



Research Article

Source memory across the lifespan: Insights from a virtual reality based neuropsychological assessment

Gema Climent¹, Joseph James Cosgrove², Fidel Rebon-Ortiz³, Irene Alice Chicchi Giglioli³ and

Unai Diaz-Orueta^{2,4} 

¹Universidad de Almería, Almería, Spain, ²Department of Psychology, Maynooth University, Maynooth, Ireland, ³Department of R+D, Giunti-Nesplora, Alcobendas, SL, Spain and ⁴Universidad Internacional de la Rioja (UNIR), Logroño, Spain

Abstract

Objective: The goal of the current study was to study the contribution of source memory, more specifically, a source memory task, on the memory performance measured with a novel virtual reality (VR)-based neuropsychological assessment test, i.e., the Suite Test. **Method:** The sample included 676 subjects (49.7% female), aged from 12 to 85 years. The Suite test comprises a 360-degree VR environment designed as a furniture shop. Participants must group specific sets of furniture items (ordered by different families of customers) by clicking on the furniture to be packed, following instructions from a voice-over. All participants were administered the full version of the test, which comprises, among others, an immediate recall task, a source memory task, a short-term delayed recall task, a long-term delayed recall task, and a recognition trial. **Results:** Performance on the VR source memory task was associated with recall across age groups, with a stronger contribution in older adults, often enhancing long-term recall. In contrast, younger individuals relied more on immediate and short-term delayed recall, with weaker relationships between source memory and the other types, suggesting that it plays a more secondary role in younger participants. **Conclusions:** The Suite Test VR-based test effectively explores source memory contributions across the lifespan. By immersing participants in a dynamic VR environment, it reveals how source memory relates to other memory types, showing age-related differences and offering valuable insights about cognitive changes, as well as about future research implications in the area of memory assessment.

Keywords: Neuropsychological assessment; virtual reality; source memory; memory interference; cognitive dysfunction; memory disorders

(Received 13 January 2025; final revision 8 September 2025; accepted 9 September 2025)

Statement of Research Significance

Research Question(s) or Topic(s): How does a source memory task contribute to memory processes across different age groups? **Main Findings:** Older adult groups benefit more from the source memory task than their younger counterparts when performing delayed recall tasks. Introducing a source memory task may help some age groups elaborate further on material to be learned and remembered. **Study Contributions:** Some age groups may be more adequately assessed using a source memory task that explores both the content and context of memory. The novelty of the study is the introduction of a source memory task that requires the processing of contextually relevant information. This contextual processing does not seek to interfere with or create confusion in the respondent, but rather to delve deeper into the coding of the material to be remembered, which appears to improve subsequent recall, especially in certain age groups.

Introduction

Memory refers to the processes by which an individual encodes, stores, and retrieves information (Strauss et al., 2006). According

to Lezak et al. (2012), many common neurological and psychiatric conditions show a decline in the efficiency of memory processes, and, consequently, assessment of memory is often a central issue in a neuropsychological examination.

With regard to this, a proliferation of assessment tools exists for the assessment of specific aspects of verbal memory (see, for example, Lezak et al., 2012; Sherman & Hrabok, 2023) and many of these tools have employed the use of interference as an integral part of standardized verbal memory assessment (Brophy et al., 2009). Some examples are the widely used Rey Auditory Verbal Learning Test (Rey, 1942) or the California Verbal Learning test (Delis et al., 2000). In these tools, there is a set of learning trials of a List A (or Monday shopping list) that is presented five times to the testee, followed by a List B (or Tuesday shopping list) that is presented in a single trial. When the testee has to repeat items from List B, according to the proactive interference theory (Underwood, 1957), the information presented before (List A, presented five times) may prevent or make the acquisition of new materials difficult, thus making it difficult to remember items from List B. Moreover, some items of List A may mistakenly be remembered as part of List B

Corresponding author: Unai Diaz-Orueta; Email: unai.diazorueta@mu.ie

Cite this article: Climent G., Cosgrove J.J., Rebon-Ortiz F., Chicchi Giglioli I.A., & Diaz-Orueta U. Source memory across the lifespan: Insights from a virtual reality based neuropsychological assessment. *Journal of the International Neuropsychological Society*, 1–10, <https://doi.org/10.1017/S1355617725101483>

© The Author(s), 2025. Published by Cambridge University Press on behalf of International Neuropsychological Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

(i.e., intrusions). After the presentation of List B or Tuesday list, the testee is usually requested to remember items from List A, which opens the door to retroactive interference, thus causing that information presented after the target material (i.e., List B presented once after List A presented five times) negatively affects recall (McGeoch, 1932).

This way of testing verbal memory has been the main paradigm for years, both in research and clinical practice, and only recently have some studies appeared to question or at least test the effects of interference paradigms in memory performance. For example, Brophy *et al.* (2009) investigated the effect of interference on delayed recall scores of the WMS-III and other commonly used memory measures. They found that the introduction of interference items during the delay affected negatively delayed recall performance on almost all sub-tests. Libon *et al.* (2011) found that, as compared to other groups, amnesic MCI patients appeared to be very susceptible to the deleterious effect of interference test conditions, along with greater penetration of list B words into subsequent list A recall. Separately Rahimi-Golkhandan *et al.* (2012) suggested that knowing about differences in susceptibility to interference across tasks may provide some important diagnostic and cognitive information for researchers and clinicians, and the choice of verbal learning tests should be guided by the knowledge of interference effects and the susceptibility of patient groups to this effect. Overall, while the use of interference could be a differential factor for specific memory profiles, it can also constitute a confounding factor for memory and recall if its effects are not properly examined.

In this context of aiming to overcome problems traditionally associated with memory assessment, virtual reality (VR) offers a potentially interesting alternative for the assessment of many cognitive processes and the inclusion of other memory processes potentially relevant for clinical diagnosis, such as source memory. VR offers several advantages in the context of neuropsychological assessment and cognitive research, such as increased ecological validity (allowing observation of behaviors in complex, dynamic settings that better reflect everyday life) and the ability to minimize examiner-related variability (Diaz-Orueta *et al.*, 2020; Parsons, 2015; Pieri *et al.*, 2023; Rizzo *et al.*, 2004). However, VR is not without its limitations. Ceccato *et al.* (2024) showed that recognition accuracy and confidence were significantly higher in real-life modality environments than in similar VR modalities, and that additional research is needed to make VR environments sufficiently comparable to real-life contexts, paying special attention to the impact of stimuli typicality and emotional valence in VR contexts. In addition, variability in user familiarity with technology and challenges in standardizing VR protocols across studies may affect the reliability and generalizability of results.

Despite these drawbacks, ongoing technological advancements and increasing empirical support suggest that VR holds significant potential as a complementary tool in neuropsychological evaluation. Recently, Climent *et al.* (2024) published the normative data for a new developed VR-based neuropsychological test, the Suite test, on a population between 12 and 85 years old. Suite is a VR-based test designed to evaluate visual memory and aims to help clinicians support the diagnosis of memory-related conditions or disorders. One of the features of Suite Test is the inclusion of a source-memory task.

Source memory can be defined as memory for details that accompany the central component of an event, or also as memory for temporal details –as well as other contextual details– of the encoded information (Palombo *et al.*, 2021). In their study, Minor

and Herzmann (2019) defined source memory in an operational way as the ability to recall both an image and its block, as opposed to item memory, which was represented by memory for the image irrespective of the memory for its block.

When assessing source memory, it can be done by asking the individual where, from where, how, or when the information was learned (like in Ceccato *et al.*, 2022, where they requested their participants to remember when each time was seen among three possible timepoints). Separately, Squire *et al.* (2007) presented a task that required individuals to be exposed to items such as a list of words, in two separate states (at the top vs. the bottom of a computer screen), and then declare whether the information originated from the top or the bottom of the screen. Symeonidou & Kuhlmann (2021) used a task with audio software to distinguish ‘sources’, alongside visual faces and corresponding names. The participants were not only asked to differentiate whether an item came from source A, source B, or was new; but they were also asked to describe their retrieval strategies. This is different from item memory, as item memory involves memory retrieval linked to the semantic features of information (Guo *et al.*, 2021), although these two forms of memory have been shown to have an almost symbiotic relationship. The retrieval of both source memory and item memory increases when the relationship between these two components is strongly linked (Guo *et al.*, 2021), and this was the rationale behind the development of the source memory task within the Suite Test.

In summary, the exploration of source memory reveals a nuanced and interconnected cognitive operation essential for recalling not just what we learn but also how and where we learn it. The symbiotic relationship between source and item memory, intricately orchestrated by the prefrontal cortex and the hippocampus (Guo *et al.*, 2021) suggest that a memory test comprising a set of tasks that measure both item and source memory could be a potential addition to tests relying on interference-based tasks or paradigms, and a way to further understand the development of memory processes across the lifespan.

For all these reasons, the goal of the current study is to focus on the source memory task of the Suite test, a recently developed VR-based neuropsychological assessment tool for memory. This research aims to contribute to building a body of knowledge about this type of memory and its relevance, even in non-pathological processes. In order to achieve that, the present study aims to evaluate how source memory, integrated into a virtual reality (VR)-based neuropsychological task (the Suite Test), contributes to overall memory performance across different age groups. In line with clinical reporting conventions, we will use *immediate recall* and *delayed recall* to label test scores. Conceptually, we will refer to *memory processes* (e.g. source/contextual memory) when discussing mechanisms, in order to avoid conflating score labels with theoretical claims about memory systems or stores.

Method

Participants

This study employs the normative sample of the SUITE test ($N = 676$) as delineated by Climent *et al.* (2024). The participant cohort included individuals aged 12 to 85 years, demonstrating a balanced gender distribution, with 49.7% of female participants (mean age = 32.65 years old) and 50.3% of male (mean age = 30.46 years old). Exclusion criteria encompassed individuals with visual,

Table 1. Normative sample age distribution

Group	Mean age (years)
12-year-olds (both genders)	12
13 – 26 Male	15.51
13 – 26 Female	15.74
27 – 44 (both)	36.03
45 – 58 (Male)	51.23
45 – 58 (Female)	50.94
59 – 85 (both)	68.31

auditory, or motor impairments that could adversely impact their ability to interact with the test. Furthermore, individuals with a history of acquired brain injury, psychiatric disorders, or neurodegenerative diseases were excluded to maintain the validity of the findings. Table 1 below shows the age distribution of the normative groups.

The original sample included 676 subjects, but a total of 29 participants were excluded from the final sample. In 14 cases, reasons for exclusion included the participant removing the VR headset during the test, misunderstanding the instructions (which led to invalid performance), or hardware malfunctions during the test administration. Additionally, another 15 subjects were excluded due to poor performance on a forced-choice recognition task (i.e., worse performance than what would be expected if the subjects answered by chance -i.e. less than half of correct answers in that task, as will be detailed later in the procedure), which was attributed to factors like faking their performance, fatigue, or tremors. As a result, the final sample comprised 647 subjects, representing a 4.29% exclusion rate.

Ethics approval statement

The study obtained ethical approval from the Human Research Ethics Committee of the University of the Basque Country in northern Spain, and all participants provided informed consent before the beginning of the study. The study was conducted following the Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

Study procedure

To administer the SUITE test, the participant sat comfortably in a chair and was provided with a VR headset and a controller operated with the dominant hand, like a joystick with buttons that allowed the participant to click and interact with various features of the virtual environment. The entire setup process, which included the participant putting on the VR headset, the evaluator's computer starting the SUITE VR desktop control application and connecting to the same network as the participant's VR headset, the evaluator entering basic participant information into the SUITE computer application (name or code, age, gender, and whether the participant was left-handed or right-handed), and starting the test administration, took a maximum of 30 minutes.

Instruments

Suite is a comprehensive neuropsychological assessment tool designed to evaluate memory functions, including immediate memory, source memory, short and long delayed free recall, recognition, and memory strategies (visual, verbal, primacy, recency), in individuals aged 12 years and older, through an immersive VR environment facilitated by an Oculus Quest headset. The virtual environment is set in a furniture store where

participants are tasked with organizing various pieces of furniture based on specific criteria for packing and shipping. Using the head-tracking technology of the VR headset, participants can explore the environment with a 360-degree view. As described in Climent et al. (2024), in the virtual scenario, there is a television screen located in a central position. On that screen, stimuli (i.e., pieces of furniture) are presented with an image of each of them and their name. A male voiceover in European Spanish names the furniture that must be packed, and the respondent has to locate each item in the virtual room, point and click on it. Additionally, the voice informs participants about different customer groups—referred to as “families”—each requesting between four and six pieces of furniture that must be remembered. The customer groups include (1) a family of four (a man, a woman, and two children), (2) a single man, (3) two men, (4) a single woman, and (5) a family of five (a man, a woman, and three children). Figure 1 below shows a sketch of the original Suite environment. To preserve the validity of the assessment, further details regarding the interface and specific stimuli or customer groups are not disclosed to prevent overexposure.

SUITE collected data related to response accuracy, defined as measures of how accurately participants completed tasks within the VR, such as correctly identifying or grouping furniture items (correct answers and errors), response times, and strategies used.

Below, the different tasks administered in Suite are displayed:

- **Task 1: Immediate Recall (Order List):** The user must select and recall furniture ordered by different families (5 groups with orders of 4–6 pieces each). Three rounds of the same order are presented to assess the learning curve. The selection strategy is analyzed (i.e., does the user select the furniture in sequential order or by category?). The recorded data are: (1) selected furniture and their order; (2) response times in milliseconds, and (3) memory strategies (primacy, recency, verbal, or visual strategies).
- **Task 2: Source Memory:** A total of eight items of previously ordered furniture are presented, and the user must remember which family requested each of them. Images of the family groups appear on the screen to select the answer. The recorded data are: (1) accuracy in matching each piece of furniture ordered with the right family; (2) response time, and (3) use of a virtual red button available on the screen if the user cannot remember the answer. Source memory, as measured here, refers to the contextual information around target items that is not explicitly requested to the individual to remember. In Suite test, that contextual information would refer to “families” who make the orders, and target items would be the specific set of pieces of furniture requested in each order that need to be learned and remembered by the testee.
- **Tasks 3: Free Delayed Recall (short delay):** Each family is shown again, and the user must select the furniture that corresponds to them, without receiving any clues from previous tasks. The recorded data are: (1) accuracy in retrieving orders; (2) response time and pauses (i.e., latencies) between selections, and (3) a path map within the environment (i.e., a graphic representation of the scanning sequence performed by the user to select the target stimuli).
- **Task 4: Prospective memory item:** The user must remember to turn off the store lights when a bell rings. It is registered whether the user remembers to act or not, and the response time.
- **Task 5. Free delayed recall (long delay):** Same task as task 3, presented after a 20-minute delay.

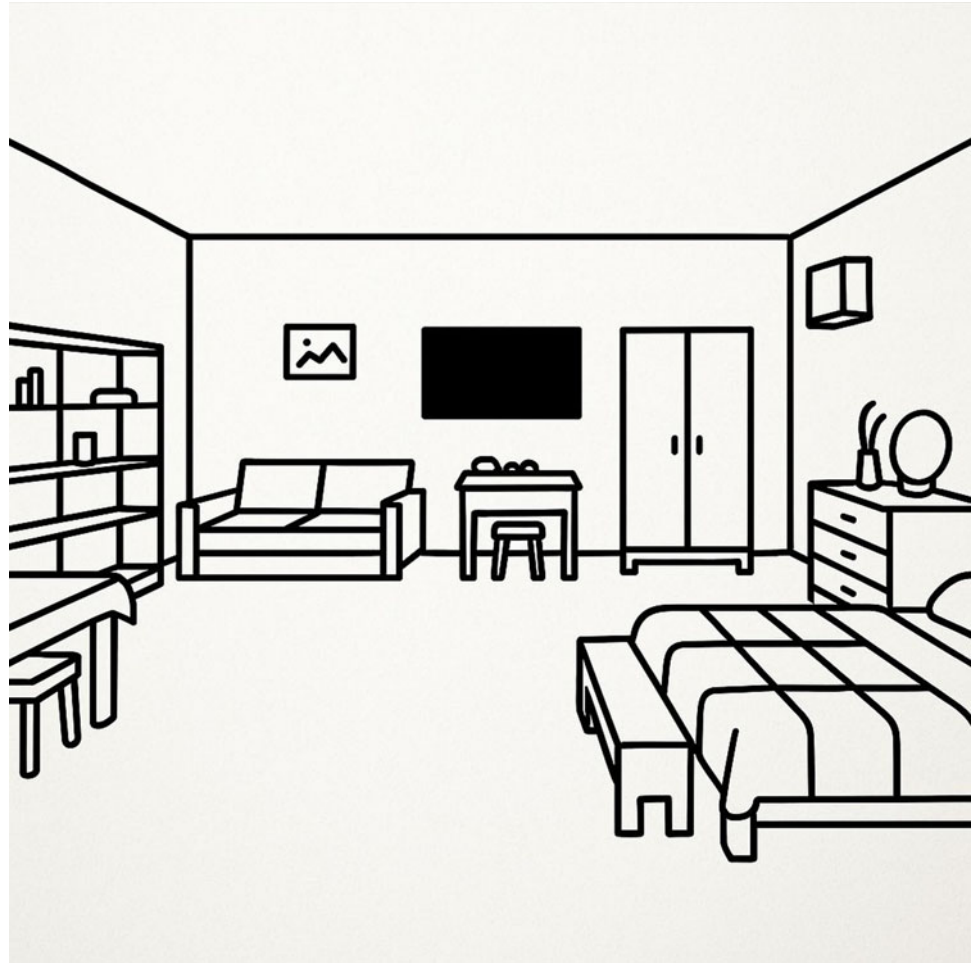


Figure 1. Sketch of the Suite virtual reality scenario.

- **Task 6. Recognition (Yes/No):** Images of furniture are presented, and the user must indicate whether they were part of the order. Targets were mixed with distractors present in the shop (seen but not ordered) and items never seen before. More specifically, a total of 18 items are presented, six of them being target stimuli, nine of them also being in the shop (but not being part of any order), and the other three are furniture items never presented before. The recorded data are (1) accuracy in recognizing the correct piece of furniture, (2) response time, and (3) false positives and negatives.
- **Task 7: Forced choice recognition task:** Each trial presented a studied target stimulus paired with a novel item from the same semantic category, equating familiarity and category cues while imposing a binary, comparative decision. More specifically, a total of 7 pairs of items are displayed (i.e., the user has to make seven decisions), seven of which are target stimuli and seven that have never been shown to the user before. Subsequently, the user must choose each one that was present. Under random responding, expected performance would obtain between 3 and 4 correct answers (out of 7). We used $\leq 2/7$ as a heuristic screen (together with ancillary indicators such as very short reaction times or inconsistent responding) to identify non-credible performance, and flagged cases were excluded. This task served as a performance validity screen and was not included in psychometric summaries of memory. The recorded data are: (1) selection accuracy, (2) response time, and (3) a discriminability index.

Previously performed reliability analyses (Climent *et al.*, 2024) showed excellent internal consistency for immediate, short, and long delay recall tasks (Cronbach's alpha of 0.93 for tasks 1, 3, and 5); moderate for source memory (Cronbach's alpha of 0.64 for task 2) and good for recognition (Cronbach's alpha of 0.79 for task 6). Reliability analyses for Task 4 showed very poor results (with less than 10% of individuals from the whole normative sample remembering to switch off the light, which led to the decision to exclude this task from the normative data), and items from Task 7 (i.e., forced choice recognition, merely used as a task to estimate performance validity and/or malingering/simulation) were not considered. Construct validity (PCA) is also detailed in the Suite normative study (Climent *et al.*, 2024).

The study also considers two other variables: the age and gender of the examinees, as Suite reveals differences in the behavior of the subjects according to age group (12 – 12, 13 – 26, 27 – 44, 45 – 58, 59 – 85) and, in some groups, even according to gender (13 – 26 and 45 – 58).

Data analysis

The following diagram in Figure 2 shows an overview of the data analysis process performed.

The statistical analyses and data management were carried out with R version 4.4.1 through several key libraries (R Core Team, 2021). The *RVAideMemoire* package facilitated the execution of

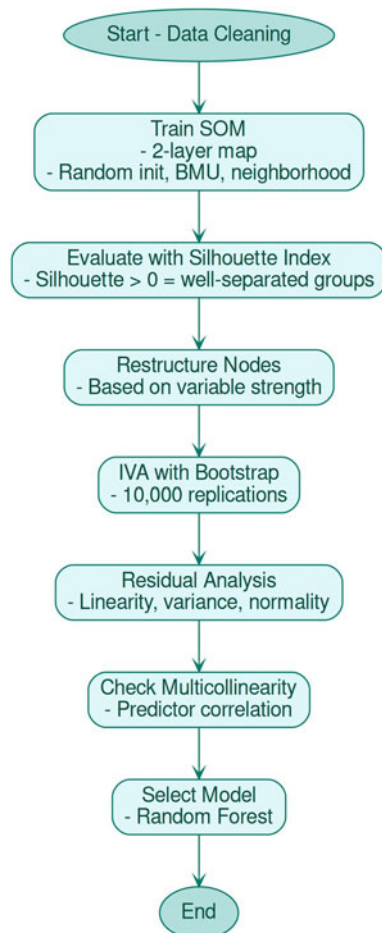


Figure 2. Graphical overview of data analysis performed. 1. Random init: Neuron weight vectors are initialized randomly, allowing the map to adapt to data without initial bias. 2. BMU (best matching unit): The neuron whose weights are most similar to the current input (usually by Euclidean distance); this “winner” is updated most strongly. 3. Neighborhood: Not only the BMU but also neighboring neurons are updated, with the degree of adjustment decreasing with distance from the BMU and over time. This preserves the topological structure of the map.

non-parametric statistical tests (HERVE, 2023), while **tidymodels** streamlined the modeling process for data consistency and efficiency (Kuhn & Wickham, 2020). The **kohonen** package was used to implement the Kohonen network algorithm (Wehrens & Buydens, 2007; Wehrens & Kruisselbrink, 2018). Additionally, the **boot** package enabled bootstrapping procedures (Canty & Ripley, 2024), and **randomForest** package was utilized for random forest analyses (Liaw & Wiener, 2002). The **cluster** package assisted in creating a silhouette index (Maechler et al., 2023), while **performance** package was employed to evaluate and diagnose statistical models (Lüdtke et al., 2021). For data visualization, **ggplot2** package was used to generate graphs (Wickham, 2016), complemented by the **magick** package for graph manipulation (Ooms, 2024). Data management and manipulation were conducted using **dplyr** package (Wickham et al., 2023), and the **tidylog** package provided insights into data transformations throughout the analysis (Elbers, 2024).

Self-Organizing Maps (SOM) were utilized to analyze memory behavior patterns during both general and scale group testing (Brereton, 2012; Oliver et al., 2018). SOM is an effective tool to visualize patterns in complex data, enabling the identification of

similar data clusters and their separation. The neuron architecture consists of two layers: the input layer that contains one neuron per input variable, and the competition layer, typically structured as a two-dimensional, low-dimensional topological grid. Each neuron in the input layer is connected to every unit in the competition layer, with each unit assigned to an N-dimensional weight vector. SOM training started with the random initialization of weight vectors assigned to each neuron. During the training phase, input data was presented as characteristic vectors that were compared to the weight vectors of all neurons within the map. The neuron that exhibited the closest match to a given input vector was designated as the “winner” or “best matching unit” (BMU). Subsequently, the BMU and its neighboring neurons updated their weight vectors to more closely align with the input vector, employing neighborhood functions and progressively decreasing learning rates. This iterative process was repeated across multiple epochs, incorporating the entire dataset. Upon completion of the training, the neurons were organized such that those with similar weight vectors were positioned adjacent to one another, thereby facilitating the visualization of patterns and the identification of clusters within the data.

Authors such as Tervonen et al. (2020) and Younger et al. (2024) performed a subsequent clustering of the nodes using k-means to identify similar nodes. Due to the limited number of variables (four types of memory processes, named here immediate recall, source memory, short-term delayed recall, and long-term delayed recall), the resulting profiles were easily interpretable, making further grouping unnecessary. To assess the quality of the generated clusters, the Silhouette index was calculated. This index measures the similarity of each object to other clusters, providing a quantitative evaluation of clustering effectiveness (Rousseeuw, 1987). In this context, the Silhouette index was applied to the clusters identified by the SOM algorithm.

Conducting several SOM analyses for the groups linked to the test (12 – 12, 13 – 26-M, 13 – 26-F, 27 – 44, 45 – 58-M, 45 – 58-F, 59 – 85) enables the monitoring of pattern evolution over time. To improve comparability, the nodes were restructured based on the strength of the defining variables. This approach ensured that the interpretation of patterns remained clear and was not influenced by the specific nodes where they appear.

After analyzing the memory profiles of each group, the source memory was subjected to an incremental validity assessment (IVA). It was used to determine whether new tests or measures provided significant additional information on a psychological variable beyond what was obtained with existing measurements. IVA is a statistical bootstrap technique that allows statistical distribution estimates (e.g., average, median, and regression coefficient) by repeating the original data with replacement (Davison & Hinkley, 1997). The statistic used is the determination coefficients (R^2), which measure the proportion of the predictable variance of independent variables in dependent variables. The difference between R^2 between the two models was calculated for more than 10,000 replications: the first model did not include source memory as an independent variable, while the second model did. Both models used long-term memory as a dependent variable because long-term memory was considered a successful fingerprint if the subject performed it correctly (Cotton & Ricker, 2022).

To ascertain the most suitable technique for the models, it was essential to conduct a study, at the very least in graphical form, on the distribution of the residuals to detect any unusual patterns. Furthermore, it was essential to evaluate the presence of

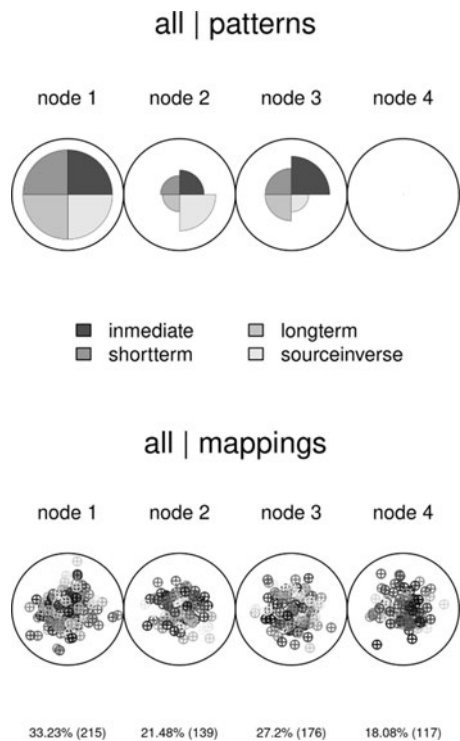


Figure 3. Profiles from the complete sample provided by SOM. Figure explanation: Each circle represents a node identified by the SOM, functioning as the centroid of a cluster. The “+” symbols indicate individual data points assigned based on similarity; the gray intensity may reflect different feature values. The proximity between symbols indicates greater similarity among observations. Percentages and numbers below each node show the proportion and number of data points assigned to each group relative to the total sample.

multicollinearity among the predictor variables, ascertaining the linear relationship between the predictor variables and the response variable, verifying whether the residuals had constant variance, and confirming whether the residuals followed a normal distribution (Osborne & Waters, 2019).

The choice of techniques depended on the results of the study. Statistical or regression models such as simple linear regression, general linear model, generalized least squares, additive model (AM), least absolute deviation (LAD), general AM, general linear mixed model (GLMM), general additive mixed model, are based on statistical and mathematical principles to model relationships between variables, or machine learning models, such as support vector machine, decision trees or random forests, which are more flexible and are used to capture complex patterns in data (Hastie *et al.*, 2017).

Results

Once the sample was obtained, the SOM algorithm was applied to the whole sample and each age group, resulting in the graphical representation of the SOM (see Figure 3 in the text and Supplemental Figure 1 in the Appendix).

In this graph, the patterns of memory behavior in the subjects can be visually verified. For example, Figure 1 shows that Node 1 is completed by 33.25% of the subjects whose four memories have a similar weight. In contrast, Node 2 is completed by 21.48% of the subjects, where the source memory has the highest weight, followed by the immediate recall, and the short-term and long-term delayed recall have a similar weight.

Table 2. Importance of variables in each node by group

group	A	B	C	D	node	pattern
12 – 12	0.9	0.8	0.8	0.9	1	A+B+C+D
	0.9	0.7	0.7	0.6	2	A+B+C
	0.8	0.6	0.6	0.8	3	A+D
	0.8	0.5	0.5	0.5	4	A
13 – 26-M	1.0	0.8	0.8	0.9	1	A+B+C+D
	0.9	0.7	0.6	0.8	2	A+B+D
	0.9	0.7	0.7	0.6	3	A+B+C
	0.9	0.5	0.5	0.4	4	A
13 – 26-F	1.0	0.8	0.8	0.9	1	A+B+C+D
	0.9	0.7	0.7	0.7	2	A+B+C+D
	0.9	0.6	0.6	0.6	3	A
	0.9	0.7	0.7	0.5	4	A+B+C
27 – 44	0.9	0.5	0.5	0.3	1	A
	0.9	0.5	0.5	0.6	2	A
	0.9	0.7	0.7	0.7	3	A+B+C+D
	1.0	0.8	0.8	0.8	4	A+B+C+D
45 – 58-M	0.9	0.7	0.7	0.8	1	A+B+C+D
	0.9	0.6	0.6	0.5	2	A
	0.8	0.5	0.5	0.6	3	A
	0.8	0.5	0.5	0.3	4	A
45 – 58-F	0.9	0.8	0.8	0.8	1	A+B+C+D
	0.9	0.6	0.6	0.7	2	A+D
	0.9	0.6	0.6	0.5	3	A
	0.8	0.5	0.5	0.2	4	A
59 – 85	0.9	0.7	0.7	0.8	1	A+B+C+D
	0.9	0.5	0.5	0.6	2	A
	0.8	0.4	0.5	0.3	3	
	0.4	0.3	0.4	0.2	4	
All	1.0	0.8	0.8	0.9	1	A+B+C+D
	0.9	0.6	0.6	0.8	2	A+D
	0.9	0.7	0.7	0.5	3	A+B+C
	0.8	0.5	0.5	0.3	4	A

Note: A (immediate recall), B (short-term delayed recall), C (long-term delayed recall), D (source memory).
M = male; F = female.

In node 3, completed by 27.2% of the subjects, immediate recall has the highest weight. Here, the relationship between short-term and long-term delayed recall is stronger than in Node 2, and the presence of source memory is lower than that of the other memories.

In node 4, which includes 18.08% of the subjects, all the memories are represented, but their strength in the distribution of the patterns in the other nodes is so low that their graphing is not observable. In order to improve the interpretation, these patterns have been tabulated (see Tables 2 and 3).

Data in Tables 2 and 3 illustrate how different memory types engage with various age groups, with younger individuals tending to exhibit stronger immediate and short-term delayed recall performance, while older adults may rely more on long-term recall and source memory. Before obtaining Table 2, it was confirmed that the identified nodes represented distinct groups. This was achieved by plotting the silhouette index for each group on the appropriate scale (see Figure 4 in the text and Supplemental Figure 2 in the Appendix). In all cases, a silhouette index value greater than zero is observed, which indicates that the groups are well delimited. However, some values are not well classified in terms of distance, suggesting the need to increase the number of clusters. Nevertheless, these subjects not classified within a group are insignificant compared to those belonging to the well-defined groups (see Supplemental Table 1 in the Appendix).

The graph illustrates the silhouette analysis for data clustering with four groups. Group 1 is the largest with 215 members (n1 = 215), followed by Group 3 (n3 = 176), Group 2 (n2 = 139), and Group 4 (n4 = 117). The silhouette width measures the similarity

Table 3. Memory patterns in the groups.

group	A+B+C+D	A+B+D	A+B+C	A+D	A
12 – 12	X		X	X	X
13 – 26-M	X	X	X		X
13 – 26-F	X		X		X
27 – 44	X				X
45 – 58-M	X				X
45 – 58-F	X			X	X
59 – 85	X				X
all	X		X	X	X

Note: A (immediate recall), B (short-term delayed recall), C (long-term delayed recall), D (source memory); M (male), F (female).

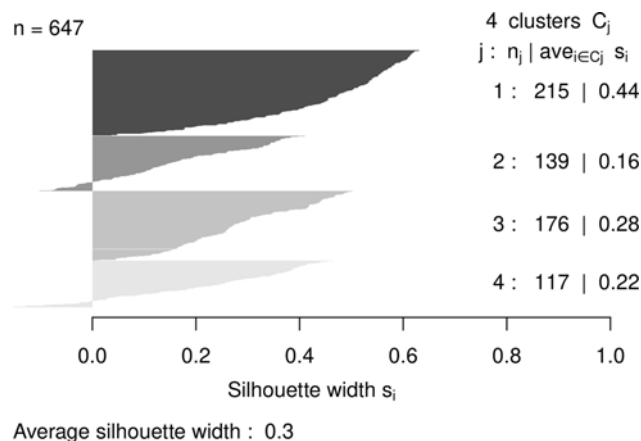


Figure 4. Silhouette indexes on the general sample. Figure explanation: Silhouette plot evaluating clustering quality. Each horizontal bar represents the silhouette width of an observation, grouped by cluster. Values near 1 indicate strong cohesion and clear separation between clusters; low or negative values suggest possible misclassifications or boundary points. Cluster sizes and averages are shown on the right.

of a data point to its own group compared to other groups. A higher value indicates better clustering. Therefore, Group 1 has the highest average silhouette width (0.44), suggesting it is well-separated from the other groups. Conversely, Group 2 has the lowest average silhouette width (0.16), indicating there may be some overlap with other groups. The overall clustering quality is moderate, suggesting that there may be room for improvement in the selection of the number of groups used, as previously explained.

After completing the study of memory profiles for each age group, the focus turned to conducting the IVA. Figure 5 in the text and Supplemental Figure 3 in the Appendix show that in none of the cases is there linearity, homogeneity of variance in the residuals, or normality. There is also no multicollinearity, as confirmed by the visual inspection of the model assumptions.

This analysis indicates that the model to be implemented in IVA should be a machine learning model. Random Forests algorithm was selected for its high accuracy and robustness to overfit, due to the combination of multiple trees. Moreover, it offers stability, is efficient in terms of training and prediction time, and provides a measure of the importance of the variables, which facilitates the interpretation of the model.

The non-parametric bootstrap analysis results presented in Table 4 reveal distinct patterns in memory performance across different age and gender groups. Notably, younger participants (ages 13 – 26) exhibit higher initial estimates along with greater imprecision and bias, indicating less stable memory performance. In contrast, the oldest age group (59 – 85) shows a negative original value and the lowest standard error, suggesting greater precision in

their memory estimates. Furthermore, significant differences between genders are particularly evident in the younger age group.

Discussion

The present study aimed to show how performance in a VR-based source memory task contributes to performance in immediate and delayed memory trials of the same test (the VR-based Suite test) across the different age groups included in the normative sample (Climent et al., 2024). The data presented here show the relevance of source memory for memory performance in individuals between 12 and 85 years old. The main variables of the tests are linked to immediate, short-term and long-term delayed recall memory processes, but instead of using interference tasks that may compromise performance (Brophy et al., 2009; Libon et al., 2011), the effects of a source memory task inserted between the immediate memory trials and the short-term delayed recall trial were investigated. While interference tasks are typically designed to assess susceptibility to forgetting and the ability to resist the intrusion of competing information, thereby draining working memory resources and measuring long-term consolidation, this new source memory task operates under a different neuropsychological principle. In Suite, participants are instructed to learn the association between each item and its corresponding ‘family’ (the contextual source). No manipulation or transformation of the context is requested. Thus, encoding is explicitly associative/contextual—via item–source binding—which may deepen the memory trace and facilitate retrieval. Neuropsychologically, this implies a greater emphasis on frontal lobe functions related to source monitoring and memory elaboration, and on hippocampal connections for the formation of item–context associations, in contrast to a primary focus on the also frontal but substantially distinct inhibition, or resistance to interference, that is key in traditional paradigms (Guo et al., 2021).

Results highlight how source memory contributes to memory performance when considered together with other memory types across age groups. Younger individuals tend to rely more on immediate and short-term delayed recall, while older adults exhibit a stronger reliance on source memory and long-term delayed recall, suggesting a shift from rapid recall in youth to more reflective processing in older age (Čepukaitytė et al., 2023). This evolution in memory aligns with theories of cognitive aging, where older adults use deeper processing strategies to elaborate and remember information (Craik et al., 1987; Light & Anderson, 1985; Salthouse et al., 1996). Additionally, older adults’ increased reliance on long-term memory supports previous research indicating more effective retrieval strategies over time (Park & Reuter-Lorenz, 2009). Gender differences were particularly noticeable in younger participants, emphasizing the complex interplay between age, gender, and memory. Incorporating source memory tasks into the assessment of memory performance provides further insight into long-term memory functioning across the lifespan, though it appears less relevant for 12-year-olds and females aged 45 – 58, where negative values were observed. These findings contribute to a deeper understanding of age-related memory changes and emphasize the crucial role of source memory in assessing memory across different demographics. As source memory seems to play a more prominent role in memory processes associated with older populations, a defective ability to rely on source memory may lead to a decrease in memory performance and thus be an important indicator of a decline in memory processes due to different etiologies. Subsequently, incorporating

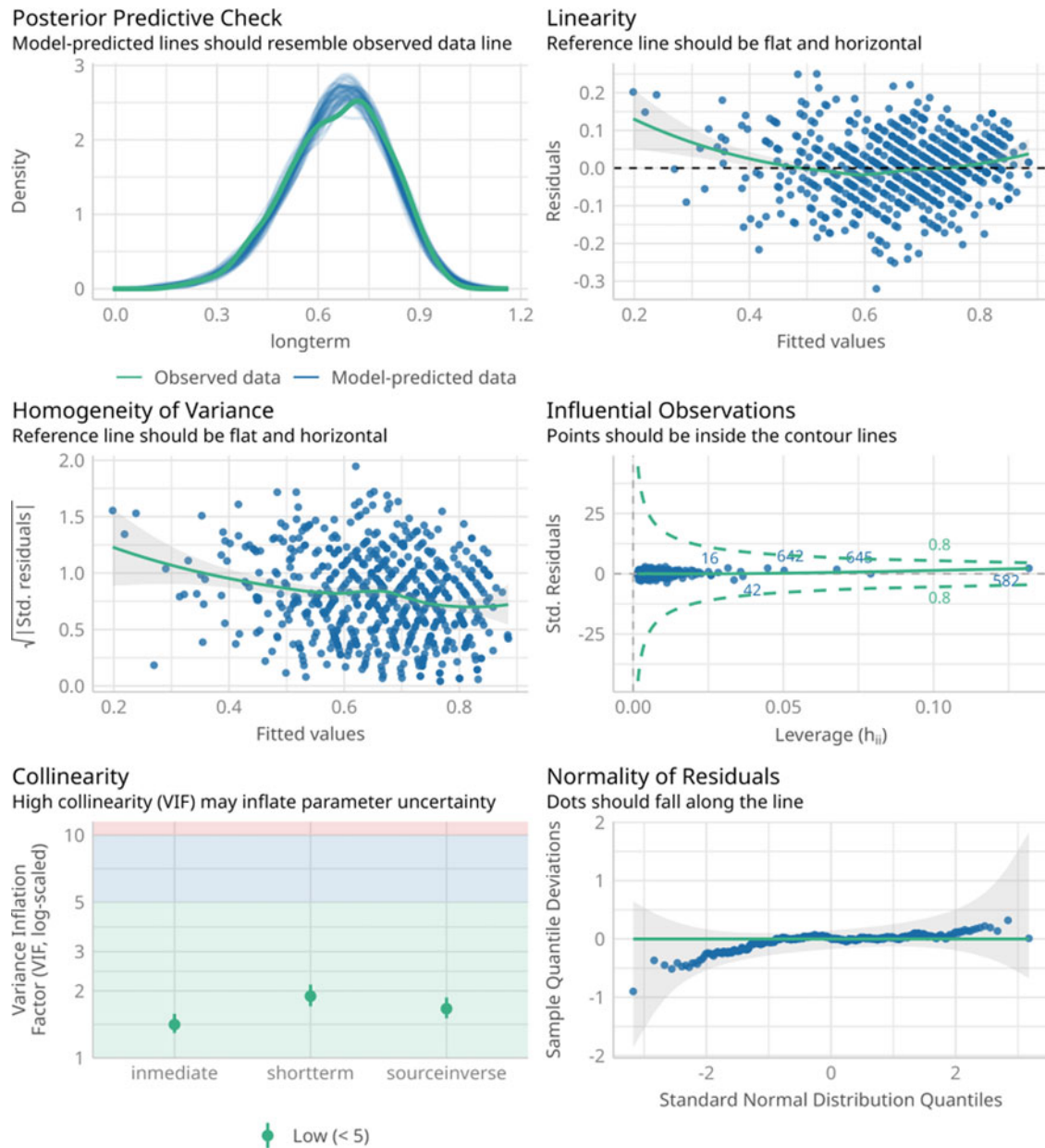


Figure 5. Memory patterns in the groups. Figure explanation: Diagnostic plots for assessing regression model assumptions and quality. Each panel displays a different aspect: fit between observed and predicted data, linearity, homogeneity of variance, presence of influential observations, collinearity among predictors, and normality of residuals. These plots help identify potential deviations from model assumptions and validate model adequacy.

assessment tasks to further evaluate the individual's ability to take advantage of source memory to boost their overall memory performance may help identify early signs of memory decline across different conditions.

The current study has obvious limitations. First, it is focused on a normative population from Spain. To enhance its generalizability, future research should involve cross-cultural validations, examining diverse populations in both community and clinical settings. Second, as literature has shown that selected clinical populations may not benefit from a memory assessment paradigm that includes interference tasks (Brophy *et al.*, 2009; Libon *et al.*, 2011), and since replacing it with source memory may potentially strengthen memory performance in certain age groups, a further confirmation of this hypothesis would require a massive administration of the test to larger samples in different clinical

settings. Third, the disparity in the number of participants in the composition of age groups and, especially, the small number of subjects in the 12-year-old age group may demand a careful interpretation of results for this age group. Finally, further studies with specific clinical populations would help to determine the sensitivity and specificity of the test.

Despite the limitations, these preliminary results on the impact of introducing source memory tasks instead of interference tasks may improve the way neuropsychological assessment of memory is performed, particularly in clinical populations where interference paradigms may exacerbate cognitive challenges. Unlike traditional interference-based approaches, which primarily highlight memory deficits, source memory tasks provide a richer understanding of both the content and context of memory retrieval, making them more versatile for diverse populations.

Table 4. Ordinary non-parametric bootstrap results by group.

group	Original	Bias	std error
12 – 12	0.009	–0.006	0.014
13 – 26-M	0.030	0.014	0.024
13 – 26-F	0.059	0.024	0.033
27 – 44	0.020	0.007	0.015
45 – 58-M	0.041	0.004	0.035
45 – 58-F	0.020	–0.007	0.012
59 – 85	–0.003	0.004	0.010
all	0.022	0.018	0.008

Note: M (male), F (female).

The novelty of the study is the introduction of a source memory task that requires the processing of contextually relevant information. This contextual processing does not seek to interfere with or create confusion in the respondent, but rather to delve deeper into the coding of the material to be remembered, which appears to improve subsequent recall, especially in certain age groups.

This distinction highlights the practical implications of the Suite Test: by integrating source memory tasks, clinicians and researchers can assess cognitive processes with greater ecological validity and adaptability. Moreover, these findings pave the way for designing interventions that leverage source memory to enhance encoding and retrieval strategies, particularly in populations affected by neurodegenerative conditions or age-related cognitive changes.

The findings of this study underscore the relevance of integrating source memory into neuropsychological assessments as an addition to interference paradigms. Interference tasks primarily assess how prior or subsequent information affects recall, often emphasizing memory limitations, particularly in clinical contexts. In contrast, source memory explores both the content and context of memory, providing a more comprehensive perspective.

In conclusion, this study emphasizes the potential of a new VR based memory test that includes a source memory task as a valuable tool in neuropsychological assessment, providing a pathway to explore diverse memory patterns across the lifespan, with the possibility to identify these patterns across clinical and neurotypical populations in upcoming research. Future studies need to emphasize its concurrent validity in comparison with other well-established tests of visual memory, as well as focus on the study of clinical samples that may add valuable information about diagnostic accuracy.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S1355617725101483>.

Author contribution. GC, UD, and FR worked in the conceptualization and planning of the paper. JC and UD wrote the first draft of the introduction, while GC, IACG, and FR worked in the methodology and results section. UD, IACG, and FR worked in the Discussion, and GC updated the introduction and discussion, as well as final refinements to the paper. All authors reviewed the final version of the paper.

Funding statement. This work was supported by the European Commission under Horizon 2020 Programme (Grant number 733901, from Project VRMIND – Virtual Reality Based Evaluation of Mental Disorders).

Competing interests. FR and IACG work in the R+D Department of Giunti-Nesplora, publisher of the test.

References

Beaumont, J. G., & Davidoff, J. B. (2019). Assessment of visuo-perceptual dysfunction. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A Handbook of Neuropsychological Assessment* (pp. 115–140). : Routledge.

- Brereton, R. (2012). Self organising maps for visualising and modelling. *Chemistry Central Journal* 6, 6, S1.
- Brophy, L. M., Jackson, M., & Crowe, S. F. (2009). Interference effects on commonly used memory tasks. *Archives of Clinical Neuropsychology*, 24, 105–112.
- Canty, A., & Ripley, B. D. (2024). Bootstrap R (S-Plus) functions. In *En Handbook of Statistical Analysis* (5th ed. pp. 105–120). Cambridge University Press.
- Ceccato, I., La Malva, P., Di Crosta, A., Palumbo, R., Gatti, M., Momi, D., Logrieco, M. G. M., Fasolo, M., Mammarella, N., Borella, E., & Di Domenico, A. (2022). When did you see it? the effect of emotional valence on temporal source memory in aging. *Cognition and Emotion*, 36, 987–994.
- Ceccato, I., Ricci, E., Mazza, C., Bartolini, E., Di Crosta, A., La Malva, P., Biondi, S., Colasanti, M., Mammarella, N., Palumbo, R., Roma, P., & Di Domenico, A. (2024). Influence of stimuli emotional features and typicality on memory performance: Insights from a virtual reality context. *Psychological Research*, 88, 257–270.
- Čepukaitytė, G., Thom, J. L., Kallmayer, M., Nobre, A. C., & Zokaei, N. (2023). The relationship between short- and long-term memory is preserved across the age range. *Brain Sciences*, 13, e106.
- Climent, G., Rebon-Ortiz, F., Saura-Carrasco, M., & Diaz-Orueta, U. (2024). Suite test. A virtual reality based neuropsychological assessment tool for memory. A normative study on immediate, short-term and delayed memory tasks. *Applied Neuropsychology: Adult*, 1–11. <https://doi.org/10.1080/23279095.2024.2306128>.
- Cotton, K., & Ricker, T. J. (2022). Examining the relationship between working memory consolidation and long-term consolidation. *Psychonomic Bulletin & Review*, 29, 1625–1648.
- Craik, F. I. M., Byrd, M., & Swanson, J. M. (1987). Patterns of memory loss in three elderly samples. *Psychology and Aging*, 2, 79–86.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap Methods and Their Applications*. Cambridge University Press, <https://doi.org/10.1017/CBO9780511802843>
- Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *Manual for the California Verbal Learning Test, Second Edition (CVLT-II)*. The Psychological Corporation.
- Diaz-Orueta, U., Blanco-Campal, A., Lamar, M., Libon, D. J., & Burke, T. (2020). Marrying past and present neuropsychology: Is the future of the process-based approach technology-based? *Frontiers in psychology*, 11, e361.
- Elbers, B. (2024). tidylog: Logging for “dplyr” and “tidyr” functions. <https://CRAN.R-project.org/package=tidylog>.
- Guo, J., Shubeck, K., & Hu, X. (2021). Relationship between item and source memory: Explanation of connection-strength model. *Frontiers in Psychology*, 12, e691577.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- Herve, M. (2023). RVAideMemoire: Testing and plotting procedures for biostatistics. <https://CRAN.R-project.org/package=RVAideMemoire>.
- Kuhn, M., & Wickham, H. (2020). Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles. <https://www.tidymodels.org>.
- Lezak, M. D., Howieson, D. B., Loring, D. W., & Fischer, J. S. (2012). *Neuropsychological Assessment* (5th edn.), Oxford University Press.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22, <https://CRAN.R-project.org/doc/Rnews/>
- Libon, D. J., Bondi, M. W., Price, C. C., Lamar, M., Eppig, J., Wambach, D. M., Nieves, C., Delano-Wood, L., Giovannetti, T., Lippa, C., Kabasakalian, A., Cosentino, S., Swenson, R., & Penney, D. L. (2011). Verbal serial list learning in mild cognitive impairment: A profile analysis of interference, forgetting, and errors. *Journal of the International Neuropsychological Society*, 17, 905–914.
- Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology*, 40, 737–747.
- Lüdecke, D., Ben-Shachar, M. S., Patil, I., Waggoner, P., & Makowski, D. (2021). Performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, 6, e3139.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2023). cluster: Cluster Analysis Basics and Extensions. <https://CRAN.R-project.org/package=cluster>.

- McGeoch, J. A. (1932). Forgetting and the law of disuse. *Psychological Review*, 39, 352–370.
- Minor, G., & Herzmann, G. (2019). Effects of negative emotion on neural correlates of item and source memory during encoding and retrieval. *Brain Research*, 1718, 32–45.
- Oliver, E., Vallés-Pérez, I., Baños, R.-M., Cebolla, A., Botella, C., & Soria-Olivas, E. (2018). Visual data mining with self-organizing maps for “self-monitoring” data analysis. *Sociological Methods & Research*, 47, 492–506.
- Ooms, J. (2024). magick: Advanced graphics and image-processing in R. <https://CRAN.R-project.org/package=magick>.
- Osborne, J. W., & Waters, E. (2019). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research, and Evaluation*, 8, 2.
- Palombo, D. J., Te, A. A., Checknita, K. J., & Madan, C. R. (2021). Exploring the Facets of Emotional Episodic Memory: Remembering “What,” “When,” and “Which”. *Psychological Science*, 32, 1104–1114.
- Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology*, 60, 173–196.
- Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience*, 9, e660.
- Pieri, L., Tosi, G., & Romano, D. (2023). Virtual reality technology in neuropsychological testing: A systematic review. *Journal of Neuropsychology*, 17, 382–399.
- R Core Team (2021). R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*, <https://www.R-project.org/>.
- Rahimi-Golkhandan, S., Maruff, P., Darby, D., & Wilson, P. (2012). Barriers to repeated assessment of verbal learning and memory: A comparison of international shopping list task and rey auditory verbal learning test on build-up of proactive interference. *Archives of Clinical Neuropsychology*, 27, 790–795.
- Rey, A. (1942). L'examen psychologique dans les cas d'encephalopathie traumatique. *Archives of Psychology*, 28, 112.
- Rizzo, A. A., Schultheis, M., Kerns, K. A., & Mateer, C. (2004). Analysis of assets for virtual reality applications in neuropsychology. *Neuropsychological Rehabilitation*, 14, 207–239.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Salthouse, T. A., Fristoe, N., & Rhee, S. H. (1996). How localized are age-related effects on neuropsychological measures? *Neuropsychology*, 10, 272–285.
- Sherman, E., & Hrabok, M. (2023). *A compendium of neuropsychological tests: Fundamentals of neuropsychological assessment and test reviews for clinical practice*. Oxford University Press.
- Squire, L. R., Wixted, J. T., & Clark, R. E. (2007). Recognition memory and the medial temporal lobe: A new perspective. *Nature Reviews Neuroscience*, 8, 872–883.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Oxford University Press.
- Symeonidou, N., & Kuhlmann, B. G. (2021). A novel paradigm to assess storage of sources in memory: The source recognition test with reinstatement. *Memory*, 29, 507–523.
- Tervonen, J., Puttonen, S., Sillanpää, M. J., Hopsu, L., Homorodi, Z., Keränen, J., Pajukanta, J., Tolonen, A., Lämsä, A., & Mäntyjärvi, J. (2020). Personalized mental stress detection with self-organizing map: From laboratory to the field. *Computers in Biology and Medicine*, 124.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, 64, 49–60.
- Wehrens, R., & Buydens, L. M. C. (2007). Self- and super-organizing maps in R: The kohonen package. *Journal of Statistical Software*, 21, 1–19.
- Wehrens, R., & Kruisselbrink, J. (2018). Flexible self-organizing maps in kohonen 3.0. *Journal of Statistical Software*, 87, 1–18.
- Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag. https://doi.org/10.1007/978-3-319-24277-4_9. <https://ggplot2.tidyverse.org>.
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>. <https://dplyr.tidyverse.org>. <https://github.com/tidyverse/dplyr>.
- Younger, J. W., Schaerlaeken, S., Anguera, J. A., & Gazzaley, A. (2024). The whole is greater than the sum of its parts: Using cognitive profiles to predict academic achievement. *Trends in Neuroscience and Education*, 36, e100237.