

BRIEF REPORT
**Demonstration of Savings on the AVLT and
Development of a Parallel Form***

J. R. Crawford, L. E. Stewart, and J. W. Moore
University of Aberdeen, Aberdeen, U.K.

ABSTRACT

The purposes of the present study were (1) to investigate whether significant savings occur with repeat testing on Rey's Auditory-Verbal Learning Test and (2) to develop a parallel version of the AVLT. Subjects ($N = 60$) were divided into pairs (individually matched for sex, age, and education) to form two groups and were administered either the AVLT or a parallel version. Results indicated no significant differences on all AVLT measures of learning, recall, and recognition, suggesting that the parallel version can be used as an equivalent form of the AVLT. Subjects were retested after an interval of 27 (± 3) days, with half receiving the same version and half receiving a different version. In contrast to subjects receiving different lists, those who were retested with the same version demonstrated a significant improvement in performance on the majority of AVLT variables.

INTRODUCTION

Rey's (1964) Auditory Verbal Learning Test (discussed in detail in Lezak, 1983, pp 422-429) is a serial learning task which uses two lists of 15 concrete nouns. Subjects are read the first list (list A) a total of five times with recall tested following each presentation. Following one presentation and immediate recall of a second list (list B) delayed recall of list A is requested. Finally an optional auditory yes/no recognition test can be presented. This consists of the previously presented word lists randomly embedded amongst semantically or phonetically related words. As has been noted by many authors (e.g., Ivnik, Sharbrough, & Laws, 1987; Lezak, 1983; Squire & Shimamura, 1986), the AVLT is a particularly useful clinical instrument because of the richness of information it provides.

For clinical or research purposes, retesting of a patient's memory performance is clearly often required to determine if change has occurred. In order to do this, parallel forms of memory tests are highly desirable since, if the same stimulus materials are used, savings from previous testing may underestimate

Requests for reprints to: John Crawford, Dept. of Psychology, University of Aberdeen, Aberdeen, U.K.

J Accepted for publication:
August, 1988

progressive decline or give an erroneous impression of stability or recovery. Such savings need not simply be the result of passive retention but may also arise from patients' attempts to reassure themselves concerning their memory by checking their retention of material. One of the strengths of the AVL T lies in the fact that it allows examination of a patient's learning curve. However, as this involves repeated presentation of the same word list, it can be seen that the possibility of savings affecting the results of retesting is greater than in most memory tests. Lezak (1982) carried out repeat testing with the AVL T and reported a significant practice effect on delayed recall of list A when retesting was carried out at 6 month and 1-year intervals.

Ryan, Geisser, Randall, and Georgemiller (1986) reported an attempt to develop a parallel form of the AVL T. They administered the original and parallel form in a counterbalanced order to a group ($N=5$) of patients referred to a psychology department for assessment. The mean test-retest interval was 140 minutes. Unfortunately, they did not report the criteria for the selection of items for the parallel form (with the exception of the replacement for list A) nor did they present the parallel word lists and recognition materials, thereby rendering them unavailable for clinical use. They also reported that the parallel version was significantly more difficult than the original on trial 4 ($p < .001$), trial 5 ($p < .001$), trial 6 ($p < .02$), and the total of trials 1 through 5 ($p < .001$). These differences may have resulted from the use of Lezak's (1983) alternative list (list 'C') as a replacement for list A since analysis of the lists using Thorndike and Lorge's (1944) word count reveals that list C contains more low-frequency words (6 below 'A' frequency) than does list A (2 below 'A' frequency). It has consistently been demonstrated that low-frequency words produce lower scores in free recall (e.g., Anderson & Bower, 1972).

Ryan et al. (1986) reported that the significant differences observed between the original AVL T and the parallel form disappeared when the scores of subjects who had been administered the parallel version first were analysed separately. This may have resulted from the short test-retest interval employed which, as the subjects studied were referred for assessment, was presumably filled with other testing. Fatigue or loss of interest may have reduced performance at retesting thereby obscuring the previously demonstrated differences in difficulty between the two versions. Thus, it would be hazardous to assume that, by administering the parallel version first, the two versions can be rendered equivalent over the test-retest intervals that clinical use would require. Finally, although as noted there were no significant differences in *recall* performance when the scores of subjects administered the parallel form first were analysed separately, a significant difference ($p < .001$) did emerge in *recognition* test performance. This difference was in the opposite direction to the other significant differences observed - i.e., *higher* recognition scores were obtained on the parallel form. This result may also be attributable to differences in word frequency since, although low-frequency words produce lower scores on free recall, they produce higher scores on recognition tests (Anderson & Bower,

1972). In view of the shortcomings noted above, it would appear that further work on a parallel version of the AVLT is desirable.

The purpose of the present study was twofold: (1) to attempt to replicate and extend Lezak's (1982) study of practice effects, and (2) to develop a parallel form which would avoid this problem.

METHOD

Construction of parallel form (see Appendix A): Two lists of 15 concrete words were constructed, each word being individually matched for frequency of occurrence in English usage (Thorndike & Lorge, 1944), word length, and serial position with a word from the original AVLT. The recognition test was constructed by replacing original AVLT words with their equivalents from the parallel list. Words semantically or phonetically related to the new lists were then inserted in the serial positions occupied by words semantically or phonetically related to the original AVLT. The balance between semantically and phonetically related words was maintained.

Sixty subjects, free of neurological, psychiatric, or sensory disability, were recruited from nonmedical health service personnel and the fire service. Subjects were divided into pairs matched for sex, age (± 3 yrs), and years of education (± 1 yr) to form two groups. As a check for possible differences in intellectual level, all subjects were administered the National Adult Reading Test (Nelson, 1982). There was no significant difference (by / test) between the groups' mean estimated IQ (106 and 108).

One group was administered the original AVLT; the other, the parallel version. In both cases the procedure set out by Lezak (1983) was followed. In the recognition test, subjects were asked to inform the examiner if the stimulus words had been contained in the previously presented word lists. A recognition score was obtained by subtracting the number of false positive identifications from the number of words correctly identified. Subjects were retested following a delay of 27 days (± 3 days) with half of each group receiving the same version; the other half, the alternative version.

RESULTS

Mean AVLT scores for the groups receiving either the original or the parallel version are presented in Table 1. It can be seen that the mean scores for the two groups are very similar.

To analyse this statistically, AVLT trials 1 to 5 were entered into a 2 (group) \times 5 (trials) analysis of variance. As would be expected, a significant main effect for trials was obtained, $F(4,232) = 164.59, p < .0001$, indicating that performance improved across trials. There was no significant main effect for group, $F(1,58) = 0.57, p = .45$, and no significant Group \times Trials interaction $F(4,232) = 0.58, p = .67$. Additional analysis (by / tests) revealed that the two groups did not differ significantly on recall of list B ($t = 0.88, p = .38$), delayed recall of list A ($t = 0.79, p = .43$) and recognition performance ($t = 0.32, p = .75$). The results of the foregoing analysis suggest that the parallel version of the AVLT can be

Table 1
Mean Scores (*SDs*) for Matched Groups Receiving the Original AVLT and the Parallel Version

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	List B	Trial 6	Recognition
Original	8.30 (1.80)	11.00 (2.11)	11.80 (2.36)	12.73 (2.02)	13.20 (1.61)	6.90 (2.22)	11.90 (2.55)	25.37 (2.68)
Parallel	7.37 (1.67)	10.50 (2.52)	11.70 (2.25)	12.63 (1.50)	13.00 (1.48)	6.43 (1.89)	11.43 (2.00)	25.13 (2.96)

considered equivalent to the original.

To investigate the effects of practice on AVLT performance, subjects were reassigned into two groups: (1) subjects who received the *same* AVLT version at test and retest, and (2) subjects who received *different* AVLT versions at test and retest. Mean AVLT scores for these two groups at test and retest are presented in Table 2. A 2 (test vs. retest) by 5 (trials) analysis of variance was carried out for the group receiving the same version at test and retest ($n = 30$). A significant main effect for trials was obtained, $F(4, 116) = 122.28, p < .0001$. A significant main effect for test vs. retest was also obtained, $F(1, 29) = 43.12, p < .0001$. Post-hoc analysis using Tukey's Honestly Significant Difference Test (Hays, 1981, pp. 434-438) revealed that performance was significantly higher at retest on all AVLT trials except for trial 5. A significant Test/Retest x Trials interaction was also obtained, $F(4, 116) = 5.72, p = .0003$. This indicates that the improvement at retesting was not uniform across AVLT trials but rather was greater on earlier

Table 2
Mean Scores (*SDs*) for Groups Receiving Either the Same or Different AVLT Version at Retest.

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	List B	Trial 6	Recognition
Test	7.87 (1.76)	11.10 (2.19)	11.93 (2.00)	13.03 (1.59)	13.33 (1.56)	6.70 (2.40)	11.93 (1.95)	25.30 (2.47)
Same Version								
Retest	10.53 (2.39)	12.87 (1.81)	13.67 (1.40)	13.90 (1.35)	14.13 (1.14)	7.70 (2.29)	13.43 (1.68)	26.67 (2.47)
Test	7.53 (1.76)	10.40 (2.43)	11.57 (2.56)	12.33 (1.88)	12.87 (1.80)	6.63 (1.70)	11.40 (2.58)	25.20 (3.15)
Different Version								
Retest	7.50 (2.13)	10.27 (2.16)	11.87 (2.00)	12.70 (1.82)	12.90 (2.11)	6.23 (1.91)	11.93 (2.77)	24.57 (3.67)

trials. A series of t tests revealed that there were significant improvements at retest on recall of list B ($t = 2.49, p = .02$), delayed recall of list A ($t = 5.47, p < .001$), and recognition performance ($t = 2.95, p = .006$). It is apparent that highly significant practice effects occurred when subjects were retested with the same version.

The same statistical procedures were used to compare the test and retest performance of subjects receiving a different AVLT version at test and retest ($n = 30$). The 2 (test vs. retest) by 5 (trials) analysis of variance yielded a significant main effect for trials, $F(4,116) = 147.19, p < .0001$. There was neither a significant main effect for test vs. retest, $F(1,29) = 0.18, p = .67$, nor a significant Test/Retest \times Trials interaction, $F(4,116) = 0.49, p = .75$. Thus, there was no significant improvement at retesting on AVLT trials when subjects were given a different version. A series of t tests revealed that there were no significant differences between test and retest performance on recall of list B ($t = 1.18, p = .25$), delayed recall of list A ($t = 1.35, p = .19$), and recognition performance ($t = 1.23, p = .23$). These results indicate that no significant improvements in performance occur when subjects are retested with a different AVLT version.

DISCUSSION

The present demonstration of highly significant practice effects on the AVLT is consistent with Lezak's (1982) earlier study and reinforces the need for a parallel form. The present attempt to develop a parallel form would appear to have been successful since none of the comparisons between the matched groups administered the original AVLT or the new lists approached statistical significance.

Finally, two factors can lead to practice effects on memory tests. First, subjects may retain the specific test material over the test-retest interval, second, a metamemory factor may operate - i.e., prior experience of the test demands may lead to improvement at retesting because the optimal memory strategy has been ascertained. This second factor poses a potential problem for clinical memory testing, as the use of a parallel form will not necessarily entirely solve the problem of practice effects. In the present study, comparison of the test-retest performance of subjects given different versions allowed an examination of this metamemory factor in isolation. As no significant differences emerged between the test and retest performance for these subjects, it would appear that this potential factor did not influence AVLT performance.

REFERENCES

- Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97-123.
- Hays, W. L. (1981). *Statistics for the Social Sciences (3rd ed.)*. New York: Holt, Rinehart & Winston.
- Ivnik, R. J., Sharbrough, F. W., & Laws, E. R. (1987). Effects of anterior temporal lobectomy on cognitive function. *Journal of Clinical Psychology*, 43, 128-137.
- Lezak, M. D. (1982 June). The test-retest stability and reliability 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.
- Nelson, H. E. (1982). *National Adult Reading Test*. Test Manual. Windsor: NFER-Nelson.
- Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.
- 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.
- Squire, L. R., Shimamura, A. P. (1986). Characterizing amnesic patients for neurobehavioral study. *Behavioral Neuroscience*, 100, 866-877.
- Thorndike, E. L., & Lorge, I. (1944). *The teacher's wordbook of 30,000 words*. New York: Teachers College, Columbia University.

DEMONSTRATION OF SAVINGS ON THE AVLT

Appendix A: Parallel version of

AVLT word lists and recognition test

RECOGNITION

NAIL (A)	ENVELOPE (SA)	LADDER (B)	FOOT(B)	WATER (B)
SAND(SA)	CAR(B)	MIRROR (A)	BREAD (PA)	JOKER (SB)
BED (A)	FACE (A)	SCREW (SA)	DESERT (A)	COAT(B)
PONY(SA)	TOAD (PA)	MUSIC (A)	STREET (SA)	CAPTAIN (SA)
JESTER (B)	SILK (PA)	DISH (B)	MACHINE (A)	TOOL (B)
MILK (A)	HILL(B)	PIE(B)	HEAD (SPA)	FLY(SPB)
PLATE (SB)	FOREST (B)	WOOD (SB)	GIRL(B)	SONG(SA)
HEART (A)	SAILOR (A)	BALL(B)	HORSE (A)	DOLL (A)
JAIL (PA)	DART (PA)	HELMET (A)	SOOT(PB)	STALL (PA)
INSECT (B)	ROAD (A)	STOOL (PB)	LETTER (A)	SHIELD (B)

Abbreviations: A = list A word; B = list B word; SA = word semantically related to list A word; PA = word phonetically related to list A word; SB = word semantically related to list B word; PB = word phonetically related to list B word; SPA = word semantically *and* phonetically related to list A word; SPB = word semantically *and* phonetically related to list B word.

WORD LISTS

<i>A</i>	<i>B</i>
DOLL	DISH
MIRROR	JESTER
NAIL	HILL
SAILOR	COAT
HEART	TOOL
DESERT	FOREST
FACE	WATER
LETTER	LADDER
BED	GIRL
MACHINE	FOOT
MILK	SHIELD
HELMET	PIE
MUSIC	INSECT
HORSE	BALL
ROAD	CAR