

On the Underestimation of Auditory Verbal Learning Impairments in Temporal Lobe Epilepsy

John M Hudson,^{1,*} Kenneth A Flowers,^{1,2} and Lauren E Morgan¹

¹School of Psychology, University of Lincoln, Lincoln, UK

²Department of Psychology, University of Hull, Hull, UK

*Corresponding author: John M Hudson, School of Psychology, College of Social Science, Brayford Pool, Lincoln LN6 7TS, UK. Tel: +44-01522886782, Fax: +44-01522886026, E-mail: jhudson@lincoln.ac.uk

Received 2016 October 17; Revised 2016 November 13; Accepted 2016 December 11.

Abstract

Background: The auditory verbal learning test (AVLT) procedure is routinely deployed in neuropsychological investigations to examine learning and memory status in research and clinical cohorts. Concerns however have been raised regarding the susceptibility of the standard AVLT procedure to ceiling effects, which may have adverse consequences for psychometric properties and result in an underestimation of true potential and differences between normal and abnormal scores.

Methods: We examined the performance of patients with temporal lobe epilepsy (TLE; n = 40) who had completed a standard 15-item AVLT and compared a group of TLE patients (n = 12) with healthy controls (n = 12) who completed an extended 24-item AVLT, which was designed to minimise the probability of ceiling scores.

Results: Ceiling effects on at least one trial (≥ 14) was achieved by 33% of patients on the 15-item test, with 60% of patients scoring within or above the average list learning total score. Increasing the list length to 24-items reduced the percentage of TLE patients scoring within the normal range to 42%. In addition, no patients but 25% of control participants achieved a maximum score on trial A5. The performance of controls was superior to patients for the best learning trial, learning rate and total learning score.

Conclusions: Increasing the list length to 24-items eliminated ceiling scores in all TLE patients and most controls and allowed the true magnitude in difference between the groups to be observed. These findings have implications for decisions relating to optimal AVLT list lengths that might be deployed for memory assessment in TLE.

Keywords: Verbal Learning, Neuropsychological Test, Memory, Temporal Lobe Epilepsy

1. Background

The neuropsychological assessment of patients with temporal lobe epilepsy (TLE) is pivotal for making clinical decisions relating to the: diagnosis (1), progression (2), effects of anti-epileptic drugs (3), lateralisation and the suitability of patients for resective surgery (4). An important feature of all assessment protocols is an evaluation of memory and the ability to acquire, consolidate, retain and recall verbal material. The most common measure of these abilities is the auditory verbal learning test (AVLT). Although there are a number of different AVLTs; the Rey AVLT (5) and word-list learning tests from the wechsler memory scale (WMS; (6)), the adult memory and information processing battery (AMIPB; (7)) and the BIRT (brain injury rehabilitation trust) memory and information processing battery (BMIPB; (8)) they generally involve a similar format. For each of these tests, the examiner reads aloud a list of 15 unrelated words (The WMS presents 12 words over four study-test trials) after which the examinee recalls as many words as possible. The process is usually repeated over five study-test trials (A1 to A5) before a novel interference list is presented in a single study-test trial (B1). Examinees then

attempt to recall the original word list in a short-delay condition (A6) and once again in a long-delay condition (A7) following a 20 minute interval.

Notwithstanding the status of the AVLT in many assessment batteries, concerns have been raised over the possibility of low ceiling effects on these measures and the implications they may have for reducing test reliability and validity (9). Ceiling effects are undesirable for many reasons, in particular they: a) preclude distinguishing between high-scoring individuals; b) constrain the range of scores and thus underestimate measures of central tendency and dispersion; c) may result in spurious research interactions; d) reduce the sensitivity of a test to identify instances of cognitive dysfunction because the difference between normal and abnormal scores are artificially contracted.

Although a number of studies report AVLT deficits in TLE (10, 11) there is evidence that ceiling effects may confound an accurate assessment of memory status in these patients (12). The aim of the current study is to compare the performance of TLE patients with healthy participants on an AVLT in which the probability of ceiling effects is

designed to be minimal. The incidence of ceiling effects across AVLT formats that differ as a function of word list length has been examined in one study (13). On the traditional 15-item test the prevalence of ceiling effects were clearly evident. Fourteen or more words were correctly recalled by 46% of participants on trial A4, 54% on trial A5 and 49% on both trials A6 and A7. The author concluded that the maximum score of a study-test word list should exceed the mean score by at least 1.5 standard deviations in order to avoid the risk of ceiling effects. The 24-item word list came closest to this criterion and was subsequently chosen for the present investigation.

2. Methods

2.1. Participants

Twelve potential surgery candidates with TLE (5 male/7 female; mean = 33 years, SD = 11.38) and 12 healthy control participants (7 male/5 female; mean = 35 years, SD = 9.96) recruited through opportunity sampling completed the 24-item AVLT. To assess standard AVLT performance in TLE we examined data from a further 40 patients (19 male/21 female; mean = 36 years, SD = 9.46) who had been referred for neuropsychological assessment and had completed the 15-item word list from the AMIPB as part of their pre-surgical evaluation. All patients with TLE were on anti-epileptic medication. MRI scans and EEG recording determined lesion focus and epileptogenic activity respectively. The 24-item AVLT group included 6 left and 6 right-sided patients; the 15-item AVLT group included 17 left, 21 right-sided, and 2 bilateral patients. All participants had normal or corrected to normal vision. The investigation was approved by the Hull area health trust ethical committee as part of the university of Hull's clinical neuroscience centre's neuropsychological research and assessment programme.

2.2. Measures

Word lists consisting of 24 nouns and verbs between three and five letters in length were constructed. Each participant was presented with one word list for trials A1 to A5 and a different word list for trial B1. Word lists were rotated across study-test trials between participants. The AVLT was administered in accordance with the abovementioned format and standardised instructions from AMIPB for word-list learning (Coughlan, & Hollows, 1985).

Other neuropsychological tests (see Table 1): Auditory attentional capacity was assessed with digit span (14). Sustained attention and executive functioning were measured with the elevator counting task (TEA-2; (15)) and letter fluency respectively (14). The NART (16) was used to estimate verbal IQ. The hospital anxiety and depression scale (HADS; (17)) provided a measure of emotional status.

3. Results

3.1. 24-Item Test

Mean recall scores for 24-item and 15-item AVLTs are displayed in Figure 1. For the 24-item AVLT a 2×5 ANOVA of learning, treating group as a between-subjects factor (TLE or control) and trials (A1 to A5) as a within-subjects factor revealed a main effect of trial indicating an increase across trials as a function of learning [$F(4, 88) = 85.199, P < 0.001$]. A main effect of group [$F(1, 22) = 11.255, P < .003, \eta^2 = 0.34$] indicated controls recalled more words than TLE patients. Learning significantly interacted with group [$F(4, 88) = 6.723, P < 0.001, \eta^2 = 0.23$] suggesting a steeper learning rate for controls. Indeed, control participants were superior in terms of the index of learning (A5 - A1) [$t(22) = 3.471, P < .002, \eta^2 = .35$], best learning trial (A5) [$t(22) = 3.746, P < .001, \eta^2 = 0.39$] and total learning (A1 to A5) [$t(22) = 3.397, P < .003, \eta^2 = 0.34$]. Notably, this pattern prevailed after the inclusion of verbal IQ as a covariate [$F(1, 21) = 1.874, P < 0.185$]. Therefore the main effect of group on total AVLT performance is independent of intellect.

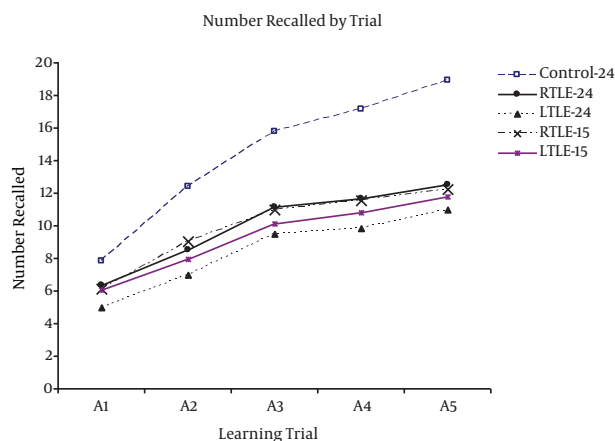


Figure 1. Mean Number of Words Recalled by Trial for Controls, Right and Left TLE Patient Groups as a Function of AVLT List Length

In addition to acquisition, we also analysed group differences for interference and forgetting. There was no significant difference for proactive interference (A6/A1) [$t(22) = 1.659, P > 0.111$], retroactive interference (A6/A5) [$t(22) = 1.004, P > 0.326$] nor rate of forgetting (A7/A6) [$t(22) = 0.382, P > 0.706$]. None of these measures interacted with lesion focus (all P 's > 0.17).

3.2. 15-Item Test

The mean total recall score from those TLE patients who completed the 15-item AVLT (mean = 47.9, range = 37 - 71) were within one standard deviation of the averaged

Table 1. Performance Means (SD) on Neuropsychological Measures for TLE Patients and Healthy Participants Who Completed the 24-Item AVLT

Measure	TLE	Controls	P <
Education (years)	12.33 (2.02)	12.08 (1.31)	0.722
General intellect			
Estimated verbal-IQ	87.92 (9.18)	115.25 (2.70)	0.001
Attention span			
Digits forwards	5.92 (0.99)	6.25 (1.14)	0.247
Digits backwards	4.67 (0.78)	4.42 (0.99)	0.041
Sustained attention			
TEA-2	6.30 (1.49)	6.76 (0.65)	0.144
Letter fluency			
FAS	29.42 (7.40)	51.75 (11.58)	0.001
Emotional status			
Anxiety	6.27 (3.93)	9.33 (3.81)	0.118
Depression	9.46 (4.41)	6.33 (2.87)	0.055

norms (7). Ceiling effects on at least one trial (≥ 14) was achieved by 33% of patients on the 15-item test.

3.3. 15 vs. 24-Item Tests

In order to examine the effects of list length in TLE a 2×5 ANOVA treating group as between-subjects factor (15-item or 24-item) and trials (A1 to A5) as a within-subjects factor revealed a main effect of trial indicating an increase across trials as a function of learning [$F(4, 200) = 89.562$, $P < .001$, $\eta^2 = 0.64$]. Strikingly, a main effect of group was not observed [$F(1, 50) = 0.222$, $P > 0.640$], presenting TLE patients with a longer list length did not lead to an increase in recall (see Figure 1). The interaction between learning and list length was not significant [$F(4, 200) = 0.254$, $P > 0.907$], indicating that the rate of learning is not influenced by list length. The total recall score for controls was on average 57% greater than that of TLE patients on the 24-item test; however when those patients who completed the 15-item test were compared to normative AMIPB scores this difference was reduced to less than 12%.

No effect of lesion laterality was found on the acquisition trials of either the 15-item or 24-item AVLT (all P 's > 0.19). Left TLE patients however recalled significantly fewer words than right-sided patients on both shorter (A6) [$t(36) = 3.126$, $P < 0.003$, $\eta^2 = 0.21$] and longer (A7) [$t(36) = 2.971$, $P < .005$, $\eta^2 = 0.20$] delayed recall trials of the 15-item AVLT.

3.4. Serial Position Effects

To examine nominal serial position effects the number of words recalled over the acquisition trials were divided as

a function of word order into eight groups. Each group corresponded to three successive words. These were analysed with a 2×8 ANOVA, treating group as a between-subjects factor (TLE or control) and nominal serial position (1 to 8) as a within-subjects factor. A main effect of position [$F(7, 154) = 15.157$, $P < .001$, $\eta^2 = 0.41$] confirmed typical primacy and recency effects. The first (serial positions 1-3; primacy) and last three words (serial positions 22-24; recency) in the list were recalled more frequently than those in middle order positions which were asymptote. Though there was no significant difference in the frequency of recall of words positioned in primacy and recency sections. Controls recalled more words at each serial position [$F(1, 22) = 12.69$, $P < 0.002$, $\eta^2 = 0.37$] than patients. The interaction between group and serial position was not significant [$F(7, 154) = 0.845$, $P < 0.552$].

4. Discussion

Although AVLTs are commonly included in memory assessment protocols, questions remain over the validity of many tests deployed in clinical practice and research studies (13). This study sought to compare the status of AVLT performance in TLE patients with healthy controls when the likelihood for ceiling effects is minimal. When the maximum score per trial was set at 24, 75% of controls recalled at least 15 words on trial A5, suggesting that 15-item AVLT procedures prevent healthy participants from demonstrating their mnemonic potential. Thus the difference in performance between the groups is markedly larger than the use of conventional word lists imply and

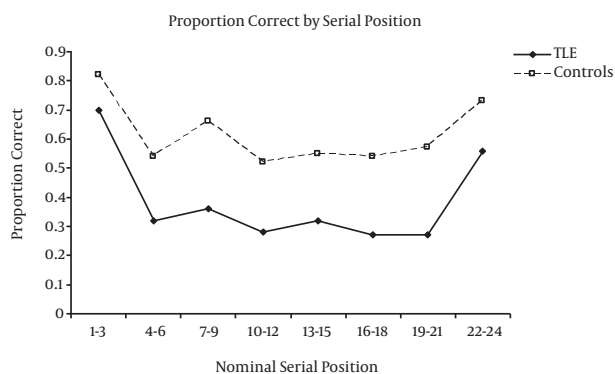


Figure 2. Mean Proportion of Correct Responses for the 24-item AVLT as a Function of Serial Position for TLE Patients and Healthy Controls

therefore the degree of memory impairment in TLE is likely to be severely underestimated with most standardised AVLT formats. Indeed, 67% of right-sided and 53% of left-sided patients taking the 15-item test scored within the normal AMPIB range. If means and standard deviations derived from healthy individuals are skewed and artificially low, the probability for false negatives is likely to increase thus reducing test sensitivity and patients with verbal learning deficits may not be identified.

It is noteworthy that TLE patients did not demonstrate the standard list length effect (18). Normally, increasing the length of a study list is accompanied by an increase in the total number but a decrease in the proportion of words recalled. In TLE across test formats total recall did not increase as a function of list length. It has been posited that list length effects are primarily the result of selective rehearsal which facilitates recollection of words towards the end of the list and thus produces an extended recency effect (19). A lack of a list length effect in TLE may therefore reflect a short-term storage deficit. There are however limitations with this interpretation. First, patients were unimpaired in digit span and TEA-2 performance and so appear to exhibit normal short-term auditory memory. Second, in relation to the asymptote portion of the serial positional curve, TLE patients demonstrated a pronounced recency effect.

Although across studies (10-12) the 15-item AVLT appears suitable for assessing memory status in most TLE patients - ceiling effects may still be apparent. Fifteen per cent of our TLE patients recalled the maximum number of words on trial A5 of the 15-item AVLT. To significantly reduce the chance of ceiling effects in TLE however, only a moderate increase in list length may be required. On the 24-item test for example, not a single patient recalled more than 18 words on any trial. Therefore patients with

TLE may not actually demonstrate a list length effect because the maximum potential of the highest performers is only marginally greater than that revealed on traditional 15-item formats. In fact the highest total score for the TLE group on the 15-item test was 68 (mean = 47.95) compared to 69 (mean = 46.25) for the 24-item test.

What then is the optimal format to assess auditory verbal learning in TLE? In the present study over five study-test trials the 24-item test quantified individual differences in memory in 87% of the sample. However, 83% of the TLE group recalled less than 17 words on trial A5. Consequently, it could be argued that a 24-item list length places a disproportionate degree of cognitive load on TLE patients, which may indeed have an adverse effect on test performance by producing fatigue and reducing motivation (14). To circumvent these caveats whilst still abating the probability of ceiling effects in controls, one may choose to reduce the number of acquisition trials and possibly list length as well. For example, here shortening the AVLT format to 17-items over three trials would restrict ceiling effects to 20% of participants (all controls). Alternatively, we advocate shortening to 22-items over three trials which would eliminate ceiling scores in the entire cohort.

One limitation of this study is the relatively small sample size for the 24-item AVLT. However, the mean recall scores across trials A1 to A5 from our healthy controls are within half a standard deviation of those reported previously (13) who tested a slightly larger and younger sample ($n = 36$; mean = 20 years). Therefore although it is important to examine the generality of the findings reported here, we don't believe the sample size is a caveat per se to the arguments put forth.

4.1. Conclusion

AVLTs are among the most frequently deployed measures of memory processing and are used extensively in the neuropsychological assessment of patients with TLE. Thus AVLT performance influences clinical and research decisions on diagnosis, management and progression. The validity and reliability of common AVLT procedures have nonetheless been shown to be compromised by ceiling effects. These restrictions preclude many healthy examinees from demonstrating their maximum potential. Traditional AVLT formats may therefore underestimate or fail to detect memory dysfunction in TLE. The results from this study confirm that extending the list length of words can circumvent ceiling effects. The presentation of more words did not at least at a group level enhance performance in TLE however and may increase cognitive load. Therefore for optimal performance across groups a reduction in the number of learning trials is a further modification to be considered. The principles highlighted here are

not of course unique to TLE but applicable to neuropsychological assessment in general. Further studies are needed to examine AVLT performance in other memory impaired cohorts using test parameters that limit ceiling effects.

Acknowledgments

The authors thank Dr Alec Ming and Dr Jayati Tarafder at the department of neurology, Hull royal infirmary, for access to patients under their care.

References

1. Jones-Gotman M, Smith ML, Risse GL, Westerveld M, Swanson SJ, Giovagnoli AR, et al. The contribution of neuropsychology to diagnostic assessment in epilepsy. *Epilepsy Behav.* 2010;**18**(1-2):3-12. doi: [10.1016/j.yebeh.2010.02.019](https://doi.org/10.1016/j.yebeh.2010.02.019). [PubMed: [20471914](https://pubmed.ncbi.nlm.nih.gov/20471914/)].
2. Strauss E, Loring D, Chelune G, Hunter M, Hermann B, Perrine K, et al. Predicting cognitive impairment in epilepsy: findings from the Bozeman epilepsy consortium. *J Clin Exp Neuropsychol.* 1995;**17**(6):909-17. doi: [10.1080/01688639508402439](https://doi.org/10.1080/01688639508402439). [PubMed: [8847396](https://pubmed.ncbi.nlm.nih.gov/8847396/)].
3. Thompson PJ, Trimble MR. Neuropsychological aspects of epilepsy. Oxford University Press; 1996.
4. Chelune GJ. Hippocampal adequacy versus functional reserve: predicting memory functions following temporal lobectomy. *Arch Clin Neuropsychol.* 1995;**10**(5):413-32. [PubMed: [14588901](https://pubmed.ncbi.nlm.nih.gov/14588901/)].
5. Rey A. L'examen psychologique dans les cas d'encéphalopathie traumatique (Les problems.). *Arch Psychol*; 1941.
6. Wechsler D. Wechsler memory scale (WMS-III). San Antonio, TX: Psychological Corporation; 1997.
7. Coughlan AK, Hollows SE. The adult memory and information processing battery (amipb): Test manual. AK Coughlin, Psychology Department, St James' Hospital; 1985.
8. Coughlan AK, Oddy M, Crawford AR. BIRT memory and information processing battery (BMIPB). *PSIGE Newsletter.* 2007;29.
9. Graf P, Uttl B. Component processes of memory: Changes across the adult lifespan. *Swiss J Psychol.* 1995;**54**(2):113-30.
10. 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**(7):693-703. doi: [10.1016/j.acn.2006.06.010](https://doi.org/10.1016/j.acn.2006.06.010). [PubMed: [16987634](https://pubmed.ncbi.nlm.nih.gov/16987634/)].
11. Taylor J, Kolamunnage-Dona R, Marson AG, Smith PE, Aldenkamp AP, Baker GA, et al. Patients with epilepsy: cognitively compromised before the start of antiepileptic drug treatment?. *Epilepsia.* 2010;**51**(1):48-56. doi: [10.1111/j.1528-1167.2009.02195.x](https://doi.org/10.1111/j.1528-1167.2009.02195.x). [PubMed: [19583779](https://pubmed.ncbi.nlm.nih.gov/19583779/)].
12. Mameniskiene R, Jatuzis D, Kaubrys G, Budrys V. The decay of memory between delayed and long-term recall in patients with temporal lobe epilepsy. *Epilepsy Behav.* 2006;**8**(1):278-88. doi: [10.1016/j.yebeh.2005.11.003](https://doi.org/10.1016/j.yebeh.2005.11.003). [PubMed: [16359927](https://pubmed.ncbi.nlm.nih.gov/16359927/)].
13. Uttl B. Measurement of individual differences: lessons from memory assessment in research and clinical practice. *Psychol Sci.* 2005;**16**(6):460-7. doi: [10.1111/j.0956-7976.2005.01557.x](https://doi.org/10.1111/j.0956-7976.2005.01557.x). [PubMed: [15943672](https://pubmed.ncbi.nlm.nih.gov/15943672/)].
14. Lezak MD. Neuropsychological assessment. USA: Oxford University Press; 2004.
15. Robertson IH, Ward T, Ridgeway V, Nimmo-Smith I, McAnespie AW. The test of everyday attention (tea). *Bury St. Edmonds, United Kingdom: Thames Valley Test Company.* 1991.
16. Nelson HE, Willison J. National adult reading test (nart). Nfer-Nelson Windsor; 1991.
17. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;**67**(6):361-70. [PubMed: [6880820](https://pubmed.ncbi.nlm.nih.gov/6880820/)].
18. Murdock BB. The serial position effect of free recall. *J Experiment Psychol.* 1962;**64**(5):482-8. doi: [10.1037/h0045106](https://doi.org/10.1037/h0045106).
19. Ward G. A recency-based account of the list length effect in free recall. *Mem Cognit.* 2002;**30**(6):885-92. [PubMed: [12450092](https://pubmed.ncbi.nlm.nih.gov/12450092/)].