Table 1) are very slight, the correlations between AVLT performance and IQ probably result from the consistency of the relationship and not the magnitude of difference in AVLT performance with varying IQ levels. While the AVLT Recall score shows some positive correlations with WAIS-R variables, the AVLT Recognition score appears to be a fairly independent measure.

Factor-analytic studies like that of Ryan et al. (1984) may be helpful in clarifying what the AVLT measures in addition to verbal learning and memory, e.g., nonverbal and verbal intelligence, perceptual organization, attention-concentration.

DISCUSSION

While the Rey AVLT has been used in several published studies, there has not previously been a report of appropriate norms for healthy adult subjects. These norms should prove useful for neuropsychologists interested in assessing performance for well-motivated, cooperative adults. Compared to a sample of medical patients with unspecified conditions (Query & Megran, 1983), the scores of this sample of healthy job applicants are consistently higher. This finding suggests that norms for healthy people may be of greater usefulness in establishing normative guidelines for assessment of impairment in brain-injured individuals. Furthermore, the present results are different (generally lower) from Rey's (1964) data reported in Lezak (1983) and may provide more up-to-date norms for contemporary Americans.

Table 4. Intercorrelations of AVLT Trials with other Variables.

	AVLT Learning Trials (Number of Words Correct)						
Variables	*I*	*II*	*III*	*IV*	*V*	*Recall*	*Recognition*
Age (Years)	—.04	—.12	—.17*	—.13	—.15*	—.14*	—.08
Education (Years)	.11	.10	.10	.18**	.16*	.03	.08
Stroop Color Naming Test							
Stroop Time (Seconds)	—.17*	—.23***	—.20**	—.29***	—.24***	—.18**	—.16*
Spielberger State-Trait Anxiety Inventory							
State Anxiety Score	.07	—.01	.03	—.05	—.02	.00	.08
Trait Anxiety Score	.04	.06	.08	.03	.01	.03	.10
Wechsler Adult Intelligence Scale-Revised (Scaled Scores)							
Information	.17*	.12	.17*	.16*	.17*	.14*	.05
Digit Span	.12	.09	.14*	.15*	.13	.11	.05
Vocabulary	.14*	.13	.10	.17*	.19**	.12	.11
Arithmetic	.13	.19**	.14**	.18**	.14*	.09	—.01
Comprehension	—.03	.00	.01	.06	.04	.01	.14*
Similarities	.16*	.23***	.29***	.24***	.21**	.22**	.19**
Picture Completion	.00	.07	.16*	.13	.12	.25***	—.02
Picture Arrangement	.05	.12*	.11*	.17*	.12	.16**	.08
Block Design	.05	.14*	.16*	.16*	.13	.19**	.03
Object Assembly	.07	.13	.22**	.18**	.17*	.24***	.11
Digit Symbol	.15*	.18**	.18**	.32***	.23***	.20**	.11
Verbal Scale IQ	.16*	.17*	.20**	.22**	.20**	.17*	.12
Performance Scale IQ	.10	.17*	.21**	.27***	.21**	.28***	.07
Full Scale IQ	.15*	.20**	.25***	.29***	.24***	.27***	.12

* $p < .05$ ** $p < .01$ *** $p < .001$

In summary, for all groups a typical learning curve characterized performance. For the last three learning trials and the postdistraction recall trial, a trend was observed for better recall at higher IQ levels. An inverse trend between age and all learning and memory trials was noted, but the magnitude of differences was small, and the range of ages was limited. These observed trends for age and IQ need to be investigated further due to small sample sizes in the present study for individuals with IQs lower than 90 and higher than 130 and for age groups older than 50. A proactive interference effect was observed for all groups where recall for the second word list was inferior to initial recall of the first word list. There are no consistent trends in frequency of errors or repetitions related to age or IQ variables.

It is of interest that more statistically significant correlations are found among the WAIS-R Performance subtests and AVLT postdistraction recall than among the WAIS-R Verbal subtests and postdistraction recall. A possible interpretation of this finding is that visual abilities such as imagery are more strongly involved in that trial than in earlier trials where a rote learning strategy may be more predominant. Additionally, the consistently high AVLT correlations with the WAIS-R Similarities subtest suggests the hypothesis that abstraction and identifying commonalities among words may also contribute to superior learning and memory on the AVLT.

The insensitivity of the Recognition trial, using the paragraph format, to Age and IQ, raises some interesting questions about the importance of this parameter. For example, it may represent one of the best measures of learning/recall in pathological cases such as dementia where one would expect that Recognition would be poor or reduced. That is, in assessing pathology, one may be better off with a measure that is not strongly sensitive to Age and IQ. One hypothesis to be tested is that, in the case of a depressive condition, one might expect Recognition to be spared and to be significantly greater than Recall. In a dementia, by contrast, there is likely to be a deficit at the level of encoding so that Recognition might show a relative deficit. While Recognition in this normal population does not differ across levels of Age or IQ, its very stability makes it a good measure to employ when studying pathology.

It is of interest to note that AVLT performance did not correlate with anxiety as measured in this study. Replication of this finding would be helpful for increasing clinicians' confidence about obtaining maximum performance during clinical and forensic evaluations. Finally, additional research on organizational processes and AVLT performance, imagery strategies, and the sensitivity of various performance measures is recommended.

Several cautions are in order for the user of these norms. First, the procedural variations in the administration of the test have been significant in past studies. The extent to which these procedural variations influence performance is an empirical question that has not been addressed. The user of these norms should be cautious to follow the administration guidelines found in Lezak (1983).

Second, the representation of females in these norms is minimal. Two-tailed t tests indicated that there were no significant differences beween males and females for most AVLT variables. However, women did make significantly fewer false identifications of words not actually on the lists ($t = 4.94$; $p < .001$). This difference deserves further investigation and therefore caution should be used in interpreting the number of errors made by female subjects.

Third, it is difficult to assess the appropriateness of these norms for older populations. In past research using other memory testing techniques, older subjects performed as well on recognition tasks but not as well on recall tasks (Schonfield & Robertson, 1966). However, Rey (1964) found that older individuals achieved lower scores on both recall and recognition tasks. Thus, caution should be used in applying these norms to older adults.

A PROPOSAL FOR A JOINT PROJECT WITH YOU, THE READER

The data that we report in this paper were collected over a period of several years. Even though we have tested several hundred subjects, this is an insufficient number to generate necessary data for a reasonably complete normative data base. In Tables 5-8, we propose a data matrix that must be completed with additional data before neuropsychologists will have reliable norms referenced for Age and FISQ against which they can evaluate their own individual test subjects.

Our proposal is that we share and pool data. If you, the reader, have tested normal subjects using both the AVLT and the WAIS-R, we are pleased to share our data with you through this publication. Conversely, we would be pleased to know about and receive your data to incorporate it with our own and to complete additional cells in the data matrix presented in Tables 5-8. In any distribution of such combined data we would, of course, acknowledge your contribution.

The utility of these normative data for clinical practice can be greatly enhanced with personal computer software that we have developed for the IBM Personal Computer. Raw scores entered by the clinician (computer operator) are compared with the normative data that are stored in a disk file and a report printout is generated. The report includes the performance centiles for learning, recall, and recognition trials based upon the age and the FSIQ of the test subject. Copies of this software will be made available without cost to readers who pool data with us to expand our Age and FSIQ data matrix.

Table 5. Rey Auditory Verbal Learning Test Scores by Age and IQ; Trial I.

	Full Scale IQ Score (WAIS-R)						
	70 - 79	80 - 89	90 - 99	100 - 109	110 - 119	120 - 129	130 +
10 - 19							
20 - 29		N=4 Min=4 Max=11 Mean=8.0 S.D.=2.9	N=18 Min=5 Max=13 Mean=7.1 S.D.=1.8	N=51 Min=3 Max=11 Mean=7.3 S.D.=1.6	N=29 Min=4 Max=11 Mean=7.4 S.D.=1.7	N=23 Min=4 Max=11 Mean=7.6 S.D.=1.8	N=1 Min= Max= Mean=8 S.D.=
30 - 39		N=1 Min= Max= Mean=8 S.D.=	N=8 Min=6 Max=8 Mean=7.0 S.D.=.93	N=23 Min=2 Max=12 Mean=7.0 S.D.=2.2	N=24 Min=4 Max=10 Mean=7.5 S.D.=1.5	N=12 Min=6 Max=12 Mean=7.6 S.D.=2.0	N=3 Min=9 Max=13 Mean=10.7 S.D.=2.1
40 - 49			N=2 Min=5 Max=9 Mean=7.0 S.D.=2.8	N=5 Min=5 Max=8 Mean=6.2 S.D.=1.6	N=2 Min=6 Max=12 Mean=9.0 S.D.=4.2	N=2 Min=9 Max=9 Mean=9 S.D.=	N=1 Min= Max= Mean=7 S.D.=
50 - 59							
60 - 69							
70 - 79							

Table 6. Rey Auditory Verbal Learning Test Scores by Age and IQ; Trial V.

	Full Scale IQ Score (WAIS-R)						
	70 - 79	80 - 89	90 - 99	100 - 109	110 - 119	120 - 129	130 +
10 - 19							
20 - 29		N=4 Min=8 Max=15 Mean=10.8 S.D.=3.4	N=18 Min=10 Max=15 Mean=13.3 S.D.=1.6	N=51 Min=9 Max=15 Mean=13.2 S.D.=1.7	N=29 Min=10 Max=15 Mean=13.7 S.D.=1.7	N=23 Min=11 Max=15 Mean=14.1 S.D.=1.3	N=1 Min= Max= Mean=15.(S.D.=
30 - 39		N=1 Min= Max= Mean=12.0 S.D.=	N=8 Min=8 Max=15 Mean=13.0 S.D.=2.4	N=23 Min=9 Max=15 Mean=12.5 S.D.=1.2	N=24 Min=11 Max=15 Mean=12.5 S.D.=1.2	N=12 Min=7 Max=15 Mean=12.8 S.D.=2.1	N=3 Min=14 Max=15 Mean=14.7 S.D.=.58
40 - 49			N=2 Min=9 Max=11 Mean=10.0 S.D.=1.4	N=5 Min=8 Max=15 Mean=11.6 S.D.=2.7	N=2 Min=14 Max=15 Mean=14.5 S.D.=.71	N=2 Min=14 Max=15 Mean=14.5 S.D.=.71	N=1 Min= Max= Mean=14 S.D.=
50 - 59							
60 - 69							
70 - 79							

Table 7 Auditory Verbal Learning Test Scores By Age and IQ; Recall.

	Full Scale IQ Score (WAIS-R)						
	70 - 79	80 - 89	90 - 99	100 - 109	110 - 119	120 - 129	130 +
10 - 19							
20 - 29		N=4 Min=8 Max=14 Mean=10.2 S.D.=2.6	N=18 Min=8 Max=15 Mean=11.3 S.D.=2.2	N=51 Min=6 Max=15 Mean=11.9 S.D.=2.1	N=29 Min=8 Max=15 Mean=12.5 S.D.=2.4	N=23 Min=8 Max=15 Mean=13.1 S.D.=1.7	N=1 Min= Max= Mean=13 S.D.=
30 - 39		N=1 Min= Max= Mean=12 S.D.=	N=8 Min=9 Max=13 Mean=11.4 S.D.=1.7	N=23 Min=6 Max=15 Mean=11.3 S.D.=2.5	N=24 Min=8 Max=15 Mean=11.6 S.D.=2.1	N=12 Min=8 Max=15 Mean=11.9 S.D.=2.2	N=3 Min=15 Max=15 Mean=15.0 S.D.=0.0
40 - 49			N=2 Min=6 Max=11 Mean=8.5 S.D.=3.5	N=5 Min=7 Max=15 Mean=14.0 S.D.=3.1	N=2 Min=13 Max=15 Mean=14.0 S.D.=1.4	N=2 Min=12 Max=14 Mean=13.0 S.D.=1.4	N=1 Min= Max= Mean=14 S.D.=
50 - 59							
60 - 69							
70 - 79							

Table 8 Auditory Verbal Learning Test Scores By Age and IQ; Recognition.

	Full Scale IQ Score (WAIS-R)						
	70 - 79	80 - 89	90 - 99	100 - 109	110 - 119	120 - 129	130 +
10 - 19							
20 - 29		N=4 Min=13 Max=14 Mean=13.8 S.D.=.5	N=18 Min=11 Max=15 Mean=14.1 S.D.=1.2	N=51 Min=12 Max=15 Mean=14.2 S.D.=.9	N=29 Min=11 Max=15 Mean=14.2 S.D.=1.0	N=23 Min=12 Max=15 Mean=14.2 S.D.=.9	N=1 Min= Max= Mean=15 S.D.=
30 - 39		N=1 Min= Max= Mean=15 S.D.=	N=8 Min=13 Max=15 Mean=13.9 S.D.=.8	N=23 Min=10 Max=15 Mean=13.9 S.D.=1.3	N=24 Min=10 Max=15 Mean=13.9 S.D.=1.3	N=12 Min=14 Max=15 Mean=14.8 S.D.=.4	N=3 Min=15 Max=15 Mean=150 S.D.=0
40 - 49			N=2 Min=13 Max=14 Mean=13.5 S.D.=.7	N=5 Min=12 Max=14 Mean=13.2 S.D.=.8	N=2 Min=14 Max=15 Mean=14.5 S.D.=.7	N=2 Min=14 Max=15 Mean=14.5 S.D.=.7	N=1 Min= Max= Mean=15 S.D.=
50 - 59							
60 - 69							
70 - 79							

REFERENCES

Chiulli, S.J., Haaland, K.Y., Ellis, H.C., & Rhodes, J.M. (1985, February). *Recall and recognition memory with a variant of the Rey Auditory-Verbal Learning Test in a clinical depressed population.* Paper presented at the meeting of the International Neuropsychological Society, San Diego, CA.

Cohen, R.E., Andres, M.L., & Smolen, R.C. (1986, August). *Rey Auditory-Verbal Learning Test performance in normal, elderly adults.* Paper presented at the meeting of the American Psychological Association in Washington, DC.

Delwaide, P.J., Devoitille, J.M., & Ylieff, M. (1980). Acute effect of drugs upon memory of patients with senile dementia. *Acta Psychiatrica Belgica, 80,* 748-754.

Lezak, M.D. (1976). *Neuropsychological assessment.* New York: Oxford University Press.

Lezak, M.D. (1979). Recovery of memory and learning functions following traumatic brain injury. *Cortex, 15,* 63-70.

Lezak, M.D. (1982, June). *The test-retest stability of some tests commonly used in neuropsychological assessment.* Paper presented at the meeting of the International Neuropsychological Society. Deauville, France.

Lezak, M.D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Miceli, G., Caltagirone, C., & Gainotti, G. (1977). Gangliosides in the treatment of mental deterioration: A double-blind comparison with placebo. *Acta Psychiatrica Scandinavica, 55,* 102-110.

Miceli, G., Caltagirone, C., Gainotti, G., Masullo, C., & Silveri, M.C. (1981). Neuropsychological correlates of localized cerebral lesions in non-aphasic brain-damaged patients. *Journal of Clinical Neuropsychology, 3,* 53-63.

Montgomery, K. & Costa, L. (1983, February). *Neuropsychological test performance of a normal elderly sample.* Paper presented at the meeting of the International Neuropsycnological Society, Mexico City.

Mungas, D. (1983). Differential clinical sensitivity of specific parameters of the Rey Auditory-Verbal Learning Test. *Journal of Consulting and Clinical Psychology, 51,* 848-855.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart and Winston.

Query, W.T., & Berger, R.A. (1980). AVLT memory scores as a function of age among general medical, neurologic, and alcoholic patients. *Journal of Clinical Psychology, 36,* 1009-1012.

Query, W.T., & Megran, J. (1983). Age-related norms for AVLT in a male patient population. *Journal of Clinical Psychology, 39,* 136-138.

Query, W.T., & Megran, J. (1984). Influence of depression and alcoholism on learning, recall, and recognition. *Journal of Clinical Psychology, 40,* 1097-1100.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie, 28,* No. 112, 286-340.

Rey, A. (1964). *L'examen clinique en psychologie.* Paris: Presses Universitaires de France.

Rosenberg, S.J., Ryan, J.J., & Prifiteria, A. (1984). Rey Auditory-Verbal Learning Test performance of patients with and without memory impairment. *Journal of Clinical Psychology, 40,* 785-787.

Ryan, J.J., Geisser, M.E., Randall, D.M. & Georgemiller, R.J. (1986). Alternate form reliability and equivalency of the Rey Auditory Verbal Learning Test. *Journal of Clinical and Experimental Neuropsychology, 8,* 611-616.

Ryan, J.J., Rosenberg, S.J., & Mittenberg, W. (1984). Factor analysis of the Rey Auditory-Verbal Learning Test. *International Journal of Clinical Neuropsychology, 6,* 239-241.

Schonfield, D., & Robertson, B. (1966). Memory storage and aging. *Canadian Journal of Psychology, 20*, 228-236.

Spielberger, D.C., Gorsuch, R.L., & Lushene, R. (1970). *Self evaluation Questionnaire.* Palo Alto, Consulting Psychologists Press, Inc.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology. 18*, 643-662.

Taylor, E.M. (1959). *The appraisal of children with cerebral deficits.* Cambridge,MA: Harvard University Press.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised. Manual.* New York: Psychological Corporation.

Wolf, M.E., Ryan, J.J., & Mosnaim, A.C. (1983). Cognitive functions in tardive dyskinesia. *Psychological Medicine, 13,* 671-674.