ELSEVIER

Archives of Clinical Neuropsychology 21 (2006) 693–703

Archives of CLINICAL NEUROPSYCHOLOGY

Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples

Mike R. Schoenberg [a,*], Kyra A. Dawson [b], Kevin Duff [c], Doyle Patton [d], James G. Scott [e], Russell L. Adams [e]

[a] *Department of Neurology, University Hospitals Case Medical Center, HH 5, 11100 Euclid Ave., Cleveland, OH 44106-5000, United States*
[b] *Department of Psychology, Cleveland State University, OH, United States*
[c] *Department of Psychiatry, University of Iowa College of Medicine, IA, United States*
[d] *Private Practice, Boca Raton, FL, United States*
[e] *Department of Psychiatry and Behavioral Sciences, University of Oklahoma Health Sciences Center, OK, United States*

Accepted 16 June 2006

Abstract

The Rey Auditory Verbal Learning Test [RAVLT; Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie*, *28*, 21] is a commonly used neuropsychological measure that assesses verbal learning and memory. Normative data have been compiled [Schmidt, M. (1996). *Rey Auditory and Verbal Learning Test: A handbook*. Los Angeles, CA: Western Psychological Services]. When assessing an individual suspected of neurological dysfunction, useful comparisons include the extent that the patient deviates from healthy peers and also how closely the subject's performance matches those with known brain injury. This study provides the means and S.D.'s of 392 individuals with documented neurological dysfunction [closed head TBI ($n = 68$), neoplasms ($n = 57$), stroke ($n = 47$), Dementia of the Alzheimer's type ($n = 158$), and presurgical epilepsy left seizure focus ($n = 28$), presurgical epilepsy right seizure focus ($n = 34$)] and 122 patients with no known neurological dysfunction and psychiatric complaints. Patients were stratified into three age groups, 16–35, 36–59, and 60–88. Data were provided for trials I–V, List B, immediate recall, 30-min delayed recall, and recognition. Classification characteristics of the RAVLT using [Schmidt, M. (1996). *Rey Auditory and Verbal Learning Test: A handbook*. Los Angeles, CA: Western Psychological Services] meta-norms found the RAVLT to best distinguish patients suspected of Alzheimer's disease from the psychiatric comparison group.

Keywords: Auditory Verbal Learning Test; RAVLT; Memory; Positive predictive power; Test scores; Diagnosis

A central tenet in neuropsychological practice is comparing the performance of a subject suspected of brain injury to scores obtained by persons without brain injury who have a similar demographic background as the patient with suspected neurological injury. When using norm referenced test data to determine the presence of brain dysfunction, errors occur when intact individuals are labeled as having brain dysfunction (false positives) or when neurological injury is not detected (false negatives). An important factor in determining the diagnostic utility of a norm referenced test in clinical neuropsychology is the degree to which the distribution of scores obtained by persons with brain injury differs from healthy individuals without known brain injury (e.g., Nunnally, 1978; Petersen et al., 2001; Retzlaff &

* Corresponding author. Tel.: +1 216 844 5820; fax: +1 216 844 1548.
E-mail address: Michael.Schoenberg@uhhs.com (M.R. Schoenberg).

0887-6177/$ – see front matter © 2006 National Academy of Neuropsychology. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.acn.2006.06.010

Gibertini, 2000). The clinical utility of a diagnostic test is a function of the test's false negative and false positive rate (e.g., Anastasi & Urbina, 1997). That is, a test's diagnostic utility may be defined by its ability to correctly identify a person with neurological dysfunction while simultaneously not classifying healthy individuals as having neurological dysfunction (Nunnally, 1978). A test's sensitivity (SENS) is the probability a test will correctly identify neurological dysfunction in a patient that is known to have neurological impairment. A test's specificity (SPEC) is the probability that a patient known to be without neurological dysfunction will have a negative test result. Unfortunately, in a diagnostic setting, the presence of neurological dysfunction is often not known, and a more useful index of a test's diagnostic utility is its positive predictive power (PPP). Given that a person has a positive test result, the PPP is the likelihood the person actually has neurological impairment. The negative predictive power (NPP) is the probability a negative test result (below cut-off criterion) correctly identifies a person without neurological impairment. Another useful index is the overall hit rate (HR), which is the ratio of true positives and true negatives compared to the total number of classifications (i.e., true positives, true negatives, false positives, and false negatives). Hence, comparing an individual's test score to scores obtained by healthy peers as well as to patients with neurological dysfunction are both important indices of a diagnostic test (Nunnally, 1978; Retzlaff & Gibertini, 2000).

The Rey Auditory Verbal Learning Test (RAVLT; Rey, 1941; Taylor, 1959) is a commonly used measure of a person's ability to encode, consolidate, store, and retrieve verbal information (see Schmidt, 1996, for a review). While the RAVLT has been found to be a sensitive test of verbal learning and memory (Bigler, Rosa, Schultz, Hall, & Harrison, 1989; Butters, Wolfe, Martone, Granholm, & Cermak, 1985; Davidoff et al., 1990; Drebing, Van Gorp, Stuck, Mitrushina, & Beck, 1994; Ivnik, Smith, Malec, Kokmen, & Tangalos, 1994; Mungas, 1983; Petersen et al., 1999; Powell, Cripe, & Dodrill, 1991; Squire & Shimamura, 1986; Tuokko, Kristjansson, & Miller, 1995), performance has also been found to be affected by age, education, intelligence, and, albeit inconsistently, by gender (e.g., Schmidt, 1996). The declines in performance with age are well documented (e.g., Bolla-Wilson & Bleecker, 1986; Crossen & Wiens, 1994; Gefen, Moar, O'Hanlon, Clark, & Geffen, 1990; Ivnik et al., 1990, 1992; Query & Megran, 1983; Savage & Gouvier, 1992). The effect of education, IQ, and gender on RAVLT performance has been mixed, but it is generally accepted education and IQ affect performance and, when there is a difference, women perform slightly better than men (e.g., Schmidt, 1996; Uchiyama et al., 1995).

Normative data for the RAVLT were compiled for healthy individuals and clinical samples by Schmidt (1996), Lezak (1995), and Geffen et al. (1990). Performance data for individuals with known neurological injury have been extensively evaluated among numerous patient samples (see Schmidt, 1996, for review). In general, these data indicated patients' with neurological impairment tend to perform worse than individuals without known neurological dysfunction. These data are less clear for patients with psychiatric illness, but performance on the RAVLT has shown to be generally insensitive to depression and anxiety with scores that do not differ meaningfully from healthy peers (e.g., Davidoff et al., 1990; Query & Megran, 1983; Schmidt, 1996). Although considerable data are available for the RAVLT in clinical samples (Estevez-Gonzalez, Kulisevsky, Boltes, Otermin, & Garcia-Sanchez, 2003; Henry & Crawford, 2004; Hoffman & Schmitt, 2004; Kurylo, Temple, Elliot, & Crawford, 2001; Smard, Rouleau, Brosseau, Laframboise, & Bojanowsky, 2003; Stefanova, Kostic, Ziropadja, Markovic, & Ocic, 2002; Tong, Yip, Lee, & Li, 2002; Vakil, Kahanshimon, & Mosche, 2000; Millis et al., 2001) these studies have been limited by small sample sizes and performance data have not been stratified by age (e.g., Guilmette & Rasile, 1995; also see Schmidt, 1996, for review). Moreover, data documenting the predictive value of the RAVLT in discriminating individuals with neurological injury from healthy controls have not consistently been provided (Neuropsychology Assessment Panel, 1996; but see Ivnik et al., 2000, 2001; Powell et al., 1991). The purpose of this study was two-fold: (1) present performance data for a large clinical sample by age, which extends available RAVLT performance data and allow clinicians to compare a patient with suspected neurological injury to healthy adults as well as discrete patient samples and (2) present classification characteristics for the RAVLT for selected patient samples as compared to a psychiatric comparison group based on cut-off scores reflecting performance 1.5 S.D. below healthy age-matched peers derived from the meta-norms complied by Schmidt (1996). The SENS, SPEC, PPP, NPP, and HR of the RAVLT were based on cut-off scores that were at the fifth percentile for age-matched healthy peers.

1. Methods

The sample was derived from a dataset from a large Midwestern tertiary medical center. The dataset consists of patients referred for a neuropsychological evaluation from a number of sources including physicians within and outside

of the medical center. Patients included in the study were those 18 years old and older whom completed the RAVLT and were not involved in litigation at the time of the evaluation. The administration of the RAVLT did not vary during the study period, and followed the procedure recommended by Schmidt (1996). Participants were administered trials I–V, trial B, and then trial VI using a presentation rate of one word per second. A free recall trial (trial VII) was administered 30 min after trial VI was completed. A recognition task was administered following the 30-min free recall trial.

1.1. RAVLT performance data

A total of 826 protocols were classified as either having neurological dysfunction or as a psychiatric comparison group (PSYCH) based on medical history, biometric data, neurological diagnosis, and review by a board certified neuropsychologist. Neurological dysfunction was documented with evidence of neurological disease (e.g., MRI, CT, abnormal neurological exam), which was rated categorically as either no abnormality, equivocal findings, or specific neurological abnormality (e.g., structural lesion, pathognomonic sign). There were 122 protocols without known neurological dysfunction and no evidence of structural abnormality on MRI or CT that were identified and these patient protocols were retained as the PSYCH group. The PSYCH group consisted of individuals referred for neuropsychological evaluation who did not have evidence of neurological dysfunction and were diagnosed by a board certified neuropsychologist as meeting diagnostic criteria for a psychiatric diagnosis of depression and/or somatoform disorder. Of the remaining 704 individuals, 312 were excluded due to one or more the following: diagnosis conflicted between physician and neuropsychologist, scored below criteria on indices of effort (see below), patient was involved in litigation, and/or absent or equivocal biometric evidence of brain injury. The resulting 392 cases included: 68 patients with closed head injury (TBI); 57 participants with neoplasms; 47 subjects with strokes; 158 patients with probable Alzheimer's disease; and 28 left temporal lobe epilepsy (LTLE) cases and 34 persons with right temporal lobe epilepsy (RTLE).

The TBI group included patients with documented closed head traumatic brain injuries (average time since injury = 14.3 months). The patients with diagnoses of neoplasm completed testing prior to surgical resection of the tumor as part of a presurgical evaluation. Patients with diagnoses of stroke ranged in age from 20 to 80 years old (mean = 54) and were evaluated an average of 34.2 months after stroke. The patients diagnosed with medically intractable epilepsy had completed video-EEG monitoring prior to the presurgical neuropsychological evaluation and had confirmed EEG seizure activity of either left or right hemisphere onset. Participants scoring below 8/15 correct on the Rey 15 item memory test (Rey, 1964) and/or scoring below 10/21 correct on the forced choice recognition trial of the 21-item memory test (Iverson, Franzen, & McCracken, 1991) were excluded (Iverson & Franzen, 1996; Lee, Loring, & Martin, 1996). Table 1 displays the age, gender, education, and average Wechsler Adult Intelligence Scale—Revised (WAIS-R;

Table 1
Demographic data by diagnostic group ($n = 514$)

Variables	TBI ($n = 68$)	Neoplasm ($n = 57$)	Stroke ($n = 47$)	Dementia ($n = 158$)	Epilepsy, left ($n = 28$)	Epilepsy, right ($n = 34$)	Psychiatric comparison ($n = 122$)
Age							
16–35	27 (39.7)	7 (12.3)	4 (8.5)	0 (0.0)	16 (57.1)	11 (32.4)	20 (16.4)
36–59	32 (47.0)	38 (66.7)	21 (44.7)	12 (7.6)	12 (42.9)	21 (61.8)	63 (51.6)
60–88	9 (13.2)	12 (21)	22 (46.8)	146 (92.4)	0 (0.0)	2 (0.6)	39 (32.0)
Gender							
Females	21 (30.8)	40 (70.2)	25 (53.2)	115 (72.8)	14 (50.0)	19 (55.9)	63 (48.3)
Males	47 (69.2)	17 (29.8)	22 (46.8)	43 (27.2)	14 (50.0)	15 (44.1)	59 (51.7)
Education							
<12	13 (19.1)	6 (10.5)	8 (17.0)	41 (25.9)	6 (21.4)	8 (23.5)	15 (12.3)
12	21 (30.9)	24 (42.1)	17 (36.2)	61 (38.6)	17 (60.7)	16 (47.1)	32 (26.2)
13.15	23 (33.8)	15 (26.3)	8 (17.0)	26 (16.5)	3 (10.7)	9 (26.5)	35 (28.7)
>16	11 (16.2)	12 (21.1)	14 (29.8)	30 (19.0)	2 (7.1)	1 (2.9)	40 (32.8)
WAIS-R FSIQ	94.2 (14.0) a	96.5 (14.0) a	99.1 (14.2) ab	90.5 (11.5) ac	86.3 (12.5) c	85.9 (11.7) c	103.6 (14.3) b

Note. Values represent frequencies with percentages in parentheses, unless otherwise noted. WAIS-R FSIQ values are means with standard deviations in parentheses. FSIQ and years of education scores were not available for all participants.

Table 2
Participants with traumatic brain injury: RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 68$)

Age	*N*	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	27									
Mean		5.19	7.63	8.93	9.22	10.00	5.15	8.30	7.44	12.59
S.D.		2.13	3.15	3.13	3.68	3.49	1.83	4.57	4.86	2.55
36–59	32									
Mean		5.75	7.81	9.12	10.06	10.62	5.50	7.94	7.50	12.41
S.D.		1.67	2.16	2.65	2.83	2.97	2.2	4.05	3.99	2.77
60–88	9									
Mean		5.56	6.89	8.44	10.00	10.00	4.22	6.56	5.67	13.44
S.D.		1.74	2.93	2.56	2.60	3.12	1.39	3.09	4.00	1.67

Note. S.D., standard deviation; List B, RAVLT List B recall trial; Trail VI, RAVLT List A short-delay recall trial; Recog, recognition trial for RAVLT (trial VIII).

Wechsler, 1981) FSIQ for each diagnostic group and the PSYCH group. Patients were separated into three age groups (e.g., 16–35, 36–59, and >60) based on previous reports of RAVLT performance (see Schmidt, 1996).

1.2. RAVLT classification characteristics

Each participant's performance on the RAVLT across trials I–VII was transformed into a *z*-score based on the age-matched meta-normative performance data complied by Schmidt (1996). Participants with scores 1.5 S.D. and below Schmidt (1996) meta-norms age-matched healthy peers were classified as neurologically impaired (i.e., score ≤ -1.5 S.D.).

Once *z*-scores were generated for each patient, the classification of each patient using only the RAVLT was compared to the patients' actual neurological biometric data to distinguish the participants with known neurological dysfunction from the PSYCH comparison group. The percent of cases accurately classified as falling below the 1.5 S.D. cut-off using Schmidt (1996) meta-norms were compared to those inaccurately classified. The SENS, SPEC, PPP, NPP, and HR was generated for RAVLT trials I, V, VI, and VII. The prevalence rate was based on the number of individuals in each diagnostic group and the PSYCH comparison group. Hence, the prevalence rate varied from 18.7% (LTLE) to 56.4% (dementia).

2. Results

2.1. RAVLT performance data

The means and S.D. for List A trials (I–V), List B, List A Immediate recall (trial VI), List A 30-min delayed recall (trial VII), and recognition (total correct) trial is displayed in Tables 2–7 for each patient group (i.e., TBI, neoplasms, stroke, probable dementia of the Alzheimer's type, LTLE, and RTLE) by age group (16–35, 36–59, and 60–88). Table 8 presents the means and S.D. for List A trials (I–V), List B, List A Immediate recall (trial VI), List A 30-min delayed recall (trial VII), and recognition (total correct) trial for the combined patient sample by age group. The PSYCH group's performance on the RAVLT is presented in Table 9 by age group for trials I–V, trial B, trial VI, trial VII, and recognition memory.

2.2. Group differences in RAVLT performance

To reflect common neuropsychological clinical practice in differentiating neurological dysfunction from psychiatric dysfunction, a comparison of RAVLT raw scores across groups was completed. There was a significant effect for age and education on RAVLT performance for each of the neurological groups and for the PSYCH group ($p < .01$). Using age and education as covariates, a MANCOVA was used to assess for differences between the PSYCH group and the

Table 3
Neoplasms: RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 57$)

Age	*N*	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	7									
Mean		6.57	8.14	8.86	9.57	10.00	4.29	7.86	5.71	10.86
S.D.		1.27	2.67	4.45	3.91	4.55	2.06	5.58	5.79	3.63
36–59	38									
Mean		5.37	7.47	8.79	9.79	10.00	4.66	7.95	7.34	12.36
S.D.		2.41	3.05	2.86	2.90	3.06	2.15	3.80	4.36	2.60
60–88	12									
Mean		4.17	6.33	6.83	7.92	8.08	4.17	6.75	6.58	13.50
S.D.		2.95	3.20	3.64	3.94	4.78	1.90	4.62	4.96	1.68

Note. Data reported for patients with a neoplasm above the Foramen magnum whom completed RAVLT prior to surgical resection and/or radiation treatment of the identified tumor.

Table 4
Stroke: RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 47$)

Age	*N*	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	4									
Mean		5.25	7.35	8.25	8.25	8.75	5.50	5.75	5.75	13.25
S.D.		1.50	2.22	3.50	2.99	4.11	2.38	4.92	5.12	2.22
36–59	21									
Mean		4.48	7.00	8.14	9.48	10.29	4.33	6.95	6.71	11.71
S.D.		1.54	2.68	2.74	3.33	2.94	1.53	4.08	3.90	3.52
60–88	22									
Mean		4.77	6.45	7.05	7.59	7.95	3.91	5.86	5.32	12.36
S.D.		1.93	2.67	2.65	3.65	4.41	1.95	3.92	4.50	2.92

groups with diagnosed neurological dysfunction. There was an overall effect of diagnosis on the number of words learned after controlling for age and education effects [$F(9, 466) = 7.29$, $p < .001$]. Subsequent univariate ANOVA for group differences on RAVLT trials I through VII were all significant ($p < .001$). Tukey post hoc analyses were used to assess for group differences, and due to the number of comparisons, a more conservative alpha level was selected

Table 5
Probable dementia of the Alzheimer's type: RAVLT mean (S.D.) number of words recalled by trial and age group ($n = 158$)

Age	*N*	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
36–59	12									
Mean		3.50	5.33	7.33	8.00	8.17	4.08	5.75	4.17	11.08
S.D.		1.24	1.61	2.15	2.59	2.25	1.98	2.67	3.41	4.40
60–88	146									
Mean		2.97	4.25	4.96	5.22	5.75	3.01	2.60	1.78	10.38
S.D.		1.49	1.68	2.10	2.24	2.57	1.62	2.59	2.73	3.80
Total	158									
Mean		3.01	4.34	5.14	5.43	5.93	3.09	2.84	1.96	10.43
S.D.		1.48	1.70	2.19	2.38	2.62	1.67	2.72	2.84	3.84

Note. Total = combined mean and S.D. for 36–59- and 60–88-year-old age bands.

Table 6
Epilepsy: left seizure focus RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 28$)

Age	N	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	16									
Mean		5.06	7.44	9.06	9.56	10.19	4.75	7.69	6.00	11.95
S.D.		1.69	2.63	2.99	2.53	3.23	1.95	3.40	4.32	2.12
36–59	12									
Mean		6.42	7.83	8.92	9.08	10.92	5.33	7.17	6.83	12.51
S.D.		1.83	1.59	1.98	1.93	1.62	1.23	2.48	3.24	2.01

Note. Data reported for patients with identified epilepsy whom completed RAVLT prior to surgical resection of suspected left temporal seizure focus.

Table 7
Epilepsy: right seizure focus RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 34$)

Age	N	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	11									
Mean		6.36	9.09	9.82	11.18	11.45	5.55	8.36	8.64	13.07
S.D.		1.36	2.12	2.14	1.78	2.30	2.25	2.77	2.54	1.79
36–59	21									
Mean		5.95	7.67	9.05	8.67	10.95	4.43	8.57	8.19	13.47
S.D.		1.88	2.42	3.34	3.41	2.97	1.75	3.71	4.13	1.81

Note. Data reported for patients with identified epilepsy whom completed RAVLT prior to surgical resection of suspected right temporal seizure focus. The two patients older than 59 were dropped due to limited data.

($p < .001$). Overall, patients with dementia scored significantly worse than the PSYCH comparison group as well as the TBI, neoplasm, LTLE, and RTLE on every RAVLT trial ($p < .001$).

2.3. RAVLT classification characteristics

2.3.1. Distinguishing neurological dysfunction from PSYCH comparison group

The SENS, SPEC, PPP, NPP, and HR of the RAVLT trials I, V, VI, and VII are presented by diagnostic group in Table 10. The SENS values of the RAVLT trials varied from 10.3 to 93.6%, in which trial V was the most sensitive. The SPEC values for the RAVLT trials ranged from 43.4 to 90.2%. The PPP rates varied from 13.8 to 89.3%. The highest

Table 8
Participants with known neurological dysfunction: RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 392$)

Age	N	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	65									
Mean		5.47	7.90	9.09	9.89	10.51	5.04	8.39	7.62	12.70
S.D.		1.88	2.63	2.94	3.18	5.04	2.01	3.98	4.39	2.57
36–59	136									
Mean		5.40	7.52	8.83	9.52	10.37	4.85	7.81	7.28	12.47
S.D.		2.04	2.61	2.84	3.03	2.97	1.95	3.81	4.09	2.86
60–88	191									
Mean		3.44	4.84	5.54	5.99	6.43	3.30	3.47	2.74	11.0
S.D.		1.85	2.22	2.52	2.92	3.23	1.73	3.35	3.65	3.64

Note. List B, RAVLT List B recall trial; Trail VI, RAVLT List A short-delay recall trial; Recog, recognition trial for RAVLT.

Table 9
Psychiatric comparison group: RAVLT mean (S.D.) number of words recalled by trial and age groups ($n = 122$)

Age	*N*	Trials					30 min			
		I	II	III	IV	V	List B	Trial VI	Delay VII	Recog VIII
16–35	20									
Mean		7.21	9.93	11.64	12.93	13.29	6.14	11.71	12.00	14.25
S.D.		2.16	2.34	2.13	2.46	2.02	2.41	2.92	2.83	1.62
36–59	63									
Mean		6.26	9.64	11.51	12.45	12.58	5.79	11.23	11.21	13.37
S.D.		1.60	2.40	2.12	1.80	1.88	1.66	2.63	3.01	2.22
60–88	39									
Mean		5.86	8.76	10.08	10.56	11.29	5.27	9.35	9.11	13.00
S.D.		1.84	2.62	2.64	2.95	2.81	1.94	3.40	3.50	2.32

Table 10
Classification statistics of the RAVLT by group for trials I, V, VI, and VII[1]

Group	SENS	SPEC	PPP	NPP	HR
TBI					
Trial I	10.3	90.2	36.8	64.3	61.6
Trial V	73.5	43.4	42.0	74.6	54.2
Trial VI	39.7	81.1	54.0	70.7	66.3
Trial VII	44.1	82.8	58.8	72.7	68.9
Neoplasm					
Trial I	26.3	90.2	55.6	72.4	69.8
Trial V	80.7	43.4	40.0	82.8	55.3
Trial VI	43.9	81.1	52.1	75.6	69.4
Trial VII	43.9	82.8	54.3	75.9	70.4
Stroke					
Trial I	19.1	90.2	42.9	74.3	70.4
Trial V	93.6	43.4	38.9	94.6	57.4
Trial VI	51.1	81.1	51.1	81.1	72.8
Trial VII	48.9	82.8	52.3	80.8	73.4
Dementia of AD type					
Trial I	63.3	90.2	89.3	65.5	75.0
Trial V	80.4	43.4	64.8	63.1	64.3
Trial VI	78.5	81.1	84.4	74.4	79.6
Trial VII	82.9	82.8	86.2	78.9	82.9
Presurgical epilepsy—left seizure focus					
Trial I	42.9	90.2	50.0	87.3	81.3
Trial V	39.3	43.4	13.8	75.7	42.7
Trial VI	57.1	81.1	41.0	89.2	76.7
Trial VII	71.4	82.8	48.8	92.7	80.7
Presurgical epilepsy—right seizure focus					
Trial I	11.8	90.2	25.0	78.6	73.1
Trial V	38.2	43.4	15.9	71.6	42.3
Trial VI	41.2	81.1	37.8	83.2	72.4
Trial VII	32.4	82.8	34.4	81.5	71.8

Note. All values are percentages; SENS, sensitivity; SPEC, specificity; PPP, positive predictive power; NPP, negative predictive power; HR, hit rate (total correctly classified); [1]Cut-off scores based on Schmidt (1996) age-matched meta-norms with RAVLT performances more than 1.5 S.D. classified as presence of neurological dysfunction.

PPP was achieved for patients with probable Alzheimer's disease. The best overall hit rate (total number of correct positives and negatives) was found for the probable Alzheimer's disease group trial VII, yielding a HR of 82.9%. The PPP rate of the stroke group ranged from 38.9 to 52.3%, and varied from 40.0 to 55.6% for the Neoplasm group. The HR of the LTLE group varied from 42.7 to 81.3% and varied from 42.3 to 73.1% for the RTLE group.

3. Discussion

This study evaluated the performance of selected patient groups on the RAVLT, and compared their performances to a psychiatric comparison group. Classification rates were evaluated using the meta-norms for healthy subjects compiled by Schmidt (1996). The diagnostic utility of a test is best evaluated when performance of persons with suspected brain dysfunction can be compared to subjects without neurological injury and to persons with known neurological dysfunction (Nunnally, 1978; Retzlaff & Gibertini, 2000). RAVLT scores on trials I–V, List B, trials VI and VII, and recognition were presented for patients with closed head TBI, neoplasms (presurgery), stroke (cerebral), and epilepsy (presurgical LTLE and RTLE). Consistent with past research, age and education were related to RAVLT performance. As expected, the groups with neurological dysfunction tended to recall fewer words on trials V, VI, and VII than did the age-matched psychiatric comparison subjects ($p < .01$). The average RAVLT scores for the patient groups was less than their age-matched healthy peers using the meta-norms of Schmidt (1996).

Overall, patients diagnosed with probable dementia of the Alzheimer's type recalled fewer words than the PSYCH comparison group using the healthy age-matched normative data provided by Schmidt (1996) meta-norms across all RAVLT trials. While the PPP of the RAVLT trials varied substantially, it was generally highest for the RAVLT trial VII. Interestingly, the SENS of RAVLT trial 1 was good while trial 5 was poor in this sample of patients diagnosed with psychiatric disorders and no neurological dysfunction (i.e., few PSYCH comparison group members scored 1.5 S.D. or below their age-matched peers using Schmidt, 1996, meta-norm data). This finding was not expected based on previous research suggesting RAVLT performance may be only reduced on trial 1 (i.e., Query & Megran, 1983; Schmidt, 1996). However, not all studies have found depression unrelated to RAVLT performance (e.g., Unkenstein & Bowden, 1991). As expected, however, the SPEC improved for the immediate and delayed recall trials of the RAVLT.

The PPP of the RAVLT for patients' in the TBI, stroke, neoplasm, LTLE, and RTLE groups were below the PPP obtained for the probable dementia of the Alzheimer's group. The reduced capacity of the RAVLT to distinguish the presence of neurological dysfunction from a PSYCH comparison group likely represents the diverse kinds of neuropsychological dysfunction that may occur with TBI (e.g., Lezak, 1995; Lucas, 1998; Williamson, Scott, & Adams, 1996), stroke (e.g., Reitan & Wolfson, 1993; Weinstein & Swenson, 1998), neoplasms (Reitan & Wolfson, 1993), and epilepsy (e.g., Rankin, Adams, & Jones, 1996; Snyder, 1998). There was no effort to select patients with brain injuries suspected to adversely affect memory specifically; rather the participants were selected based on documentation of neurological dysfunction. Nevertheless, the classification rates of the RAVLT for the TBI, stroke, neoplasm, and presurgical epilepsy patients were consistent with other research (e.g., see Schmidt, 1996, for review and also Ivnik et al., 2000, 2001). Results for the probable Alzheimer's dementia group generally mirrored other studies examining diagnostic accuracy of list learning tests, revealing SENS and SPEC rates of 62–98% (e.g., Chen et al., 2000; Ivnik et al., 2000; Petersen et al., 1999, 2001; Salmon et al., 2002).

Although a direct comparison between the diagnostic efficacy of brain imaging and RAVLT performance was not possible with the current data, the classification characteristics for structural neuroimaging (i.e., MRI) have reported SENS rates ranging from 77 to 92% and SPEC varying from 49 to 95%, although rates varying from 45 to 100% have been reported (e.g., Gosche, Mortimer, Smith, Markesbery, & Snowdon, 2002; Jack et al., 1997, 1999; Knopman et al., 2001; Laakso et al., 1998; O'Brien et al., 1997). For example, Laakso et al. (1998) evaluated the hippocampal volumes of 55 patients with probable AD compared to 43 patients meeting diagnosis for age associated memory impairment, 42 cognitively healthy elderly controls, and 20 cognitively normal younger control subjects. With a 34% prevalence rate of probable Alzheimer's disease, MRI volumetric evaluation yielded an overall hit rate of 92%. Another study by Gosche et al. (2002) in a sample of 56 older adults compared the diagnosis of dementia based on MRI hippocampal volumes to neuropathic confirmed Alzheimer's disease pathology, and found a PPP of 90%, a NPP of 77%, a SENS of 82%, and a SPEC of 87%, when the prevalence rate for neuropathologically confirmed Alzheimer's pathology was 58.9%. Given this study's prevalence rate of probable dementia of the Alzheimer's type (56.4%), the obtained HR rate would argue in support of the RAVLT as a screening measure for probable dementia of the Alzheimer's type.

Several limitations of this study should be highlighted. The average FSIQ of the patients with known brain injury was less than that of the PSYCH comparison group. This finding, however, was expected (as there was neurological dysfunction), and is not thought to account for the poorer performance of the neurologically impaired groups. Indeed, estimates of premorbid FSIQ using Vanderploeg & Schinka (1995) BEST-3 algorithms found no meaningful differences between patient groups (Cohen's $d < 0.3$). A second limitation is the small sample sizes in some of the cells. For example, older TBI and epilepsy patients as well as younger stroke patients are less well represented in the current sample, so the reported classification characteristics might not accurately capture these patient groups. Of particular note, is that the biomedical information was limited to reports from the referring physicians and did not include the actual MRI/CT films, EEG recordings, etc. Similarly, in the case of the probable Alzheimer's disease patient group, no pathologic confirmation of Alzheimer's disease was available on these patients and the diagnosis was based on biomedical, neurological, and neuropsychological data. Thus, the determination of dementia was not independent of the patient's performance on the RAVLT. All other patient groups were classified as having neurological dysfunction based on biomedical and a neurological examination by a physician, which was independent of their performance on neuropsychological tests, however this may be seen as a limitation as the reliability of physician diagnoses were unknown.

Despite these limitations, these data provide useful information for clinicians using the RAVLT. These data indicate word list learning was generally impaired for patients with known neurological dysfunction compared to age-matched peers using the meta-norms of Schmidt (1996). In addition, the performance of patients with neurological dysfunction differed from that of a psychiatric comparison group. However, poor performance alone on the RAVLT is not sufficient for one to diagnose the presence of neurological dysfunction. Indeed, interpretation should be in context of a comprehensive neuropsychological evaluation along with medical, psychiatric, and historical data. Independent validation of RAVLT classification characteristics in more specific neurological groups is needed, in which the clinical samples are more specifically described (e.g., severity of TBI, location and extent of stroke, type of dementia).

References

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Bigler, E. D., Rosa, L., Schultz, F., Hall, S., & Harris, J. (1989). Rey Auditory Verbal Learning and Rey-Osterrieth Complex Figure design performance in Alzheimer's disease and closed head injury. *Journal of Clinical Psychology*, *45*, 277–280.

Bolla-Wilson, K., & Bleecker, M. L. (1986). Influence of verbal intelligence, sex, age, and education on the Rey Auditory Verbal Learning Test. *Developmental Neuropsychology*, *2*, 203–211.

Butters, N., Wolfe, J., Martone, M., Granholm, E., & Cermak, L. S. (1985). Memory disorders associated with Huntington's disease: Verbal recall, verbal recognition, and procedural memory. *Neuropsychologia*, *23*, 729–743.

Chen, P., Ratcliff, G., Belle, S. H., Cauley, J. A., DeKosky, S. T., & Ganguli, M. (2000). Cognitive tests that best discriminate between presymptomatic AD and those who remain nondemented. *Neurology*, *55*, 1847–1853.

Crossen, J.R., & Wiens, A.N. (1994). Comparison of the Auditory-Verbal Learning Test (AVLT) and California Verbal Learning Test (CVLT) in a sample of normal subjects. *Journal of Clinical and Experimental Neuropsychology*, *16*, 190–194. Erratum. (1994). *Journal of Clinical and Experimental Neuropsychology*, *16*, 649.

Davidoff, G., Roth, E., Thomas, P., Doljanac, R., Kijkers, M., Berent, S., et al. (1990). Depression and neuropsychological test performance in acute spinal cord injury: Lack of correlation. *Archives of Clinical Neuropsychology*, *5*, 77–88.

Drebing, C. E., Van Gorp, W. G., Stuck, A. E., Mitrushina, M., & Beck, J. (1994). Early detection of cognitive decline in higher cognitively functioning older adults: Sensitivity and specificity of a neuropsychological screening battery. *Neuropsychology*, *8*, 31–37.

Estevez-Gonzalez, A., Kulisevsky, J., Boltes, A., Otermin, P., & Garcia-Sanchez, C. (2003). Rey Verbal Learning Test is a useful tool for differential diagnosis in the pre-clinical phase of Alzheimer's disease: Comparison with mild cognitive impairment in normal aging. *International Journal of Geriatric Psychiatry*, *18*, 1021–1028.

Geffen, G., Moar, K. J., O'Hanlon, A. P., Clark, C. R., & Geffen, L. B. (1990). Performance measures of 16- to 86-year-old males and females on the auditory verbal learning test. *The Clinical Neuropsychologist*, *4*, 45–63.

Gosche, K. M., Mortimer, J. A., Smith, C. D., Markesbery, W. R., & Snowdon, D. A. (2002). Hippocampal volume as an index of Alzheimer neuropathology: Findings from the Nun Study. *Neurology*, *58*, 1476–1482.

Guilmette, T. J., & Rasile, D. (1995). Sensitivity, specificity, and diagnostic accuracy of three verbal memory measures in the assessment of mild brain injury. *Neuropsychology*, *9*(3), 338–344.

Henry, J. D., & Crawford, J. R. (2004). A meta-analytic review of verbal fluency performance in patients with traumatic brain injury. *Neuropsychology*, *18*(4), 621–628.

Hoffman, M., & Schmitt, F. (2004). Cognitive impairment in isolated subtentorial stroke. *Acta Neurologica Scandavica*, *109*, 14–24.

Iverson, G. L., & Franzen, M. D. (1996). Using multiple objective memory procedures to detect simulated malingering. *Journal of Clinical and Experimental Neuropsychology*, *18*, 38–51.

Iverson, G. L., Franzen, M. D., & McCracken, L. M. (1991). Evaluation of an objective assessment technique for the detection of malingered memory deficits. *Law and Human Behavior*, *15*, 667–676.

Ivnik, R. J., Malec, J. F., Tangalos, E. G., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1990). The Auditory-Verbal Learning Test (AVLT): Norms for ages 55 years and older. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *2*, 304–312.

Ivnik, R. J., Malec, J. F., Tangalos, E. G., Petersen, R. C., Kokmen, E., & Kurland, L. T. (1992). Mayo's older Americans normative studies: Updated AVLT norms for ages 56 to 97. *The Clinical Neuropsychologist*, *6*(Suppl.), 83–104.

Ivnik, R. J., Smith, G. E., Cerhan, J. H., Boeve, B. F., Tangalos, E. G., & Petersen, R. C. (2001). Understanding the diagnostic capabilities of cognitive tests. *The Clinical Neuropsychologist*, *15*, 114–124.

Ivnik, R. J., Smith, G. E., Malec, J. F., Kokmen, E., & Tangalos, E. G. (1994). Mayo Cognitive Factor Scales: Distinguishing normal and clinical samples by profile variability. *Neuropsychology*, *8*, 203–209.

Ivnik, R. J., Smith, G. E., Petersen, R. C., Boeve, B. F., Kokmen, E., & Tangalos, E. G. (2000). Diagnostic accuracy of four approaches to interpreting neuropsychological test data. *Neuropsychology*, *14*, 163–177.

Jack, C. R., Jr., Petersen, R. C., Xu, Y. C., O'Brien, P. C., Smith, G. E., Ivnik, R. J., et al. (1999). Prediction of AD with MRI-based hippocampal volume in mild cognitive impairment. *Neurology*, *52*, 1397–1403.

Jack, C. R., Petersen, R. C., Xu, Y. C., Waring, S. C., O'Brien, P. C., Tangalos, E. G., et al. (1997). Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease. *Neurology*, *49*, 786–794.

Knopman, D. S., DeKosky, S. T., Cummings, J. L., Chui, H., Corey-Bloom, J., Relkin, N., et al. (2001). Practice parameter: Diagnosis of dementia (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*, 1143–1153.

Kurylo, M., Temple, R. O., Elliot, T. R., & Crawford, D. (2001). Rey Auditory Verbal Learning Test (AVLT) performance in individuals with recent onset spinal cord injury. *Rehabilitation Psychology*, *46*(3), 247–261.

Laakso, M. P., Soininen, H. K., Partanen, K., Lehtovirta, M., Hallikainen, M., Hänninen, T., et al. (1998). MRI of the hippocampus in Alzheimer's disease: Sensitivity, specificity, and analysis of the incorrectly classified subjects. *Neurobiology of Aging*, *19*, 23–31.

Lee, G. P., Loring, D. W., & Martin, R. C. (1992). Rey's 15-Item Visual Memory Test for the detection of malingering: Normative observations on patients with neurological disorders. *Psychological Assessment*, *4*, 43–46.

Lezak, M. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lucas, J. A. (1998). Traumatic brain injury and postconcussive syndrome. In P. J. Snyder, & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment*. Washington, DC: American Psychological Association.

Millis, S. R., Rosenthal, M., Novack, T. A., Sherer, M., Nick, T. G., Kreutzer, J. S., et al. (2001). Long term neuropsychological outcome after traumatic brain injury. *Journal of Head Trauma Rehabilitation*, *16*(4), 343–355.

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.

Neuropsychology Assessment Panel. (1996). Assessment: Neuropsychological testing of adults. Considerations for neurologists. *Neurology*, *34*, 592–599.

Nunnally, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.

O'Brien, J. T., Desmond, P., Ames, D., Schweitzer, I., Chiu, E., & Tress, B. (1997). Temporal lobe magnetic resonance imaging can differentiate Alzheimer's disease from normal aging, depression, vascular dementia, and other causes of cognitive impairment. *Psychological Medicine*, *27*, 1267–1275.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokman, E. (1999). Mild cognitive impairment: Clinical characterization and outcome. *Archives of Neurology*, *56*, 303–308.

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidence-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology*, *56*, 1133–1142.

Powell, J. G., Cripe, L. I., & Dodrill, C. B. (1991). Assessment of brain impairment with the Rey Auditory Verbal Learning Test: A comparison with other neuropsychological measures. *Archives of Clinical Neuropsychology*, *6*, 241–249.

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

Rankin, E. J., Adams, R. L., & Jones, H. E. (1996). Epilepsy and nonepileptic attack disorder. In R. L. Adams, O. A. Parsons, J. L. Culbertson, & S. J. Nixon (Eds.), *Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders*. Washington, DC: American Psychological Association.

Reitan, R. M., & Wolfson, D. (1993). *The Halstead-Reitan neuropsychological test battery: Theory and clinical interpretation* (2nd ed.). Tucson, Arizona: Neuropsychology Press.

Retzlaff, P. D., & Gibertini, M. (2000). Neuropsychometric issues and problems. In R. D. Vanderploeg (Ed.), *Clinician's guide to neuropsychological assessment* (2nd ed., pp. 277–299). Mahwah, NJ: Lawrence Erlbaum Associates.

Rey, A. (1941). L'examen psychologique dans les cas d'encéphalopathie traumatique. *Archives de Psychologie*, *28*, 21.

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

Salmon, D. P., Thomas, R. G., Pay, M. M., Booth, A., Hofstetter, C. R., Thal, L. J., et al. (2002). Alzheimer's disease can be accurately diagnosed in very mildly impaired individuals. *Neurology*, *59*, 1022–1028.

Savage, R. M., & Gouvier, W. D. (1992). Rey Auditory Verbal Learning Test: The effects of age and gender, and norms for delayed recall and story recognition trials. *Archives of Clinical Neuropsychology*, *7*, 407–414.

Schmidt, M. (1996). *Rey Auditory and Verbal Learning Test: A handbook*. Los Angeles, CA: Western Psychological Services.

Smard, S., Rouleau, I., Brosseau, J., Laframboise, M., & Bojanowsky, M. (2003). Impact of executive dysfunctions on episodic memory abilities in patients with ruptured aneurysm of the anterior communicating artery. *Brain and Cognition*, *53*, 354–358.

Snyder, P. J. (1998). Epilepsy. In P. J. Snyder, & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment*. Washington, DC: American Psychological Association.

Squire, L. R., & Shimamura, A. P. (1986). Characterizing amnesic patients for neurobehavioral study. *Behavioral Neuroscience*, *100*, 866–877.

Stefanova, E. D., Kostic, V. S., Ziropadja, L., Markovic, M., & Ocic, G. (2002). Serial position learn effects in patients with aneurysm of the anterior communicating artery. *Journal of Clinical and Experimental Neuropsyology*, *24*(5), 687–694.

Taylor, E. M. (1959). *Psychological appraisal of children with cerebral defects*. Cambridge, MA: Harvard University Press, pp. 423–428.

Tong, B. S. K., Yip, J. T. H., Lee, T. M. C., & Lee, L. S. W. (2002). Frontal fluency and memory functioning among multiple sclerosis patients in Hong Kong. *Brain Injury*, *16*(11), 987–995.

Tuokko, H., Kristjansson, E., & Miller, J. (1995). Neuropsychological detection of dementia: An overview of the neuropsychological component of the Canadian study of health and aging. *Journal of Clinical and Experimental Neuropsychology*, *17*, 352–373.

Uchiyama, C. L., D'Elia, L. F., Dellinger, A. M., Becker, J. T., Selnes, O. A., Wesch, J. E., et al. (1995). Alternate forms of the auditory-verbal learning test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology*, *10*, 133–145.

Unkenstein, A. E., & Bowden, S. C. (1991). Predicting the course of neuropsychological status in recently abstinent alcoholics: A pilot study. *The Clinical Neuropsychologist*, *5*, 24–32.

Vakil, E., Kahanshimon, H., & Mosche, O. A. (2000). Motor and non-motor sequence learning in patients with Basal Ganglia lesions: The case of serial reaction time (SRT). *Neuropsychologia*, *38*, 1–10.

Vanderploeg, R. D., & Schinka, J. A. (1995). Predicting WAIS-R IQ premorbid ability: Combining subtest performance and demographic variable predictors. *Archives of Clinical Neuropsychology*, *10*, 225–239.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised (WAIS-R)*. San Antonio, TX: The Psychological Corporation.

Weinstein, A., & Swenson, R. A. (1998). Cerebrovascular disease. In P. J. Snyder, & P. D. Nussbaum (Eds.), *Clinical neuropsychology: A pocket handbook for assessment*. Washington, DC: American Psychological Association.

Williamson, D. J. G., Scott, J. G., & Adams, R. L. (1996). Traumatic brain injury. In R. L. Adams, O. A. Parsons, J. L. Culbertson, & S. J. Nixon (Eds.), *Neuropsychology for clinical practice: Etiology, assessment, and treatment of common neurological disorders*. Washington, DC: American Psychological Association.