



## OPEN ACCESS

EDITED BY  
Ashok Kumar,  
University of Florida, United States

REVIEWED BY  
Erin Rutherford Hascup,  
Southern Illinois University School  
of Medicine, United States  
Nadia Justel,  
Consejo Nacional de Investigaciones  
Científicas y Técnicas (CONICET),  
Argentina  
Jason Soble,  
University of Illinois at Chicago,  
United States

\*CORRESPONDENCE  
Meir Plotnik  
meir.plotnik@sheba.health.gov.il

SPECIALTY SECTION  
This article was submitted to  
Neurocognitive Aging and Behavior,  
a section of the journal  
Frontiers in Aging Neuroscience

RECEIVED 28 June 2022  
ACCEPTED 09 August 2022  
PUBLISHED 14 September 2022

CITATION  
Gottlieb A, Doniger GM, Kimel-Naor S,  
Ben-Gal O, Cohen M, Iny H, Beeri MS  
and Plotnik M (2022) Development  
and validation of virtual reality-based  
Rey Auditory Verbal Learning Test.  
*Front. Aging Neurosci.* 14:980093.  
doi: 10.3389/fnagi.2022.980093

COPYRIGHT  
© 2022 Gottlieb, Doniger, Kimel-Naor,  
Ben-Gal, Cohen, Iny, Beeri and Plotnik.  
This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s)  
are credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Development and validation of virtual reality-based Rey Auditory Verbal Learning Test

Amihai Gottlieb<sup>1</sup>, Glen M. Doniger<sup>1,2</sup>, Shani Kimel-Naor<sup>1</sup>,  
Oran Ben-Gal<sup>1</sup>, Maya Cohen<sup>1</sup>, Hila Iny<sup>1</sup>,  
Michal Schnaider Beeri<sup>2,3</sup> and Meir Plotnik<sup>1,4,5\*</sup>

<sup>1</sup>Center of Advanced Technologies in Rehabilitation, Sheba Medical Center, Ramat Gan, Israel, <sup>2</sup>Joseph Sagol Neuroscience Center, Sheba Medical Center, Ramat Gan, Israel, <sup>3</sup>Department of Psychiatry, The Icahn School of Medicine at Mount Sinai, New York, NY, United States, <sup>4</sup>Department of Physiology and Pharmacology, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel, <sup>5</sup>Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

**Objective:** Translations and adaptations of traditional neuropsychological tests to virtual reality (VR) technology bear the potential to increase their ecological validity since the technology enables simulating everyday life conditions in a controlled manner. The current paper describes our translation of a commonly used neuropsychological test to VR, the Rey Auditory Verbal Learning Test (RAVLT). For this aim, we developed a VR adaptation of the RAVLT (VR-RAVLT) which is based on a conversation with a secretary in a virtual office using a fully immersive VR system. To validate the VR-RAVLT, we tested its construct validity, its age-related discriminant validity and its test-retest validity in reference to the original gold standard RAVLT (GS-RAVLT).

**Method:** Seventy-eight participants from different age groups performed the GS-RAVLT and the VR-RAVLT tests in a counterbalanced order in addition to other neuropsychological tests. Construct validity was validated using Pearson's correlations coefficients and serial position effects; discriminant validity was validated using receiver operating characteristic area under the curve values and test-retest reliability was validated using intraclass correlation coefficients.

**Results:** Comparing both RAVLTs' format results indicates that the VR-RAVLT has comparable construct, discriminant and test-retest validities.

**Conclusion:** the novel VR-RAVLT and the GS-RAVLT share similar psychometric properties suggesting that the two tests measure the same cognitive construct. This is an indication of the feasibility of adapting the RAVLT to the VR environment. Future developments will employ this approach for clinical diagnosis and treatment.

## KEYWORDS

memory and learning tests, neuropsychological tests, reliability and validity, Rey Auditory Verbal Learning test, validation study, virtual reality

## Introduction

The term “ecological validity” in regard to neuropsychological tests refers to the relation of the performance on the tests and the performance in everyday life (Chaytor et al., 2006). Although it was assumed that traditional neuropsychological tests are able to tap the same executive functions that are used in day-to-day actions, there is evidence that this is not necessarily the case since poor performance on the tests was not ineludibly reflected in poor performance in everyday life (Wilson, 1993; Manchester et al., 2004; Sbordone, 2008; Bottari et al., 2009). This situation impairs the ability to rely on the results from traditional tests in order to intervene and appropriately support individuals with cognitive impairments.

Translations and adaptations of traditional neuropsychological tests to virtual reality (VR) technology bear the potential to increase their ecological validity since the technology enables simulating everyday life conditions in a controlled manner (Parsons et al., 2017; Gal et al., 2019; Plotnik et al., 2021). However, to date, attempts to translate neuropsychological tests to VR are still rare (Elkind et al., 2001; Pollak et al., 2009; Armstrong et al., 2013; Parsons et al., 2013).

Recently we reported on our successful study to translate the Color Trails Test, which is a cultural fair variant of the widely used Trail Making Test, to a VR environment (Plotnik et al., 2021). The current paper describes our study to translate another commonly used neuropsychological test to VR, the Rey Auditory Verbal Learning Test (RAVLT).

The RAVLT is a widely used verbal memory test (Rey, 1964; Vakil and Blachstein, 1997). Its main components consist of three parts: first, a list of 15-semantically unrelated items (list A) is read five times to the participants, whom are asked to recall as many items as they can remember after each time (trials 1–5). Then, a new interference list is introduced (list B) after which the participants are asked to repeat the original list (trial 6). In the third part the participants are asked to recall the original list (list A) after a 20–40 min interval (trial 7 – delayed recall). There is a large number of potential outcome measures that can be computed from the results of the RAVLT (Bowler, 2013) whereas we decided to use three common measures that were frequently mentioned in the literature (Geffen et al., 1990; Vakil and Blachstein, 1993; Guilmette and Rasile, 1995; Groth-Marnat, 2009; Vakil et al., 2010; Bowler, 2013; Vlahou et al., 2013): acquisition (ACQUISITION; sum of the scores on trials 1–5), retroactive interference (RI; trial 5 vs. the score in trial 6) and retention (RETENTION—trial 5 vs. the score on trial 7).

The RAVLT has shown good construct, test-retest and discriminant validity (Vakil and Blachstein, 1993; Tierney et al., 1996; Estévez-González et al., 2003; Schoenberg et al., 2006; Balthazar et al., 2010; Fichman et al., 2010; Ricci et al., 2012; Bowler, 2013) and high correlation with other neuropsychological tests such as Stroop (McMinn et al., 1988),

Trail Making Test (Callahan and Johnstone, 1994), subscales of the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (McMinn et al., 1988), and the Benton Visual Retention Test (de Sousa Magalhães et al., 2012). Additionally, the RAVLT’s results demonstrate robust serial positions effects of primacy and recency i.e., a tendency to better recall the first few and last few words in the lists (Bernard, 1991; Crockett et al., 1992).

As part of our overarching goal to create VR- based ecological valid adaptation of widely used neuropsychological tests (Plotnik et al., 2021), we created a VR adaptation of the RAVLT (VR-RAVLT). Our objective was to test its construct validity and its age-related discriminant validity in reference to the original gold standard RAVLT (GS-RAVLT) and its test-retest validity, among 78 participants from different age groups. We hypothesized that the tests would show that the new VR-RAVLT similarly evaluates verbal memory as the GS-RAVLT and has comparable, or even better, discriminant and test-retest validities.

## Materials and methods

### Creating the virtual reality adaptation of the Rey Auditory Verbal Learning Test

Virtual reality adaptation of the Rey Auditory Verbal Learning Test is designed to maintain the core features of the original GS-RAVLT (i.e., conserving the core task of reciting as many words out of a 15 words list read to the participant). As first stage, the current format goes beyond the “in clinic” evaluator reading lists of pseudo random words to the participant, by creating a typical real-life scenario during which the need to memorize words (i.e., places—see below) is intended to assist the participant in his/her daily function (i.e., the need to actually visit these places).

The VR-RAVLT places the participant in a virtual office with a virtual personal assistant (avatar) seated behind a desk. The avatar tells the participant a list of 15 places s/he needs to visit on the same day (see **Supplementary Table 1**), and that s/he must recall as many as possible. The avatar informs the participant that as she will be leaving early for the day, she will repeat the list to ensure the participant remembers all the places (i.e., similar to the procedure employed in the GS-RAVLT). List B consists of 15 places that the participant would need to visit on the next day. Participant responses are recorded by a research assistant in a form similar to what is used in the GS-RAVLT. Lists of places were matched for Hebrew word frequency using a 165-million word database (Linzen, 2009). Since the GS-RAVLT consists of alternate forms, two alternate lists were also created for the VR-RAVLT and tested in the study (see **Supplementary Table 1**). VR adaptation was performed in a full immersive VR system



FIGURE 1

The virtual reality adaptation of the Rey Auditory Verbal Learning Test (VR-RAVLT). For RAVLT episodic, verbal memory test—recall of 15 items is tested five times sequentially (learning curve), once again after a new interference list (retroactive interference) and then again after 20 min (delayed recall). In the VR-adapted test, recall of the non-semantically related items is replaced by recall of real-life places-to-go dictated by a virtual personal assistant (avatar).

(HTC-Vive; New Taipei City, Taiwan). **Figure 1** depicts the VR-RAVLT as viewed by the participant. Video demo is provided in the [Supplementary material](#).

## Sample size justification

As we hypothesized that the VR-RAVLT would measure verbal memory similarly to the GS-RAVLT, we calculated the sample size based on bivariate normal model correlation with  $r \geq 0.6$  and alpha of 5% to achieve a power of 80% using Gpower (Erdfelder et al., 1996). It was determined that a total sample size of 15 participants was required. Therefore, we set a minimum of 15 participants in each of the study groups, i.e., young adults (YA), middle aged (MA), and older-adults (OLD).

## Participants

A total of 78 participants comprising of three cohorts of healthy participants were included: (1) YA;  $n = 29$ ; (2) MA;  $n = 29$ ; and (3) Cognitively normal OLD;  $n = 20$ . **Table 1** depicts their demographic data. Inclusion criterion was “healthy men and women ages 18–90” while exclusion criterion was “Having any motor, balance, psychiatric, or cognitive impairment that may affect the ability to understand instructions or perform the tasks required.” Participants’ health status was confirmed by questioning. Two participants were excluded from the study as they failed to meet these criteria (one due to motor impairment and one due to a suspected cognitive impairment). The experimental protocol was approved by the local institutional review board (IRB). All participants signed a written informed

consent prior to entering the study. The research was completed in accordance with Helsinki Declaration.

## Procedure

Participants performed GS-RAVLT and the VR-RAVLT test in a full counter-balanced order. The GS-RAVLT was performed according to its handbook instructions (Schmidt, 1996), i.e., by using two lists of 15-semantically unrelated items (two lists are also presented in the VR-RAVLT, recall the “Creating the virtual reality adaptation of the Rey Auditory Verbal Learning Test” section). In order to test the construct validity (concurrent validity) of the VR-RAVLT, the participants were also asked to perform four additional neuropsychological tests: the Montreal Cognitive Assessment (MOCA) (Nasreddine et al., 2005), the WAIS-R Digit Symbol Test (Wechsler and De Lemos, 1981), a Verbal Fluency Test (Lezak et al., 2004) and the WMS III digit span test (Wechsler, 1997). The MOCA is a highly valid and reliable, 10-min cognitive test that is widely used in numerous clinical and research setting and that taps a wide number of putative cognitive domains (Nasreddine et al., 2005). In the WAIS-R Digit Symbol Test the participants are required to draw as many symbols as they can according to a digit-symbol pairs index during 2 min and the test is considered to be sensitive to age, depression, brain damage and dementia (Wechsler and De Lemos, 1981). In the Verbal fluency test the participants are required to produce as many words as possible from specific categories in a time frame of 2 min and the test is considered as a valid method to detect cognitive impairment and dementia in clinical and research settings (Lezak et al., 2004). Finally, the forward and backward WMS III digit span test is part of the Wechsler Memory Scale (WMS) which requires the participants to repeat numbers in the same order as read aloud by the examiner, and in a reverse order and is considered as a valid measure of working memory which is one of the elements underlying general intelligence (Wechsler, 1997). The four tests were performed according to their handbook instructions and were done sequentially by all of the participants in between the performance of the two RAVLT formats. In order to examine the test-retest reliability, the GS-RAVLT and the VR-RAVLT were performed by the same participants also during a second visit, 2–3 weeks after the first one. As participants awaited the delayed recall component in both RAVLT test versions, they were asked to complete the additional pen and paper neuropsychological tests, which included removing and reapplying the VR goggles in the VR-RAVLT.

## Outcome measures and analyses

For the GS-RAVLT and the VR-RAVLT, the number of correct remembered items were recorded. Summary statistics

TABLE 1 Demographic data.

	Young adults ( <i>n</i> = 29)	Middle aged adults ( <i>n</i> = 29)	Older-adults ( <i>n</i> = 20)	<i>P</i> (ANOVA/Chi square)
Age (mean ± SD, range)	24.1 ± 2.9 (19–29)	57.7 ± 4.0 (43–63)	71.8 ± 5.1 (66–86)	<i>p</i> < 0.0001
Gender (females, percentage)	17, 58.6%	21, 72.4%	13, 65.0%	n.s.
Education years (mean ± SD, range)	13.9 ± 2.0 (12–19)	16.8 ± 3.89 (12–30)	13.9 ± 3.0 (8–18)	<i>p</i> = 0.001

SD, standard deviation; ANOVA, analysis of variance; age was found to be significant different among all groups; the middle-aged group had significantly more education years as opposed to the young adults and the older-adults groups.

(mean ± SD) were computed for each outcome measure (ACQUISITION, RI, and RETENTION). Before conducting the statistical analyses, we performed two *a priori* analyses. First, to verify suitability of parametric statistics, Shapiro–Wilk normality tests were run on the residuals that were calculated using analyses of variance (ANOVA) with a three-level independent group variable that were performed on each of the three outcome measures, per test. Of the six normality tests, five indicated non-normal distributions (Shapiro–Wilk statistic  $\leq 0.91$ ;  $p \leq 0.007$ ). Thus, we used non-parametric Kruskal–Wallis tests to assess effects of group (YA, MA, and OLD) within each test format and Wilcoxon signed rank tests to assess the effect of Format (GS vs. VR). Second, we tested the comparability of the two VR-RAVLT alternate forms by performing an independent samples Mann–Whitney test comparing both forms' three main outcome measures. These comparisons revealed no significance results ( $p > 0.43$ ) confirming the forms comparability and allowing to combine their results in the further analyses.

Construct validity (Concurrent validity) of the VR-RAVLT was evaluated by computing Pearson's correlations coefficient between the GS-RAVLT and the VR-RAVLT tests' outcome measures (ACQUISITION, RI, and RETENTION) and by comparing the Pearson's correlations coefficients between each test's outcome measures and the main outcome measures of the four additional neuropsychological tests (the MOCA test, the WAIS-R Digit Symbol Test, the Verbal Fluency Test, and the WMS III digit span test).

To calculate serial position effects, the participant's recalled words for each test format were divided into three segments: Primacy (words 1–5), Middle (words 6–10), and Recency (words 11–15) in line with the literature (Bernard, 1991; Suhr, 2002). Then, these segments were summed and submitted to a repeated measures ANOVA to compare segment and test format effects.

Discriminant validity of the VR-RAVLT (i.e., ability to separate between the age groups) was evaluated by comparing the receiver operating characteristic (ROC) area under the curve values (AUC; range: 0–1, higher values reflect better discriminability) for the main outcome measures for each test format.

Test-retest reliability of the VR-RAVLT was validated by comparing the intraclass correlation coefficients [ICC; two-way mixed, effects, absolute agreement, (Koo and Li, 2016)] of the three outcome measures (ACQUISITION, RI, and

RETENTION) computed from the results of the GS-RAVLT and the VR-RAVLT that were performed by the same participant cohorts at two visits, 2–3 weeks apart. By convention ICC > 0.75 is considered good reliability (Koo and Li, 2016).

Level of statistical significance was set at 0.05. Statistical analyses were run using SPSS software (SPSS Ver. 24, IBM).

## Results

All of the participants complied with the VR-RAVLT platform, with no complains on any discomfort or inability to perform the instructions related to the verbal memory test.

### Main outcome measures

Table 2 presents the Summary statistics (mean ± SD) that were computed for each outcome measure.

Group effects were observed for the ACQUISITION and the RETENTION variables in the VR-RAVLT,  $H(2) = 19.49$ ,  $p < 0.0001$ ,  $H(2) = 8.31$ ,  $p = 0.01$ , respectively with poorer scores for the older-adults group. A similar pattern of results was observed in the GS-RAVLT,  $H(2) = 26.13$ ,  $p < 0.0001$ ,  $H(2) = 15.44$ ,  $p < 0.0001$ , respectively.

In regards to Format effects, only the RETENTION variable was found to be significant,  $Z = 2.49$ ,  $p = 0.01$ , resulting from higher retention rates in the VR-RAVLT. None of the other test's results were significant ( $p > 0.13$ ).

Since the middle-aged group had significantly more education years in comparison to the young- and older-adults (see Table 1), we conducted a control analysis by omitting the participants with the highest years of education ( $\geq 20$ ) from the MA group ( $n = 6$ ) so that this parameter was no longer different between groups. We repeated the statistical analysis and found similar results (see details in section 1 of the Supplementary material).

### Construct validity

Statistically significant correlations were found between the GS-RAVLT and the VR-RAVLT for two of the three outcome measures (ACQUISITION:  $r = 0.721$ ,  $p < 0.0001$ ;

TABLE 2 Summary statistics for each outcome measure – Overall and within the age groups.

	ACQUISITION (0–75)		RI (0–15)		RETENTION (0–15)	
VR-RAVLT (mean ± SD, range)	53.1 ± 8.8 (25–71)		1.0 ± 1.7 (–3 to 7)		0.08 ± 1.6 (–3 to 5)*	
YA MA OLD	57.3 ± 7.8 (33–71) <sup>α</sup>	45.8 ± 9.7 (25–61) <sup>β</sup>	1.2 ± 1.6 (–2 to 6)	1.4 ± 2.2 (–2 to 7)	0.6 ± 1.3 (–2 to 5) <sup>α</sup>	1.5 ± 1.7 (–3 to 5) <sup>β</sup>
GS-RAVLT (mean ± SD, range)	54.3 ± 10.6 (26–71)		1.3 ± 2.1 (–2 to 8)		1.6 ± 2.5 (–3 to 10)*	
YA MA OLD	59.5 ± 7.2 (34–71) <sup>α</sup>	43.6 ± 10.4 (26–59) <sup>β</sup>	0.9 ± 1.6 (–1 to 6)	2.0 ± 2.0 (–1 to 6)	0.4 ± 1.3 (–2 to 3) <sup>α</sup>	3.5 ± 2.8 (0–10) <sup>β</sup>

ACQUISITION: sum of trial 1–5; RI, Retroactive Interference, trial 5 – trial 7; RETENTION: trial 5 – Delayed Recall;  $\alpha, \beta, p \leq 0.01$  (between age groups effect, within outcome measure); \* $p = 0.01$  (between test version effect, overall).

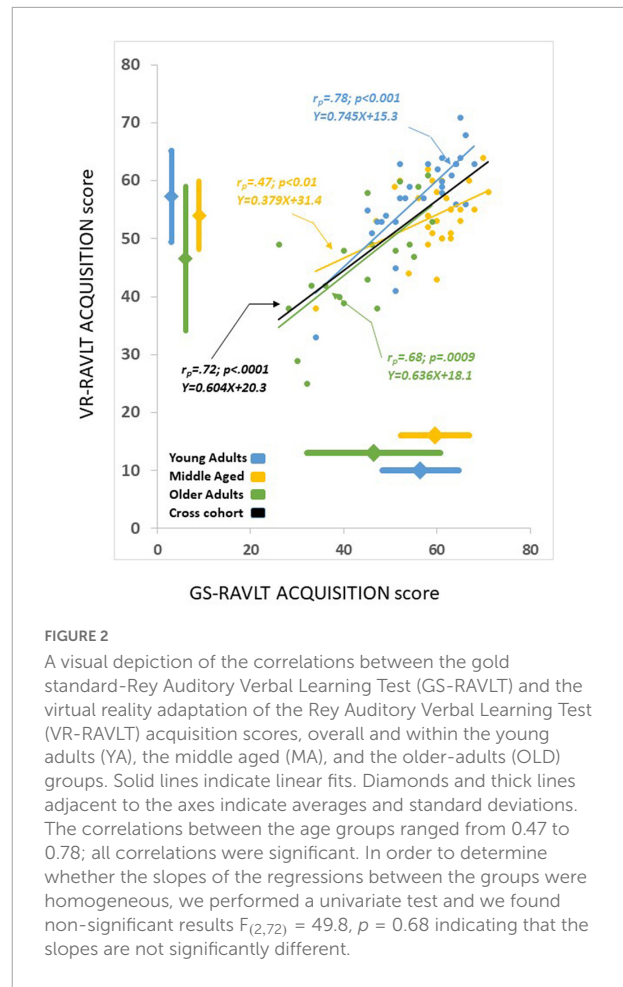


FIGURE 2

A visual depiction of the correlations between the gold standard-Rey Auditory Verbal Learning Test (GS-RAVLT) and the virtual reality adaptation of the Rey Auditory Verbal Learning Test (VR-RAVLT) acquisition scores, overall and within the young adults (YA), the middle aged (MA), and the older-adults (OLD) groups. Solid lines indicate linear fits. Diamonds and thick lines adjacent to the axes indicate averages and standard deviations. The correlations between the age groups ranged from 0.47 to 0.78; all correlations were significant. In order to determine whether the slopes of the regressions between the groups were homogeneous, we performed a univariate test and we found non-significant results  $F_{(2,72)} = 49.8, p = 0.68$  indicating that the slopes are not significantly different.

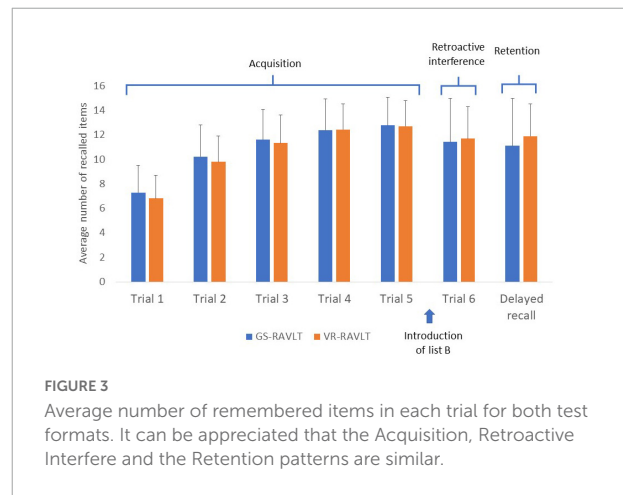


FIGURE 3

Average number of remembered items in each trial for both test formats. It can be appreciated that the Acquisition, Retroactive Interference and the Retention patterns are similar.

RETENTION,  $r = 0.31, p = 0.005$ ; RI,  $r = 0.18, p = 0.10$ ; all 78 participants were included in this analysis, *Post hoc* power calculations are provided in section 2 of the **Supplementary material**, suggesting construct validity of the VR-RAVLT (see, e.g., **Figure 2**). To demonstrate these correlations, **Figure 3** depicts the average number of remembered items on each trial for both test formats.

TABLE 3 Pearson correlations between the RAVLT outcome measures and four other neuropsychological tests.

	MOCA score	WAIS-R digit symbol test score	WMS III digit span test Total	Verbal fluency test total
GS-RAVLT ACQUISITION	0.62**	0.57**	0.38**	0.24**
VR-RAVLT ACQUISITION	0.49**	0.54**	0.34**	0.28**
GS-RAVLT RI	-0.25*	-0.25*	-0.07	-0.13
VR-RAVLT RI	-0.09	-0.11	0.11	0.02
GS-RAVLT RETENTION	0.41**	0.30**	-0.05	0.14
VR-RAVLT RETENTION	0.29**	0.24*	0.14	0.23*

RI, retroactive Interference; YA, Young Adults; MA, Middle Aged; MOCA, Montreal Cognitive Assessment; WAIS-R, Wechsler Adult Intelligence Scale-Revised; WMS, Wechsler Memory Scale.

\* $p \leq 0.05$ ; \*\* $p < 0.01$ .

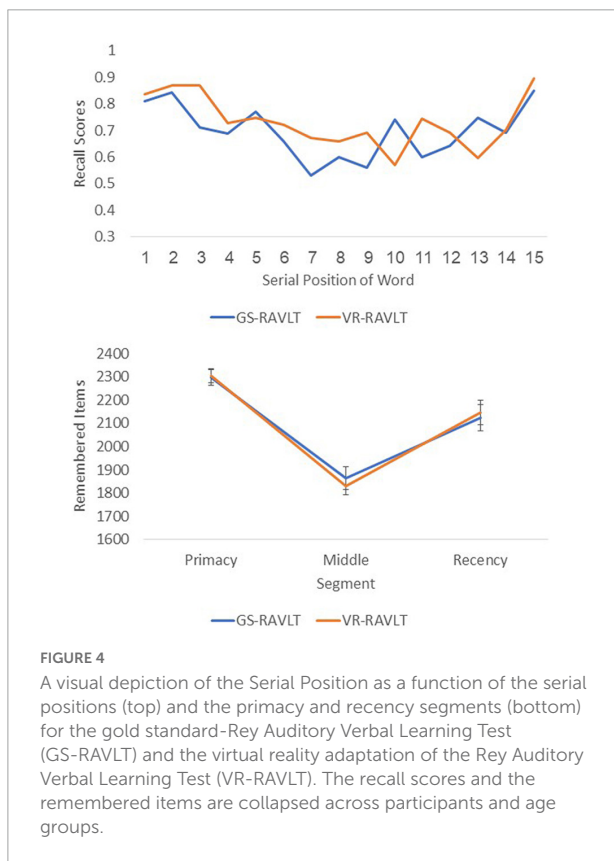


FIGURE 4

A visual depiction of the Serial Position as a function of the serial positions (top) and the primacy and recency segments (bottom) for the gold standard-Rey Auditory Verbal Learning Test (GS-RAVLT) and the virtual reality adaptation of the Rey Auditory Verbal Learning Test (VR-RAVLT). The recall scores and the remembered items are collapsed across participants and age groups.

Table 3 presents the overall Pearson correlations coefficients between the RAVLT tests' main outcome measures and the main outcome measures of other four neuropsychological tests.

It can be appreciated that the overall correlation patterns between the neuropsychological tests and both formats of the RAVLT test are similar, also confirming the Construct validity of the VR-RAVLT. These correlations within each age group, scoring and Group effects related to the additional neuropsychological tests are reported in section 3 and (Tables 2, 3) of the Supplementary material.

Result of the serial position effect analysis yielded a significant main effect of segments ( $F_{(2,12)} = 7.1$ ,  $p = 0.009$ ,  $\eta^2 = 0.54$ ), but no main test format effect ( $F_{(1,12)} = 0.002$ ,  $p = 0.97$ ,  $\eta^2 = 0.00$ ) and no interaction effect ( $F_{(2,12)} = 0.08$ ,

$p = 0.92$ ,  $\eta^2 = 0.01$ ). These results portray robust primacy and recency effects that were similarly presented in both test formats, a pattern that also point to the VR-RAVLT's good construct validity. Figure 4 depicts a visual impression of the serial position effects as a function of the serial positions and the primacy and recency segments.

## Discriminant validity

Table 4 depicts the ROC AUC values for discriminating between the three cohorts (bilateral discrimination). It can be appreciated that overall, the two formats have roughly similar discriminability, and consistent with the results presented in the "Main outcome measures" section (the source of the group effect are the values of the older-adults group). RI, however does not contribute to the ability to discriminate between the groups.

## Test-retest reliability

For the VR-RAVLT (retest interval of 2–3 weeks), good reliability was yielded for the ACQUISITION ( $ICC = 0.732$ ,  $p < 0.0001$ ) and the RETENTION ( $ICC = 0.321$ ,  $p = 0.050$ ) and poor reliability was found for the RI ( $ICC = -0.185$ ,  $p = 0.844$ ) variable. For the GS-RAVLT a similar pattern was observed; good reliability was found for the ACQUISITION ( $ICC = 0.822$ ,  $p < 0.0001$ ) and the RETENTION ( $ICC = 0.669$ ,  $p < 0.0001$ ) variables, and poor reliability was found for the RI ( $ICC = -0.379$ ,  $p = 0.992$ ) variable.

## Discussion

The objective of the current work was to test the construct validity, the age-related discriminant validity and the test-retest validity of a novel, virtual reality format of the RAVLT, the VR-RAVLT, in reference to the original gold standard RAVLT. For this objective, seventy-eight healthy participants from three age groups performed both tests.

Overall, the results confirm construct validity of the VR-RAVLT, as reflected by significant correlations between the

TABLE 4 AUC values from ROC curves.

	GS-RAVLT			VR-RAVLT		
	Acquisition	RI	Retention	Acquisition	RI	Retention
YA vs. MA	0.59	0.51	0.62	0.67*	0.62	0.52
YA vs. OLD	0.83**	0.63	0.72*	0.82**	0.61	0.71*
MA vs. OLD	0.91**	0.65	0.85**	0.76*	0.51	0.72*

AUC, area under the curve; ROC, receiver operating characteristic curves; YA, young adults; MA, middle aged adults; OLD, Older-adults; RI, retroactive Interference. Details on actual scores are not shown.

\* $p \leq 0.05$ ; \*\* $p < 0.0001$ .

two formats in terms of the outcome measures constituting the main verbal memory performance (i.e., ACQUISITION, RETENTION). At odd was the retrospective interference—RI (see more below). Additionally, significant serial position effects (primacy and recency) were similarly demonstrated for both test formats, and comparable to those that were reported in the literature (Sullivan et al., 2002; Powell et al., 2004). Reviewing the correlations between the GS-RAVLT, the VR-RVLT and the other neuropsychological tests (Table 3), only the ACQUISITION and the RETENTION variables were highly correlated with other tests, which is also in agreement with the literature (McMinn et al., 1988; de Sousa Magalhães et al., 2012), while the RI seems to be less correlated. To the best of our knowledge, we are the first to report significant and high correlations between the MOCA scores, the Verbal Fluency Test scores and the WMS III digit span test scores and between the GS-RAVLT's main outcome measures, further confirming its construct validity. In addition, similarly to the younger adults, among the older adults group we found high correlations between the VR-RVLT and the other neuropsychological tests which were comparable to the GS-RAVLT's correlations with these tests (see Table 3 of the Supplementary material), suggesting that this group can understand, use, and be tested with the VR apparatus and indicating that this more technical and novel modality of cognitive assessment can be used for older adults.

Based on our samples, the RI measure does not seem to be similarly correlated with other neuropsychological tests' outcomes and to be able to properly discriminate between healthy age cohorts, as the other RAVLT's outcome measures. These results are in line with the few other cases in which this measure was calculated and reported [(Porter et al., 2003; Vakil et al., 2010; de Sousa Magalhães et al., 2012), but see Ricci et al. (2012) for a different result]. Considering these results, it seems that RI is not a sensitive index as the ACQUISITION or the RETENTION and its use in regards to measuring verbal memory should be re-visited.

In terms of discriminant validity, the VR-RAVLT was equivalent to the GS-RAVLT as being able to discriminate between older healthy adults and healthy young and middle-aged persons. When comparing the results to the literature,

it should be noted that the RAVLT's discriminant validity was mainly assessed by comparing clinical cohorts to healthy controls e.g., (Tierney et al., 1996; Estévez-González et al., 2003; Schoenberg et al., 2006; Balthazar et al., 2010; Ricci et al., 2012). At the same time, our AUC values are comparable to those obtained in studies that reported AUC (Ricci et al., 2012; Li et al., 2018). Notably, differentiating a clinical condition such as dementia or traumatic brain injury from normal cognition [i.e., (Schoenberg et al., 2006; Li et al., 2018)] is less difficult than differentiating between different age groups of cognitively normal individuals, as the cognitive changes related to aging are more subtle. Our results suggest that the RAVLT, including the VR-RAVLT, have the potential for discrimination of subtle early changes in cognition.

Finally, as to test-retest reliability, the VR-RAVLT yielded similar ICC values compared to the GS-RAVLT, and to those that were reported in the literature (Shapiro and Harrison, 1990; Vakil and Blachstein, 1997; Rezvanfard et al., 2011; de Sousa Magalhães et al., 2012; Nightingale et al., 2019).

The finding that the variance of both the VR-RAVLT and the GS-RAVLT ACQUISITION scores is greater among the OLD group than among the other groups (see Figure 2) could be attributed to some of the older participants having an incipient cognitive decline. In addition, variability in cognitive functioning increases with age (Glisky, 2007). We argue that, given the high correlation between the VR-RAVLT's and the GS-RAVLT's outcome measures found among this group, our results and conclusions are strengthened by these findings.

Re-framing the GS-RAVLT's words remembering task as a "places-to-go" remembering task in the VR-RAVLT, generated a constraint of using 15 words that only describe places as the items to remember, as opposed to the 15 non-semantically related words that are used in the GS-RAVLT. This change may have increased the words semantic relatedness and thus may have caused a potential advantage in remembering them due to semantic priming (Foss, 1982) or by a better chance of using semantic encoding (Tulving, 1972). As a result, one possible outcome of our study was that the participants would score higher on the VR-RAVLT than the GS-RAVLT, an outcome that has not been observed. It is possible that the semantic distance of some of the places in the VR-RAVLT (e.g., coffee

shop and airport) is similar to that of some words in the GS-RAVLT (e.g., house and river). An investigation of this hypothesis is warranted.

The current results do not demonstrate superiority of the VR-RAVLT over the GS-RAVLT in terms of age-related discriminant validity and test-retest validity, which may raise the question of the added value of the new test. In this context, the narrative (list of places to visit) and settings (an office) were designed to be ecological. However, the current task is still somewhat artificial i.e., to recall the same 15 item list seven times. This limitation was due to the requirement to translate the RAVLT to VR settings while not changing the task too much to keep the test's core psychometric properties. On the other hand, our findings point to excellent construct validity of the VR format. Therefore, we posit that the current VR-RAVLT should evolve into more "real life" VR versions, while still maintaining the GS validity. This will provide additional benefits such as increased availability and usability by enabling to use them in various settings including patients' homes. A next step will be to create a truly ecological verbal memory task which is more representative of daily tasks that would be valid and reliable while assessing verbal memory abilities, such as an everyday attempt to remember a grocery list or tasks. The combination of these advantages may make the VR version of the RAVLT more sensitive for the early ascertainment of incipient cognitive impairment.

As one of the limitations of the current study, the VR-RAVLT was not validated among non-healthy individuals such as those with cognitive impairment whereas one of RAVLT's main purposes is to detect such impairment. Ongoing VR-RAVLT studies are recruiting cognitively impaired individuals.

In summary, our results suggest that the novel VR-RAVLT and the GS-RAVLT share similar psychometric properties (e.g., an increase in scores between trials 1 and 5, and then reduced scores in trials 6 and delayed). Coupled with the general high correlations between corresponding outcome measures and similar serial position effects, this suggests that the two tests measure the same cognitive construct (verbal memory). Taken together these results are an indication of the feasibility of adapting the RAVLT to the VR environment while preserving its core feature. This work, along with our recent work in the Trails Making Test (Plotnik et al., 2021) support that additional neuropsychological tests can be adapted to VR format as well. Finally, this work consists of one more milestone in the process of creating ecological valid versions of classic neuropsychological tests, which may yield clinical benefits by providing opportunities to understand neuropsychological functioning of patients in "close" to real life settings.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, upon reasonable request.

## Ethics statement

The studies involving human participants were reviewed and approved by the Sheba Medical Center local IRB Committee. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

GD, MB, and MP: conceptualization, supervision, project administration, and writing—review and editing. OB-G: software. SK-N, MC, and HI: investigation. AG: data curation, formal analysis, and writing—original draft. All authors contributed to the article and approved the submitted version.

## Funding

This work was funded in part by Dr. Marina Nissim and Ms. Judie Stein.

## Acknowledgments

We wish to thank Mr. Yotam Hagur-Bahat for technical support.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnagi.2022.980093/full#supplementary-material>



## References

- Armstrong, C. M., Reger, G. M., Edwards, J., Rizzo, A. A., Courtney, C. G., and Parsons, T. D. (2013). Validity of the virtual reality stroop task (VRST) in active duty military. *J. Clin. Exp. Neuropsychol.* 35, 113–123. doi: 10.1080/13803395.2012.740002
- Balthazar, M. L., Yasuda, C. L., Cendes, F., and Damasceno, B. P. (2010). Learning, retrieval, and recognition are compromised in aMCI and mild AD: Are distinct episodic memory processes mediated by the same anatomical structures? *J. Int. Neuropsychol. Soc.* 16, 205–209. doi: 10.1017/S1355617709990956
- Bernard, L. C. (1991). The detection of faked deficits on the rey auditory verbal learning test: The effect of serial position. *Arch. Clin. Neuropsychol.* 6, 81–88. doi: 10.1093/arclin/6.1-2.81
- Bottari, C., Dassa, C., Rainville, C., and Dutil, E. (2009). The factorial validity and internal consistency of the instrumental activities of daily living profile in individuals with a traumatic brain injury. *Neuropsychol. Rehabil.* 19, 177–207. doi: 10.1080/09602010802188435
- Bowler, D. (2013). “Rey auditory verbal learning test (Rey AVLT),” in *Encyclopedia of autism spectrum disorders*, ed. F. R. Volkmar (New York, NY: Springer New York), 2591–2595.
- Callahan, C. D., and Johnstone, B. (1994). The clinical utility of the rey auditory-verbal learning test in medical rehabilitation. *J. Clin. Psychol. Med. Settings* 1, 261–268. doi: 10.1007/BF01989627
- Chaytor, N., Schmitter-Edgecombe, M., and Burr, R. (2006). Improving the ecological validity of executive functioning assessment. *Arch. Clin. Neuropsychol.* 21, 217–227. doi: 10.1016/j.acn.2005.12.002
- Crockett, D. J., Hadjistavropoulos, T., and Hurwitz, T. (1992). Primacy and recency effects in the assessment of memory using the rey auditory verbal learning test. *Arch. Clin. Neuropsychol.* 7, 97–107. doi: 10.1093/arclin/7.1.97
- de Sousa Magalhães, S., Fernandes Malloy-Diniz, L., and Cavalheiro Hamdan, A. (2012). Validity convergent and reliability test-retest of the rey auditory verbal learning test. *Clin. Neuropsychiatry* 9, 129–137.
- Elkind, J. S., Rubin, E., Rosenthal, S., Skoff, B., and Prather, P. (2001). A simulated reality scenario compared with the computerized Wisconsin card sorting test: An analysis of preliminary results. *Cyberpsychol. Behav.* 4, 489–496. doi: 10.1089/109493101750527042
- Erdfelder, E., Faul, F., and Buchner, A. (1996). GPOWER: A general power analysis program. *Behav. Res. Methods Instr. Comput.* 28, 1–11. doi: 10.3758/BF03203630
- Estévez-González, A., Kulisevsky, J., Boltes, A., Otermin, P., and García-Sánchez, C. (2003). Rey verbal learning test is a useful tool for differential diagnosis in the preclinical phase of Alzheimer’s disease: Comparison with mild cognitive impairment and normal aging. *Int. J. Geriatr. Psychiatry* 18, 1021–1028. doi: 10.1002/gps.1010
- Fichman, H. C., Dias, L. B. T., Fernandes, C. S., Lourenço, R., Caramelli, P., and Nitrini, R. (2010). Normative data and construct validity of the rey auditory verbal learning test in a Brazilian elderly population. *Psychol. Neurosci.* 3, 79–84. doi: 10.3922/j.pns.2010.1.010
- Foss, D. J. (1982). A discourse on semantic priming. *Cogn. Psychol.* 14, 590–607. doi: 10.1016/0010-0285(82)90020-2
- Gal, O. B., Doniger, G. M., Cohen, M., Bahat, Y., Beeri, M. S., Zeilig, G., et al. (2019). “Cognitive-motor interaction during virtual reality trail making,” in *Proceedings of the 2019 International Conference on Virtual Rehabilitation (ICVR)*, (Tel Aviv: IEEE), 1–6.
- Geffen, G., Moar, K., O’hanlon, A., Clark, C., and Geffen, L. (1990). Performance measures of 16- to 86-year-old males and females on the auditory verbal learning test. *Clin. Neuropsychol.* 4, 45–63. doi: 10.1080/13854049008401496
- Glisky, E. L. (2007). “Changes in cognitive function in human aging,” in *Brain aging: Models, methods, and mechanisms*, ed. D. R. Riddle (Boca Raton FL: CRC Press), 3–20. doi: 10.1201/9781420005523-1
- Groth-Marnat, G. (2009). *Handbook of psychological assessment*. Hoboken, NJ: John Wiley & Sons.
- Guilmette, T. J., and Rasile, D. (1995). Sensitivity, specificity, and diagnostic accuracy of three verbal memory measures in the assessment of mild brain injury. *Neuropsychology* 9:338. doi: 10.1037/0894-4105.9.3.338
- Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropr. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012
- Lezak, M. D., Howieson, D. B., Loring, D. W., and Fischer, J. S. (2004). *Neuropsychological assessment*. New York, NY: Oxford University Press.
- Li, X., Jia, S., Zhou, Z., Jin, Y., Zhang, X., Hou, C., et al. (2018). The role of the Montreal cognitive Assessment (MoCA) and its memory tasks for detecting mild cognitive impairment. *Neurol. Sci.* 39, 1029–1034. doi: 10.1007/s10072-018-3319-0
- Linzen, T. (2009). *Corpus of blog postings collected from the Israblog website*. Tel Aviv: Tel Aviv University.
- Manchester, D., Priestley, N., and Jackson, H. (2004). The assessment of executive functions: Coming out of the office. *Brain Injury* 18, 1067–1081. doi: 10.1080/02699050410001672387
- McMinn, M. R., Wiens, A. N., and Crossen, J. R. (1988). Rey auditory-verbal learning test: Development of norms for healthy young adults. *Clin. Neuropsychol.* 2, 67–87. doi: 10.1080/13854048808520087
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x
- Nightingale, T. E., Lim, C. A. R., Sachdeva, R., Zheng, M. M. Z., Phillips, A. A., and Krassioukov, A. (2019). Reliability of cognitive measures in individuals with a chronic spinal cord injury. *PM R* 11, 1278–1286. doi: 10.1002/pmrj.12161
- Parsons, T. D., Carlew, A. R., Magtoto, J., and Stonecipher, K. (2017). The potential of function-led virtual environments for ecologically valid measures of executive function in experimental and clinical neuropsychology. *Neuropsychol. Rehabil.* 27, 777–807. doi: 10.1080/09602011.2015.1109524
- Parsons, T. D., Courtney, C. G., and Dawson, M. E. (2013). Virtual reality stroop task for assessment of supervisory attentional processing. *J. Clin. Exp. Neuropsychol.* 35, 812–826. doi: 10.1080/13803395.2013.824556
- Plotnik, M., Ben-Gal, O., Doniger, G. M., Gottlieb, A., Bahat, Y., Cohen, M., et al. (2021). Multimodal immersive trail making—virtual reality paradigm to study cognitive-motor interactions. *J. Neuroeng. Rehabil.* 18:82. doi: 10.1186/s12984-021-00849-9
- Pollak, Y., Weiss, P. L., Rizzo, A. A., Weizer, M., Shriki, L., Shalev, R. S., et al. (2009). The utility of a continuous performance test embedded in virtual reality in measuring ADHD-related deficits. *J. Dev. Behav. Pediatr.* 30, 2–6. doi: 10.1097/DBP.0b013e3181969b22
- Porter, R. J., Gallagher, P., Thompson, J. M., and Young, A. H. (2003). Neurocognitive impairment in drug-free patients with major depressive disorder. *Br. J. Psychiatry* 182, 214–220. doi: 10.1192/bjp.182.3.214
- Powell, M. R., Gfeller, J. D., Oliveri, M. V., Stanton, S., and Hendricks, B. (2004). The Rey AVLT serial position effect: A useful indicator of symptom exaggeration? *Clin. Neuropsychol.* 18, 465–476. doi: 10.1080/1385404049052409
- Rey, A. (1964). *Rey auditory verbal learning test (RAVLT). L’Examen clinique en psychologie*. Paris: PUF.
- Rezvafard, M., Ekhtiari, H., Rezvanfar, A., Noroozian, M., Nilipour, R., and Javan, G. K. (2011). The Rey auditory verbal learning test: Alternate forms equivalency and reliability for the iranian adult population (Persian version). *Archives of Iranian medicine* 14, 104–109.
- Ricci, M., Graef, S., Blundo, C., and Miller, L. A. (2012). Using the Rey Auditory Verbal Learning Test (RAVLT) to differentiate Alzheimer’s dementia and behavioural variant fronto-temporal dementia. *Clin. Neuropsychol.* 26, 926–941. doi: 10.1080/13854046.2012.704073
- Sbordone, R. J. (2008). Ecological validity of neuropsychological testing: Critical issues. *Neuropsychol. Handb.* 367:394.
- Schmidt, M. (1996). *Rey auditory verbal learning test: A handbook*. Los Angeles, CA: Western Psychological Services.
- Schoenberg, M. R., Dawson, K. A., Duff, K., Patton, D., Scott, J. G., and Adams, R. L. (2006). Test performance and classification statistics for the Rey Auditory Verbal Learning Test in selected clinical samples. *Arch. Clin. Neuropsychol.* 21, 693–703. doi: 10.1016/j.acn.2006.06.010
- Shapiro, D. M., and Harrison, D. W. (1990). Alternate forms of the AVLT: A procedure and test of form equivalency. *Arch. Clin. Neuropsychol.* 5, 405–410. doi: 10.1093/arclin/5.4.405
- Suhr, J. A. (2002). Malingering, coaching, and the serial position effect. *Arch. Clin. Neuropsychol.* 17, 69–77. doi: 10.1093/arclin/17.1.69
- Sullivan, K., Deffenti, C., and Keane, B. (2002). Malingering on the RAVLT Part II. Detection strategies. *Arch. Clin. Neuropsychol.* 17, 223–233. doi: 10.1093/arclin/17.3.223

- Tierney, M., Szalai, J., Snow, W., Fisher, R., Nores, A., Nadon, G., et al. (1996). Prediction of probable Alzheimer's disease in memory-impaired patients: A prospective longitudinal study. *Neurology* 46, 661–665. doi: 10.1212/WNL.46.3.661
- Tulving, E. (1972). "Episodic and semantic memory," in *Organization of memory*, eds E. Tulving and W. Donaldson (Cambridge, MA: Academic Press).
- Vakil, E., and Blachstein, H. (1993). Rey auditory-verbal learning test: Structure analysis. *J. Clin. Psychol.* 49, 883–890. doi: 10.1002/1097-4679(199311)49:6<883::AID-JCLP2270490616>3.0.CO;2-6
- Vakil, E., and Blachstein, H. (1997). Rey AVLT: Developmental norms for adults and the sensitivity of different memory measures to age. *Clin. Neuropsychol.* 11, 356–369. doi: 10.1080/13854049708400464
- Vakil, E., Greenstein, Y., and Blachstein, H. (2010). Normative data for composite scores for children and adults derived from the Rey Auditory Verbal Learning Test. *Clin. Neuropsychol.* 24, 662–677. doi: 10.1080/13854040903493522
- Vlahou, C. H., Kosmidis, M. H., Dardagani, A., Tsotsi, S., Giannakou, M., Giatzoulidou, A., et al. (2013). Development of the Greek verbal learning test: Reliability, construct validity, and normative standards. *Arch. Clin. Neuropsychol.* 28, 52–64. doi: 10.1093/arclin/acs099
- Wechsler, D. (1997). *Wechsler memory scale*, 3rd Edn. San Antonio, TX: The Psychological Corporation.
- Wechsler, D., and De Lemos, M. M. (1981). *Wechsler adult intelligence scale-revised*. New York, NY: Harcourt Brace Jovanovich.
- Wilson, B. A. (1993). Ecological validity of neuropsychological assessment: Do neuropsychological indexes predict performance in everyday activities? *Appl. Prevent. Psychol.* 2, 209–215. doi: 10.1016/S0962-1849(05)80091-5