

Exploring Memories of Everyday Life: A Validation Study of L-RECAP

Running title: L-RECAP validation

Maud Billet^{1,2*}, Lisa Dussard¹, Charline Colson¹, Marie Geurten^{1,2†}, & Sylvie Willems^{1†}

¹Department of Psychology, Psychology and Neuroscience of Cognition (PsyNCog) Research Unit, University of Liège, 4000 Liège, Belgium

²Fund for Scientific Research (F.R.S. – FNRS), 1000 Bruxelles, Belgium

[†]Jointly last author

ORCID:

Maud Billet: 0000-0003-3067-0091

Charline Colson: 0009-0004-1450-9921

Marie Geurten: 0000-0003-3743-5246

Sylvie Willems: 0000-0001-5924-0680

***Corresponding author:**

Maud Billet

Email: maud.billet@uliege.be

Phone number: +32 4 366 59 43

Address: B33 – Quartier Agora, Place des Orateurs 1, 4000 Liège, Belgium

Supplementary material: Figure S1. Screenshots of the questions sent on the mobile app; Table S1. Example of specificity scoring; Table S2. Demographic information across studies; Table S3. Detailed descriptions of the different neuropsychological tasks and memory questionnaires; Table S4. Scoring grid for the recall of the previous testing session.

Abstract

Objective. In six studies of healthy individuals, we examined the feasibility, validity, and reliability of L-RECAP, a new ecological measure of memory functioning based on the sampling of everyday experiences.

Method. Participants received five notifications per day for seven days via a mobile app, asking about their ongoing activities. Two days after the last notification, the accuracy, richness, specificity, and phenomenology of their memory for five of these activities were assessed with verbal recall. L-RECAP's feasibility and acceptability were examined through the notifications response rate and explicit feedback ($n = 135$; S1). Correlations with traditional neuropsychological assessments ($n = 79$; S2) and participants' memories of a previous session ($n = 20$; S3) were used to document convergent and divergent validity. Predictive validity was assessed using self-report memory questionnaires ($n = 135$; S4) and discriminant validity was tested by comparing individuals who were satisfied vs. dissatisfied with their memory functioning ($n = 56$; S5). Finally, test-retest reliability over six weeks was examined ($n = 30$; S6).

Results. Participants reported good acceptability for L-RECAP. It correlated moderately with standard memory tests (r from .29 to .50) and self-report memory measures (r from .19 to .38). The specificity score correlated strongly with recall of a previous session ($r = .54$). The accuracy score distinguished between satisfied from dissatisfied participants ($d = .34$) and the task showed satisfactory test-retest reliability.

Conclusions. Although further validation in clinical populations is needed, L-RECAP appears to be a feasible, reliable, and ecologically valid measure of everyday memory functioning.

Keywords: Episodic memory; event memory; neuropsychological assessment; ecological validity

Introduction

It is challenging to predict memory functioning in everyday life using the validated neuropsychological memory tests that are currently used in clinical practice (Chaytor & Schmitter-Edgecombe, 2003; Spooner & Pachana, 2006). Indeed, it is not uncommon for individuals who experience difficulties in remembering everyday events to obtain normal scores on traditional memory tasks (Milton et al., 2021; Palombo et al., 2015). This phenomenon may be explained, at least in part, by the fact that several processes known to be involved in the memory of everyday events are neglected, or even completely absent, from traditional evaluations.

An important characteristic of episodic memory is that it concerns personal and complex events experienced in a continuous flow. To be encoded in memory, these experiences must first be segmented into discrete meaningful units (for a review, see Zacks, 2020), suggesting that the assessment of everyday memory should ideally involve continuous events rather than already segmented materials. Remembering such events also requires the integration of multiple sources of information – including actions, spatial and temporal context, perceptual details, thoughts, and emotions – into coherent memory representations (Moscovitch et al., 2016). These associative memories are themselves integrated into broader autobiographical knowledge that may serve as frameworks for reconstructing past events (Conway, 2001; Grilli & Sheldon, 2022), thus underscoring the importance of assessing memories of personal experiences rather than material irrelevant for the self. In addition, the consolidation of such episodic memories unfolds over extended periods of time (Frankland & Bontempi, 2005; Squire et al., 2015). Unfortunately, most standard memory tests only use a 20- to 30-minute delay between learning and delayed recall, making it unlikely to detect consolidation deficits that may emerge days after initial successful encoding (Landry et al., 2022).

Several approaches have addressed some of these limitations, although each tends to capture only specific aspects of everyday episodic memory or introduces additional practical constraints. For example, wearable cameras paradigms allow the assessment of memory for naturalistic events experienced in a continuous flow but their use in clinical settings raises practical and ethical questions (Allé et al., 2017). Virtual reality (VR) paradigms can be easier to implement while immersing participants into continuous, dynamic, and multisensory contexts (Plancher & Piolino, 2017). Unfortunately, cybersickness is frequently reported (for reviews, see Caserman et al., 2021; Ramaseri Chandra et al., 2022) and they do not allow assessment of personal memories. This is also the case of other ecological assessments that incorporate everyday-like activities (e.g. Beaver & Schmitter-Edgecombe, 2017). Even if that kind of ecological task showed predictive value for both self-reported functional status and memory abilities, they rely on standardized activities proposed by the experimenter and do not really capture memory for events that naturally unfold in daily life.

Autobiographical memory interviews (e.g., Kopelman et al., 1989; Levine et al., 2002; Piolino et al., 2003) offer the advantages of targeting memories of personal real-life events, integrated into general conceptual knowledge. However, such a measure generally does not account for the accuracy of the reported information and often focuses on remote and already consolidated memories. Conversely, other studies have examined consolidation by extending retention intervals of traditional memory tasks but these tests still focus on relatively simple and decontextualized materials (e.g., Mameniskiene et al., 2006; Visser et al., 2019). Consequently, although existing approaches can capture some important aspects of everyday memory functioning, no single tool combines the assessment of continuous, personally experienced, and contextually rich events with measures of memory accuracy over retention intervals that are more representative of real-world memory demands, while maintaining clinical feasibility.

Another essential aspect inherent to episodic memory, yet often overlooked in its assessment, is phenomenology, the subjective experience accompanying the retrieval of personal memories (Klein & Nichols, 2012; Tulving, 1985). Phenomenological characteristics such as vividness, reliving, coherence, or visual imagery were identified as central components of autobiographical remembering by Berntsen et al. (2019). Such subjective feelings are used to make memory decisions (*e.g.*, “*I don’t need to check, I have a vivid memory of that day*”; Johnson, 2006) and have been linked to well-being (Werner-Seidler & Moulds, 2011). However, they are rarely assessed in neuropsychological memory tests and, when they are, the focus is typically restricted to a limited part of the broader subjective experience (*i.e.*, most often vividness or reliving while other dimensions remain largely neglected; *e.g.*, Levine et al., 2002; Piolino et al., 2003).

Considering these current conceptions, this study aimed to introduce a new neuropsychological task that captures key aspects of real-life event memory: L-RECAP (Liège – Recall of Events via Continuous Assessment of Personal experiences). Through the sampling of everyday events via a mobile application, this episodic memory task has been designed to assess memories of several daily events, which are personally experienced in a continuous flow, encompass a large amount of contextual and associative information, occur on different days (and thus, must be retrieved after different delays), and are accompanied by a subjective experience of remembering. Importantly, L-RECAP was also specifically conceived to be easily applicable in clinical settings, with minimal constraints for both participants and practitioners. Our goal was to integrate into a single tool the qualities of autobiographical interviews – which can assess the richness, specificity and subjective quality of personal memories – and laboratory-based tasks – which evaluate the accuracy of the information reported. Several studies have already proposed a memory evaluation based on the sampling of everyday experiences, but they tend to assess only specific aspects of

memory, such as the recall of the “when”, “where”, or affects (Laliberte et al., 2021; McKenzie et al., 2020; Yim et al., 2018). To our knowledge, L-RECAP is the first instrument to propose a more comprehensive assessment of everyday memory functioning.

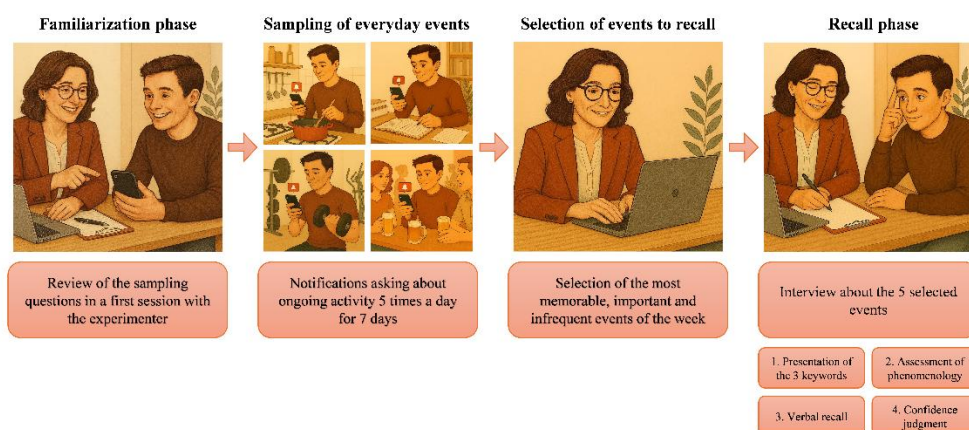
In this paper, our aim was to investigate the feasibility, validity, and reliability of L-RECAP in six studies conducted in healthy individuals, an important first step before considering its use in clinical settings. Before describing the specific objectives of each study, we first describe the L-RECAP memory task.

Description of L-RECAP

L-RECAP administration and scoring manuals, along with normative data, can be requested by emailing the authors. Although the present validation paper concerns the French version of the task, an English version that has not yet been validated is available upon request. Authors will also provide guidance on translation procedures for other languages.

The administration of L-RECAP followed four main phases (see Figure 1): (1) the familiarization phase, during which participants were introduced to the task and taught how to use the mobile application; (2) the sampling of everyday events lasting one week; (3) the experimenter’s selection of the events to recall; and finally (4) the recall phase, during which participants were interviewed about the selected events from their week.

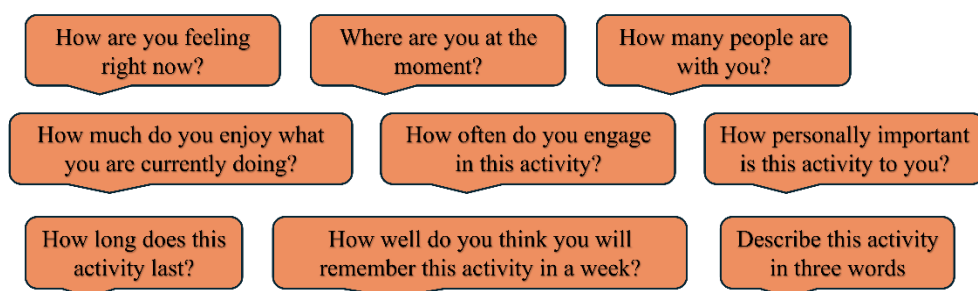
Figure 1. Illustration of the -RECAP procedure



Familiarization Phase

L-RECAP allows the assessment of memories of daily events using the mobile application m-Path* (Mestdagh et al., 2023). During an initial session, participants were invited to install this application on their smartphone and reviewed the questions with the experimenter. They were told they would have to answer the same questions five times a day, every day for seven days. These questions concerned the activity the participants were engaged in at the time, their mood, where they were, and who was present, but also their appreciation of the event, its frequency, importance, duration, and memorability. They also had to describe the activity in three words, which served as retrieval cues for the recall phase. All the questions are illustrated in Figure 2 (see Figure S1 in the supplementary material for screenshots of the app interface).

Figure 2. L-RECAP Sampling Questions on Everyday Events



Sampling of Everyday Events

The day after the familiarization session, individuals started to receive notifications, each including the same questions presented in Figure 2. They were scheduled between the individual's wake-up and bedtime hours and were sent at semi-random intervals – there must be at least one hour between notifications to increase the likelihood that they concern different activities. The date and exact time of response to each notification were recorded on m-Path.

* The task can be adapted to other experience sampling platforms.

Selection of the Events to Recall

At the end of the week, five events were selected for the recall phase based on a composite score averaging participants' ratings of rarity, importance, and memorability. The highest-scoring event from each day was retained, yielding five events that represented the least frequent, most important, and most memorable experiences reported during the week. This procedure minimized the inclusion of trivial or mundane events and enhanced the relevance of the selected experiences for the assessment of memory functioning. Importantly, participants were not informed of the event selection criteria or the number of events they would be interviewed about during the recall phase, in order to avoid intentionality bias whereby participants might selectively respond based on which events they wished to recall later.

Recall Phase

The recall phase ideally took place two days after the last notification (ranging from one to three days later), giving the experimenter sufficient time to select the five events to interview before meeting the participant again. The general structure of this recall phase is summarized in Figure 1. For each event selected, individuals were presented with the three words they had generated to describe this event in the sampling questions (i.e., “*Do you remember the activity you described with the three words: word1, word2, word3?*”).

Once participants remembered which event these words referred to, they had to rate the phenomenological quality of this specific memory before recalling the event aloud. More specifically, they had to use visual analog scales ranging from 0 (not at all) to 100 (very strongly) to rate the vividness, coherence, reliving, rehearsal, scene, and visual details of this memory (see Table 1). The choice of these phenomenological questions was based on the items that loaded the best on the factors of the French version of the Autobiographical

Recollection Test (ART; Billet et al., 2023), developed to assess the phenomenology of autobiographical memories, which were adapted here to assess specific memories.

Table 1. Phenomenological Assessment of L-RECAP

Factor	Item
Vividness	My memories of this event are vivid.
Coherence	My memories of this event are coherent, not disconnected fragments.
Reliving	When I remember this event, it's as if I'm reliving it.
Rehearsal	I've thought about this event, reflected on it, or talked about it.
Scene allocentric	When I remember this event, I recall the location of actions, objects, and people.
Scene egocentric	When I remember this event, I can situate myself in relation to the various things that were there.
Visual details	My memories of this event include clear visual details.

After the phenomenological assessment, individuals were asked to recall the event in as much detail as possible. To guide their report, they were given a piece of paper showing the information they should include in their recall: “What happened?”, “When?”, “Where?”, “With whom?”, “Perceptual details?”, and “Internal details (mood, thoughts, feelings)?”. If one of these expected details was not reported spontaneously, the experimenter provided additional prompts (e.g., *Could you specify when this event occurred?*). Participants were also presented with the following example to guide their recall: *“For the words “bakery – friend – chocolate”, it reminds me that on Thursday afternoon (when), I went with my friend Madeleine (who) to have a coffee (what happened) at a bakery located in the street behind her house (where). As we entered, we saw the display case filled with a wide variety of*

pastries (visual detail). I particularly remember being drawn to the colorful macarons, which included green, pink, brown, and orange varieties (visual details). After considering the options, I chose the chocolate macarons, which I found delicious (taste detail). The bakery was quite busy, and there was a lot of noise throughout the room (auditory detail). I remember wondering why so many people were there that afternoon (thought). I also felt very happy to spend time with my friend, whom I had not seen for quite some time (emotion).” This verbal recall session was recorded for subsequent coding. At the end of their report, individuals had to provide a confidence judgment about the accuracy of their memory (*How confident are you in the accuracy of what you have just described?*) on a visual analog scale ranging from 0 (not at all) to 100 (very strongly).

This procedure was then repeated for the four remaining events. If a participant could not retrieve an event using the three cue words after one minute, the experimenter provided the specific day on which the event occurred. If this temporal cue led to the retrieval of the event, the procedure continued as described above with the phenomenological assessment, verbal recall, and confidence judgment. If the participant still could not retrieve the event after the temporal cue, it was considered unrecovered, a score of zero was awarded, and the experimenter went on to the next event.

Scoring

Four main scores were calculated: accuracy, specificity, richness, and phenomenology. For each of these four scores, we averaged the five events interviewed.

The **accuracy score** was determined by comparing whether the (1) day of the event, (2) time of day, (3) location, (4) people present, and (5) mood the participant recalled matched the information recorded in the mobile application. Each correctly reported piece of information received one point, so individuals initially received a 5-point score per event. Then, a proportional accuracy score was calculated by dividing the number of correctly

recalled details by the total number of scorable items, resulting in a value between 0 and 1 (e.g., a score of 0.8 for the correct recall of 4 out of the 5 pieces of information requested).

The **specificity score** measures the episodicity of the events recalled. The precision of the reported information about time, location, perceptual details, and internal details was evaluated on four scales ranging from 0 to 3. This scoring system was inspired by the one used by Levine et al. (2002), initially developed to assess specificity of remote autobiographical memories, but adapted here for memories of recent events with criteria reflecting their greater temporal precision and richer episodic content. A detailed description of the scoring system is provided in Table 2. Three points corresponded to a rich and very specific description. Two points corresponded to a less detailed description. One point represented a description that was limited to general, non-specific information. No points were allocated when no episodic information was mentioned. The specificity score obtained for each of the four pieces of information (i.e., time, location, perceptual details, and internal details) was then averaged to obtain a global specificity score for each event. An example of specificity scoring is provided in supplementary material (Table S1).

Table 2. Specificity Scoring Grid Inspired by Levine et al. (2002)

Spatial information	0 = no information provided 1 = vague location 2 = building mentioned 3 = room or location in room mentioned
Temporal information	0 = no information provided 1 = vague response (no mention of the day or time of day) 2 = day or time of day mentioned

	3 = day and time of day reported
Perceptual details	0 = no information provided 1 = maximum 3 details from the same modality but without a detailed description 2 = (a) minimum 2 details from the same modality with a detailed description, (b) more than 3 details from the same modality but without a detailed description, or (c) several details from different modalities without a detailed description 3 = several details from different modalities with a detailed description
Internal states	0 = no information provided 1 = only one thought or feeling reported 2 = several thoughts/feelings expressed but not explained 3 = several thoughts expressed, and the reasons for their occurrence explained

The **episodic richness score** corresponds to the number of details reported about the events. To calculate this score, individuals' verbal recall was recorded, transcribed, and coded following Levine et al.'s (2002) scoring grid. Two broad categories of details were distinguished: internal versus external. Internal details pertained directly to the main event examined; they were specific in time and place, and thus reflected another measure of episodicity. External details referred to information about other events, semantic information, metacognitive statements, or repetitions. Descriptions of the different subcategories of information are reported in Table 3. In this paper, we focus on two scores: (1) internal richness (used to reflect the episodic richness of the specific events covered), corresponding

to the sum of all the internal details reported; and (2) total richness, corresponding to the sum of all the internal and external details reported.

Table 3. Richness Scoring Grid Inspired by Levine et al. (2002)

Category	Information
Internal	
Event	Happenings, individuals present, weather conditions, physical/emotional actions, or reactions of others
Time	Week, day, time of day
Place	Localization of the event including the city, street, building, room, part of room
Perceptual	Auditory, olfactory, tactile, taste, and visual details, body position, duration
Thought/emotion	Emotional states, thoughts, implications
External	
Event	
Time	
Place	Specific details from other incidents, external to the main event recalled
Perceptual	
Thought/emotion	
Semantic	General knowledge or facts, ongoing events, extended states of being
Metacognition	Metacognitive statements
Repetitions	Unsolicited repetition of details

Finally, **the phenomenology score** of each event was calculated by averaging the ratings provided by the participant for the six items of the phenomenological assessment (see Table 1).

Objectives of the different studies

In Study 1, data were first collected to examine the feasibility of L-RECAP through the ability to use the app, response rate, and number of retrieved events. In Study 2, we documented its convergent and divergent validity by examining its relations with traditional memory and cognitive tasks and, in Study 3, with participants' memory for a standardized past event. Predictive validity was then determined in Study 4 by investigating correlations between L-RECAP scores and participants' self-reports of memory functioning. Discriminant validity was investigated in Study 5 by determining the task's ability to detect memory problems in healthy individuals reporting memory difficulties. Finally, Study 6 documented its test-retest reliability.

For these purposes, the L-RECAP task was incorporated into several research protocols at our lab, which were initially designed to address broader objectives than the L-RECAP validation. Here, data from these different samples are reported according to the specific psychometric properties examined. This approach accounts for the varying sample sizes reported across studies, the inclusion of some participants in multiple data subsets, and the use of different memory measures for comparison. A summary of the demographic information from the different studies is provided in supplementary material (Table S2). All studies received approval from the Ethics Committee of the Faculty of Psychology, Speech Therapy, and Education Sciences at the University of Liège (2122-080), and the Liège University Hospital Ethics Committee (2022/293; 2019/253).

Study 1: L-RECAP's Feasibility

Objectives

This first study aimed to determine the acceptability and feasibility of L-RECAP in healthy subjects by assessing participants' ability to (a) use the mobile application through which the notifications were sent; (b) respond to a sufficient number of notifications to allow

us to select five relevant events (one per day); and (c) retrieve an adequate number of these selected events. To do this, participants' feedback on the task and application usage, their response rate to the notifications, and their retrieval rate during recall were collected, and potential age-related effects were explored.

Materials and Methods

Participants

A total of 135 native French-speaking participants (73 females, 62 males), aged between 19 and 80 years ($M = 40.36$, $SD = 16.33$), were included in this study. Participants had a mean of 14.96 years of education ($SD = 2.19$) and reported no history of neurological or psychiatric conditions. Recruitment was conducted via word of mouth, posters, and social media advertisements. Given the descriptive nature of the analyses planned, no a priori power analysis was conducted. All available individuals who wanted to participate and who met the inclusion criteria were therefore included in the study in order to obtain the largest possible sample. Correlational analyses with age were subsequently added as exploratory analyses. A post hoc sensitivity analysis was performed using G*Power and indicated that the sample size was sufficient to reveal medium effect sizes ($r = .21$).

Procedure

The procedure consisted of the full administration of L-RECAP, which required two appointments. The first session began with obtaining informed consent and collecting demographic information, followed by the familiarization phase in which participants installed the application on their smartphones, and the experimenter guided them through the sampling questions. This session lasted approximately 10 minutes and could be conducted in person or via video-conference, depending on participants' preference. Participants then received the five daily notifications for seven days. Approximately two days after the final notifications (range: 1–3 days), they completed the recall phase and provided their feedback

by answering a questionnaire at the end of the session. This second session lasted between 15 to 25 minutes, depending on the length of participants' recalls. Again, this second session could be conducted in person or via video-conference, depending on participants' preference.

Measures

After completing the L-RECAP sampling of everyday events, we examined the response rate to the notifications to ensure that participants provided enough responses to select five events occurring on different days. We then administered the recall phase and documented the number of events retrieved.

Feedback on L-RECAP

After the recall phase, participants were asked to answer eight questions assessing their experience with the task. Each question was answered using a visual analog scale ranging from 0 to 100. The first question concerned the difficulties encountered when using the application (*How many practical difficulties did you encounter while using the application?*). Three questions concerned the intrusiveness of the task (*How much did responding to notifications influence (1) your mood? (2) your social interactions? (3) your activities?*). Three other questions focused on the number of interactions (*What did you think of the number of (1) notifications per day? (2) questions per notification? (3) days you received notifications?*). The last question concerned the difficulty of the recall phase (*How difficult did you find the memory task?*).

Statistical Analyses

Descriptive statistics were used to examine the response rate throughout the week, the number of events retrieved, and participants' feedback on L-RECAP. In addition, we conducted exploratory correlational analyses to investigate the relationship between these variables and participants' chronological age. Given that all variables deviated from

normality, Spearman correlations were used. Analyses were performed using JAMOVİ (The jamovi project, 2025).

Results and Discussion

Descriptive Statistics

Table 4. Descriptive Statistics Concerning the Feasibility of L-RECAP

	Mean	Median	Standard Deviation	Min – Max
Difficulties using the application				
	9.14	2	17.05	0–77
(0 = no difficulty; 100 = a lot of difficulties)				
Influence on mood				
	28.93	20	27.44	0–100
(0 = not at all; 100 = very strong influence)				
Influence on social interactions				
	31.27	21	27.63	0–100
(0 = not at all; 100 = very strong influence)				
Influence on activities				
	35	29	28.09	0–95
(0 = not at all; 100 = very strong influence)				
Number of notifications per day				
	40.99	40	30.70	0–100
(0 = acceptable; 100 = too many)				
Number of questions per notification				
	23.61	15	24.83	0–87
(0 = acceptable; 100 = too many)				
Number of days receiving the notifications				
	24.76	15	25.92	0–89
(0 = acceptable; 100 = too many)				
Difficulty of the memory task				
	30.27	25	21.91	0–82

(0 = not difficult at all; 100 = too difficult)

Response rate (%)	76.29	83	21.33	11–100
Number of events retrieved (from 0 to 5)	4.83	5	0.48	2–5

Few difficulties were reported with using the application, as the mean difficulty rating fell within the lowest 10% of the difficulty scale, and the median was close to zero. Regarding the intrusiveness of the task, both mean and median ratings were in the lower half of all scales, suggesting that participants considered that responding to the notifications had an acceptable impact on their mood, social interactions, or daily activities. Similarly, participants rated the number of notifications, questions, and days of participation as closer to “acceptable” than “too many.”

Moreover, in most cases, response rates were sufficient to select five events occurring on different days for the recall phase. Only three subjects did not respond to enough notifications to provide five events to interview them about (response rate below 27%). Additionally, all participants successfully retrieved at least two of the five events interviewed. Less than 1% ($n = 1$) retrieved only two events, 2% ($n = 3$) retrieved three, 10% ($n = 13$) retrieved four, and 85% ($n = 115$) retrieved all five. Finally, participants rated the difficulty of the memory task in the lower half of the scale, indicating that it was generally perceived as not particularly difficult.

Correlations with Age

Table 5 presents the results of the correlational analyses between participants’ age and their experience with L-RECAP. To get a clearer picture of the effects, we added the descriptive results for participants over the age of 60.

Table 5. Spearman Correlations Between Age and Feasibility of L-RECAP

	Age	Descriptive statistics for participants			
		over 60 years old ($n = 20$)			
		Mean	Median	SD	Min – Max
$r_s (p)$					
Difficulties using the application	.37 (<.001)	23.15	9.5	25.33	0–68
Influence on mood	.24 (.01)	52.6	52.5	32.29	1–100
Influence on social interactions	.10 (.25)	39.5	25	31.19	0–100
Influence on activities	.22 (.01)	48.10	53	28.08	3–94
Number of notifications per day	.12 (.15)	44.40	51	31.11	0–85
Number of questions per notification	.24 (.01)	32.30	29	23.9	0–79
Number of days receiving the notifications	.25 (.003)	39.5	38	29.18	0–79
Difficulty of the memory task	–.02 (.79)	34	26.5	29.22	0–82
Response rate	–.02 (.74)	75.9	89	24.7	11–97
Number of events retrieved	–.17 (.04)	4.84	5	.37	4–5

Note: Significant results are highlighted in bold.

These correlational analyses revealed a significant positive correlation between age and the perceived difficulty of using the application. However, the level of difficulty remained quite low even for participants over 60 years old, with a mean rating of 23.15 and a median of 9.5 on a 100-point scale. Importantly, this perceived difficulty did not impact their response rate, as no correlation was found between age and response rate.

The results also showed that, with increasing age, participants reported a greater perceived impact of responding to notifications on their mood and their activities, but not on their social interactions. The same positive correlations were observed with the number of questions per notification and the number of days during which they received the notifications, which are perceived as higher with advancing age.

Finally, a significant correlation was also found between age and the number of events retrieved out of the five covered in the interview, even though all participants over 60 years old retrieved a minimum of four events. However, this did not seem to influence the perceived difficulty of the memory task, which was not associated with age.

In conclusion, this study demonstrated that L-RECAP proved feasible and had good acceptability among healthy individuals. Although older adults reported slightly lower ease of use, the overall level of difficulty remained minimal and did not interfere with task completion, as response rates were high enough to allow the selection of five events for the recall phase. This finding supports evidence that older individuals generally hold a positive attitude toward smartphone use (Harris et al., 2022), especially when they receive appropriate training (Yeh et al., 2025). This highlights the importance of the familiarization phase during which participants are taught how to use the mobile app. The other results indicate that both the event-sampling and recall phases were well accepted, although some individuals still found the notifications somewhat disturbing. To address this issue, it is important to clearly explain to participants how these notifications will help to provide valuable information about their memory of daily events. Given that participants recruited here agreed to take part in full knowledge of the protocol's constraints, the L-RECAP feasibility will need to be further explored in a real-world clinical setting.

Study 2: L-RECAP's Convergent and Divergent Validity

Objectives

In this second study, we examined L-RECAP's convergent and divergent validity. Convergent validity was determined by comparing L-RECAP scores with validated verbal and visuospatial memory assessments. Because memories of everyday events involve both verbal and visual components, we hypothesized that L-RECAP performance would correlate with both types of memory assessments, but only moderately given the distinctions between static, decontextualized material and memory of personal events discussed in the introduction. Divergent validity was evaluated using executive functioning tasks, specifically verbal fluency and inhibition tests. Although executive functions are known to be involved throughout the encoding and retrieval of episodic memory, identifying cognitive measures entirely unrelated to memory was challenging. Therefore, we selected standard executive function tasks that are assumed to capture only a minor part of the executive processes involved in memory functioning. We thus expected a modest association with phonemic fluency, due to the involvement of strategic retrieval processes that may partially overlap with memory performance, although this association was expected to be weaker than those observed with memory measures. In contrast, minimal associations were predicted with the inhibition task, as inhibitory control is presumed to play a more limited role in the encoding and retrieval of everyday episodic memories.

Materials and Methods

Participants

Seventy-nine healthy adults (48 females, 31 males) participated in this study. They were aged between 19 and 79 years ($M = 46.60$, $SD = 16.35$), and their mean number of years of education was 14.58 ($SD = 2.14$). The inclusion criteria required participants to be native French speakers, over 18 years of age, with no history of neurological or psychiatric

disorders, and no uncorrected hearing or visual impairments. Participants were recruited via word of mouth, posters, and social media advertisements. The sample size was determined for the purpose of a larger study, but a post hoc power analysis conducted on G*Power indicated that a sample of 67 participants was sufficient to detect a medium effect size ($r = .30$) with a power of .80 at an α level of .05.

Procedure

The study was conducted in three sessions. During the first session, participants provided informed consent and demographic information and completed neuropsychological tasks, including story recall (Wechsler, 2012), phonological fluency (Bayard et al., 2023), and the Stroop test (Bayard et al., 2009). During the second session, participants completed the Brief Visual Memory Test – Revised (BVMT-R; Benedict et al., 1996) and the familiarization phase of L-RECAP. The last session was dedicated to the recall phase of L-RECAP. Sessions were scheduled one week apart, with participants completing the sampling of everyday events between the second and third sessions. Each session lasted from 45 minutes to one hour and was supplemented by other neuropsychological tests for the purposes of another study.

Measures

L-RECAP. Participants completed the L-RECAP memory task and two experimenters independently coded the richness and specificity of 20% of the transcripts to determine interrater reliability. Calculation of Cohen's κ (Cohen, 1960) resulted in an average interrater correlation of 0.83 for both richness and specificity. Interrater reliability was not calculated for the accuracy score or for the phenomenology score as they involve little or no subjectivity.

Neuropsychological Assessment. Table 6 summarizes the tests and indices selected to explore the different functions of interest. Two tasks were administered to explore convergent validity with verbal and visuospatial components of episodic memory. The other two investigated divergent validity with assessment of executive functioning and, more

specifically, verbal fluency and inhibition. A more detailed description of the different tasks is provided in supplementary material (Table S3). For the verbal episodic memory measure, only Story B from the Wechsler Memory Scale – IV (WMS-IV; Wechsler, 2012) was administered to limit the testing duration.

Table 6. Functions Assessed with the Neuropsychological Assessment and Indices Selected

Functions	Tests	Indices selected (raw scores)
Verbal episodic memory	Story B from Logical Memory Test (WMS-IV; French version of Wechsler, 2012)	Total immediate (IR) and delayed recall (DR)
Visuospatial episodic memory	Brief Visual Memory Test – Revised (BVMT-R; Benedict et al., 1996)	Sum of the three immediate trials (IR) and delayed recall (DR)
Spontaneous flexibility	Phonological fluency (French version by Bayard et al., 2023)	Total number of different correct words generated
Inhibition	Stroop Victoria (French version by Bayard et al., 2009)	Time difference between the interference plank and the color plank

Statistical Analyses

Convergent and divergent validity were examined by computing correlational analyses between the different L-RECAP scores and the scores obtained for the different memory and executive tasks. Given the effect of age on some of these cognitive performances, partial correlations were computed to control for chronological age. Spearman correlations were used because the relations between some of the variables were not linear. The significance level ($p < .05$) was corrected using the Benjamini and Hochberg procedure (Benjamini & Hochberg,

1995) considering the large number of comparisons. After the correction, α was set at .03.

Analyses were performed using JAMOV (The jamovi project, 2025).

Results and Discussion

Table 7 presents the results of the correlational analyses.

Table 7. Descriptive statistics (Mean, Standard Deviations, and Range) for the Logical Memory Test, BVMT-R, Phonological Fluency, Stroop Test, and Spearman Correlations with L-RECAP (r_s (p))

	Logical memory (IR)	Logical memory (DR)	BVMT-R (IR)	BVMT-R (DR)	Verbal fluency	Stroop time inter-color
<i>Mean (SD)</i>	14.06 (3.80)	12.91 (3.86)	26.15 (6.12)	10.05 (2.21)	25.04 (6.61)	10.62 (6.51)
Range	6 – 24	5 – 22	7 – 34	3 – 12	12 – 44	0 – 35
Accuracy	.31 (.006)	.35 (.002)	.37 (.001)	.29 (.009)	.14 (.21)	.002 (.99)
Internal richness	.41 (<.001)	.47 (<.001)	.46 (<.001)	.29 (.01)	.31 (.006)	.12 (.29)
Total richness	.45 (<.001)	.50 (<.001)	.46 (<.001)	.31 (.006)	.32 (.005)	.10 (.37)
Specificity	.43 (<.001)	.46 (<.001)	.21 (.07)	.17 (.14)	.22 (.14)	.04 (.73)
Phenomenolog y	.14 (.21)	.25 (.03)	.16 (.17)	.05 (.67)	.05 (.69)	.02 (.86)

Note: Significant results are highlighted in bold.

L-RECAP's accuracy, richness, and specificity scores were significantly correlated with both immediate and delayed recall on the logical memory test. Only accuracy and richness scores were significantly correlated with the immediate and delayed recall on the BVMT-R. All significant correlations were of medium size, which confirms L-RECAP's convergent validity with both verbal and visual memory assessments.

Regarding divergent validity, the results revealed significant correlations only between the L-RECAP richness scores and performance on the phonological fluency test, which is consistent since both involve strategic retrieval processes and require lexical access.

Additionally, as expected, no significant correlation was found with the Stroop Victoria task.

In conclusion, this study provides evidence of convergent validity by showing moderate correlations between different L-RECAP scores and those obtained with validated verbal and visual memory assessments. These findings support the idea that remembering everyday events relies on both verbal and visual components (Janssen et al., 2015). Although the L-RECAP recall phase is a verbal task, retrieving a personal event involves scene reconstruction processes, which recruit brain regions supporting visual memory (Hassabis et al., 2007). Moreover, memories of personal events are often represented in the form of visual images (Conway, 2009; Rubin et al., 2003). Consistent with this, lower visual imagery abilities, whether due to brain damage or aphantasia (i.e., inability to form mental images), have been linked to poorer episodic and autobiographical memory performance (Palombo et al., 2018; Sheldon et al., 2017). Once again, this evidence underscores the importance of using more comprehensive assessment tools to capture the multifaceted nature of personal memories.

In addition, positive correlations were also observed with the phonological fluency test, which has previously been linked to episodic memory performance (Ruff et al., 1997). This finding was expected, given that both tasks engage similar processes, such as strategic retrieval and lexical access. However, as the other scores of L-RECAP were not significantly correlated with this task, it suggests that lexical access and verbal productivity are the primary explanation for this link between verbal fluency and L-RECAP richness scores rather than the involvement of retrieval processes. In contrast, divergent validity is supported by the absence of a significant correlation with Stroop performance.

Study 3: L-RECAP's Correlation with Naturalistic Event Memory Performance

Objective

The convergent validity of L-RECAP was established in Study 2 through its correlations with traditional, validated memory measures of story recall and figures reproduction. However, this approach can be criticized, given that the primary goal of developing new ecological tasks is to capture memory processes that are neglected by traditional assessments. In Study 3, we therefore examined the relationship between L-RECAP and another assessment of real event memory. Several research paradigms have proposed assessment of event memory, such as those asking participants to recall an event they personally experienced while wearing a portable camera (e.g., Jeunehomme & D'Argembeau, 2020). Another approach is to ask participants to recall more or less standardized events produced during a session. For example, the What-Where-When Memory Test (Smulders et al., 2017) involves hiding various objects in different locations on two occasions. Beaver and Schmitter-Edgecombe (2017) also proposed to assess memories of daily-like activities (e.g., sweeping and dusting, cooking a soap) in a campus apartment. Mimicking a common clinical practice, Lemesle et al. (2022) asked their participants about what they recalled of the content of a previous neuropsychological assessment session. We opted for this procedure here. To ensure consistent administration across participants, we standardized the session-recall questions and then examined whether L-RECAP performance was associated with memories of the prior session, which represents a personally experienced event occurring in a natural context. As L-RECAP was designed to assess memory for real-life experiences, we hypothesized that L-RECAP scores would be positively associated with recall of the prior session. Furthermore, we explored whether this association was stronger than with a list-of words memory test.

Materials and Methods

Participants

This exploratory study included 20 French-speaking women aged between 20 and 51 years ($M = 35.4$; $SD = 10.76$), with a mean of 14.15 years of education ($SD = 1.04$). They were healthy individuals with no history of neurological, psychiatric, or alcohol-related disorders, and no uncorrected visual or hearing impairments. They were recruited through word of mouth and announcements on social media. The sample size was not determined a priori because the tasks of interest were drawn from a larger protocol designed for different objectives.

Procedure

Participants took part in two sessions scheduled one week apart. During the first session, they provided informed consent and demographic information. The session followed the structure of a standard neuropsychological assessment, beginning with a discussion about the participants' history, followed by the administration of five standard neuropsychological tests and the familiarization phase of L-RECAP. In addition, two standardized events were incorporated.

The entire first session – including the anamnesis, neuropsychological assessment, standardized events, and familiarization with L-RECAP – constituted a personally experienced event situated in a specific space and time, which participants memorized incidentally. Their memory of this whole session was assessed in session two, with the same experimenter, during which they also completed the recall phase of L-RECAP. They answered the sampling questions of L-RECAP between the two sessions.

Materials

Memory of the Previous Testing Session. The first session took the form of a traditional neuropsychological assessment and incorporated 10 elements that participants were

asked to recall during the second session. The first session began with (1) an anamnesis (i.e., collection of socio-demographic information), followed by the administration of five standard cognitive tests: the California Verbal Learning Test (CVLT; Poitrenaud et al., 2007), (2) immediate and (3) delayed recall, (4) the Digit Span (Wechsler, 2011), (5) the Stroop test (Stroop, 1935), (6) the Trail Making Test (Reitan, 1971), and (7) the Hayling Sentence Completion test (Burgess & Shallice, 1996). In addition, two standardized events were embedded within the session: the experimenter (8) offered the participant a glass of water, and (9) requested a favor, asking the participant to hand over a green folder, under the pretext that it contained the protocol of the next test. Finally, (10) the m-Path application was installed on participants' smartphones to explain the L-RECAP task, which enabled them to respond to notifications between sessions.

During the second session, participants were asked to freely recall the entire first session. They were given one point for each of the ten general events correctly mentioned and additional points for the associated details. For example, for the CVLT, one point was awarded if participants reported that the experimenter read a list of items to remember, and other points were given for the mention that the list was read multiple times or that there was a second list. A detailed description of the expected details and number of points allocated is provided in the supplementary material (Table S4). A cued recall session took place directly after the free recall. Participants were given cues for events that they did not spontaneously recall (e.g., for the CVLT, the cue is "I asked you to recall a list of items") and were asked to add everything they remembered about the event. A maximum of 41 points was awarded (10 for the general events and 31 for the associated details).

Statistical Analyses

Correlational analyses were performed to investigate the link between the recall of the previous testing session and (1) L-RECAP, and (2) the CVLT (i.e., the delayed recall). Due to

the small sample size and departure from normality of multiple variables, Spearman correlations were used. Since age did not correlate with any of the variables of interest and given the limited age range in this sample, the effect of age was not controlled in these correlation analyses.

Results and Discussion

Table 8 presents the results of the correlations between the recall of the previous session and L-RECAP and CVLT scores.

Table 8. Spearman Correlations Between Memory of the Previous Session, L-RECAP, and the CVLT

	Recall of the previous session
L-RECAP – Accuracy	.24 (.30)
L-RECAP – Internal richness	.19 (.41)
L-RECAP – Total richness	.06 (.80)
L-RECAP – Specificity	.54 (.01)
L-RECAP – Phenomenology	.19 (.49)
CVLT – Delayed recall	.23 (.34)

Note: Significant results are highlighted in bold.

These results indicate a strong positive correlation between memories of the previous session and the L-RECAP specificity score, whereas the other correlations did not reach significance. The strong association with specificity suggests that this L-RECAP score effectively assesses event memory, showing a much stronger correlation than the one observed with the CVLT. The fact that this association emerged for only the specificity score and not the other L-RECAP scores may be explained by the level of precision required when recalling the previous session. Indeed, participants had to provide specific, precise details and temporal organization about the events that occurred, which is precisely what the L-RECAP specificity score assesses about everyday events. The accuracy score, on the other hand,

focuses only on whether the reported information was correct, without requiring a high level of detail. The absence of significant correlations with the richness and phenomenological scores can also be easily explained, since neither the amount of information recalled nor the phenomenological qualities of the memory were considered in the recall of the previous session.

These findings should nevertheless be interpreted with caution, given the exploratory nature of the study and the small and homogeneous sample, which consisted exclusively of women relatively young. Knowing the influence that both sex and age could have on memory performance, the present results should be considered as a preliminary investigation of the convergent validity of L-RECAP that need replication with a larger and more representative sample.

Study 4: L-RECAP's Predictive Validity

Objective

The fourth study was designed to evaluate L-RECAP's predictive validity. More specifically, it examined the extent to which the measure is associated with self-reported daily-life functioning. To this end, two self-assessment questionnaires were administered to evaluate memory satisfaction, perceived memory ability (Multifactorial Memory Questionnaire, MMQ; Troyer & Rich, 2002), but also the phenomenological qualities of autobiographical memories (ART; Berntsen et al., 2019). The MMQ and the ART were selected because they cover different aspects of the self-assessment of everyday memory. Within the MMQ, the Satisfaction scale reflects a global appraisal of memory functioning that may rely primarily on metacognitive semantic knowledge, whereas the Ability scale requires a more experience-based assessment through the retrieval of specific experiences of memory difficulties, thereby engaging episodic memory processes to a greater extent. The ART, on the other hand, assesses phenomenological aspects of autobiographical remembering and thus

provides information about the experiential quality of memory retrieval. Together, these measures allowed us to assess predictive validity across a broad range of memory self-assessment outcomes.

Materials and Methods

Participants

A total of 135 participants were included in this study (79 females and 78 males). They were aged from 19 to 79 years, with a mean age of 43.6 years ($SD = 16.5$). Their mean number of years of education was 14.7 ($SD = 2.19$; range: 9–25). They were recruited through word of mouth and social media announcements. The inclusion criteria required participants to be over 18 years of age, native French speakers, and have no history of neurological, psychiatric, or alcohol-related disorders. They were also required to have no suspected current cognitive disorders and no uncorrected hearing or visual impairments. The sample consisted of individuals who had previously been enrolled in Study 2, as well as additional participants recruited for the purpose of another study in the lab who completed the same tasks (i.e., L-RECAP and the memory questionnaires of interest). A sensitivity analysis conducted in G*Power indicated that, with 135 participants, α of .05, and power of .80, the minimum detectable correlation was $r = .21$.

Procedure

Across the different study protocols, participants completed two sessions: one for the L-RECAP familiarization phase and questionnaires completion, and another for the L-RECAP recall phase. They answered the sampling questions during the week between the two sessions.

Materials

In addition to L-RECAP, participants completed two questionnaires: the French version of the MMQ (Fort et al., 2004) to assess memory satisfaction and self-perceived

ability, and the French version of the ART (Billet et al., 2023) to explore the phenomenology of autobiographical memories with seven factors (i.e., vividness, coherence, reliving, rehearsal, scene, visual details, and life story). The life story scale of the ART was not analyzed in this study since we had no reason to expect that memories from the last week (explored with L-RECAP) would already have become part of participants' life history. The indices selected for these questionnaires are presented in Table 9 and a more detailed description of the ART is provided in the supplementary material (Table S3).

Table 9. Questionnaires Used in Study 4

Variables of interest	Tests	Indices selected (raw scores)
Self-assessment of memory satisfaction and ability	Multifactorial Memory Questionnaire (MMQ; Troyer & Rich, 2002; French version by Fort et al., 2004)	Total scores on the satisfaction and ability scales
Phenomenology of autobiographical memories	Autobiographical Recollection Test (ART; Berntsen et al., 2019; French version by Billet et al., 2023)	Mean of the 3 items for each scale (except life story)

Statistical Analyses

Correlational analyses were performed to explore the link between L-RECAP performance and self-assessment in questionnaires. All the L-RECAP scores (accuracy, richness, specificity, and phenomenology) were used to explore their associations with the MMQ satisfaction and ability scales. However, only the phenomenological scores of L-RECAP were used in the analyses with the ART since this questionnaire was created to assess the phenomenology of autobiographical memories. All analyses were performed with JAMOVİ (The jamovi project, 2025).

Results and Discussion

Table 10 presents the results of the correlational analysis between L-RECAP and the MMQ.

Table 10. Correlations Between L-RECAP and the Satisfaction and Ability Scales of the MMQ

	MMQ satisfaction	MMQ ability
Accuracy	.21 (.02)	.22 (.01)
Internal richness	.12 (.18)	.23 (.007)
Specificity	.11 (.22)	.12 (.15)
Phenomenology	.13 (.14)	.19 (.03)

Note: Significant results are highlighted in bold.

The results revealed significant correlations of small size between all L-RECAP scores (except specificity) and self-perceived memory ability. Accuracy was also significantly correlated with the satisfaction scale, but not the richness, specificity, and phenomenological scores.

Table 11 presents the correlational analyses between the subscales of the L-RECAP phenomenological score and the same scales of the ART.

Table 11. Correlations Between L-RECAP Phenomenology Scores and the ART Scales

	ART – Vividness	ART – Coherence	ART – Reliving	ART – Rehearsal	ART – Scene	ART – Visual
L-RECAP Vividness	.21 (.01)					
L-RECAP Coherence		.26 (.002)				

L-RECAP Reliving	.38 (<.001)
L-RECAP Rehearsal	.25 (.005)
L-RECAP Scene allocentric	.25 (.004)
L-RECAP Scene egocentric	.24 (.007)
L-RECAP Visual	.35 (<.001)

Note: Significant results are highlighted in bold.

All factors of the ART showed significant positive correlations with the equivalent L-RECAP phenomenological scores, suggesting that the phenomenology of recent memories matches the phenomenology of autobiographical memories in general.

In conclusion, these results support the predictive validity of L-RECAP by revealing significant associations with self-perceived memory satisfaction and ability. Given these correlations, L-RECAP should be able to distinguish between individuals who are satisfied with their memory functioning and those who are unsatisfied, which is examined in Study 5. Although the present correlations remain weak, this pattern is nonetheless encouraging and not unexpected, given that this sample consisted of healthy adults who typically have satisfactory memory functioning. Examining this relation in clinical nosognosic patients would likely be more revealing, as greater variability in both objective performance and subjective complaints could lead to stronger associations.

Moreover, correlations between the phenomenological characteristics of memories from the last few days (assessed with L-RECAP) and those of autobiographical memories in general revealed significant associations. These findings replicate the results reported by

Gehrt et al. (2022), who demonstrated that the general subjective experience of autobiographical remembering was closely related to the phenomenology of specific memories. However, it remains difficult to determine what participants rely on when making their general judgments, which may depend on both general beliefs about their memory functioning but also situational factors, such as specific memories that are salient at the time they complete the questionnaire (Aytürk et al., 2024).

Study 5: L-RECAP's Discriminant Validity

Objectives

The objective of Study 5 was to explore whether L-RECAP was sensitive to subtle episodic memory difficulties that may go undetected in standard neuropsychological assessments. Episodic memory is characterized by a significant interindividual variability (Palombo et al., 2018), with a substantial proportion of healthy adults reporting personal memory dissatisfaction (Luck et al., 2018; Pinho et al., 2023). These differences have been linked to various processes underpinning episodic memory (e.g., visual imagery) that may vary across the population (Palombo et al., 2018).

In this study, we tested individuals who reported such subjective memory complaints despite the absence of any known neurological disorder (i.e., dissatisfied group). Their L-RECAP performance was compared to that of a group who said they were satisfied with their memory (i.e., satisfied group), for which we expected higher performances. We further evaluated whether L-RECAP could discriminate between these two groups better than a validated memory battery by comparing participants' performance across different indices of the WMS-IV (Wechsler, 2012).

Materials and Methods

Participants

The sample consisted of 56 adults, divided into two groups ($n = 28$) based on their self-reported satisfaction with their everyday memory functioning. Participants in the dissatisfied group were recruited via social media advertisements targeting individuals who reported being very dissatisfied with their memory in daily life or who felt that their personal memories were poorer and lacked detail compared to the general population. Individuals who recognize themselves with this description were invited to contact the researchers leading the study. The satisfied group was then recruited and matched at the group level for age, gender, and educational level. Sample characteristics are presented in Table 12. Inclusion criteria were (1) to be over 18 years old; (2) not have uncorrected hearing or visual problems; and (3) not have a history of psychiatric, neurological, or alcohol-related disorders. Participants were recruited through word of mouth, posters, and social media announcements. The sample size was determined pragmatically, according to the number of dissatisfied individuals that could be recruited during the year allocated to this study. Two participants (one from each group) did not complete the recall phase of L-RECAP, resulting in 27 participants per group. Moreover, four participants from the satisfied group did not complete the different tasks from the WMS-IV ($n = 24$). Sensitivity analyses indicated that this sample size allowed us to detect an effect of Cohen's $d = .34$.

Table 12. Sample Characteristics for Study 5

	Dissatisfied group	Satisfied group	Group comparison
Age	$M = 37.11$; $SD = 14.98$	$M = 36.79$; $SD = 14.57$	$U = 389$; $p = .97$
Years of education	$M = 15.14$; $SD = 2.72$	$M = 14.5$; $SD = 1.58$	$U = 369$; $p = .71$
Gender	16 females; 12 males	15 females; 13 males	$\chi^2 = .07$; $p = .79$

Procedure

The study was conducted in two sessions. In the first session, participants provided informed consent and demographic information, completed the selected subtests of the WMS-IV (Wechsler, 2012), and underwent the familiarization phase of L-RECAP. The second session was dedicated to the recall phase of L-RECAP. Between sessions, participants engaged in the sampling of everyday events and completed questionnaires.

Materials

L-RECAP. A new interrater reliability analysis was conducted for the richness and specificity scores by two independent raters based on 20% of the transcripts (one rater was the same as in the reliability assessment in Study 2, and the other was new). Calculation of Cohen's κ (Cohen, 1960) resulted in an average interrater correlation of 0.86 for richness and .79 for specificity.

Self-Assessment of Memory Functioning. The MMQ (Troyer & Rich, 2002; French version by Fort et al., 2004) was administered to explore participants' subjective perceptions of their memory functioning and to confirm the groups' difference in terms of satisfaction with their memory functioning. Table 13 presents the scores used in the analyses. A more detailed description of the questionnaire is provided in the supplementary material (Table S3).

Subtests of the Wechsler Memory Scale – IV (WMS-IV; Wechsler, 2012). Four subtests of the WMS-IV were administered: Logical Memory, Verbal Paired Associates, Visual Reproduction and Designs. These tests were used to calculate the auditory, visual, and delayed memory indices (see Table 13). This battery was selected to assess various aspects of episodic memory for comparison with the integrative L-RECAP task. More detailed descriptions of these tests are provided in the supplementary material (Table S3).

Table 13. Variables of Interest, Tests Used, and Indices Selected in Study 5

Variables of interest	Tests	Indices selected
-----------------------	-------	------------------

Self-assessment of memory functioning	Multifactorial Memory Questionnaire (MMQ; Troyer & Rich, 2002; French version by Fort et al., 2004)	Total score for each of the satisfaction and ability scales
Auditory episodic memory	Logical Memory and Verbal Paired Associates (WMS-IV; Wechsler, 2012)	Auditory Memory Index (AMI)
Visual episodic memory	Visual Reproduction and Designs (WMS-IV; Wechsler, 2012)	Visual Memory Index (VMI)
Delayed memory	Logical Memory, Verbal Paired Associates, Visual Reproduction, and Designs (WMS-IV; Wechsler, 2012)	Delayed Memory Index (DMI)

Statistical Analyses

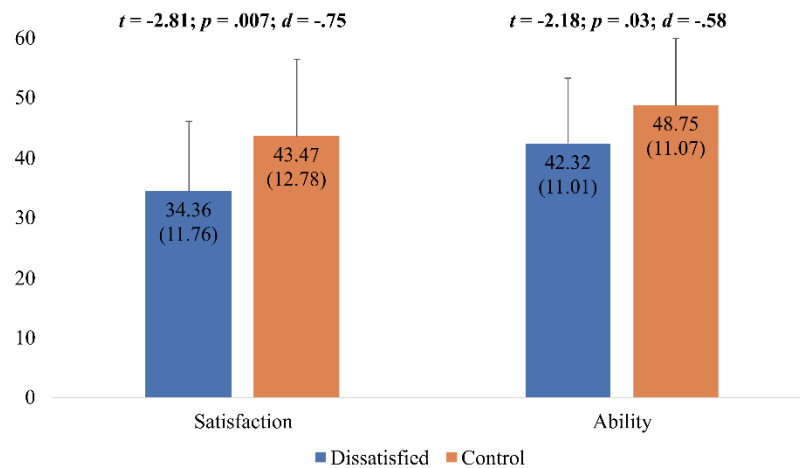
Independent sample *t*-tests were conducted to compare the two groups on all the different tasks (MMQ, L-RECAP, and WMS-IV indices). When a departure from normality was observed, Mann-Whitney *U* tests were performed.

Results and Discussion

Group Comparisons on Self-Reported Memory Complaints

Figure 3 presents the results of the *t*-test analyses on the two MMQ subscales.

Figure 3. Group Comparisons on the subscales of the Multifactorial Memory Questionnaire (MMQ)



The results confirmed that our recruitment method allows us to truly include participants with various memory satisfaction levels in our two groups. Indeed, the dissatisfied group reported significantly lower memory satisfaction and ability than the satisfied group.

Group Comparisons on Memory Performance

Figure 4 presents the *t*-test analyses on the WMS-IV scores and Figure 5 presents the Mann-Whitney *U*-test analyses on the different L-RECAP scores.

Figure 4. Group Comparisons on Memory Performance on the WMS-IV Indices ($n = 52$)

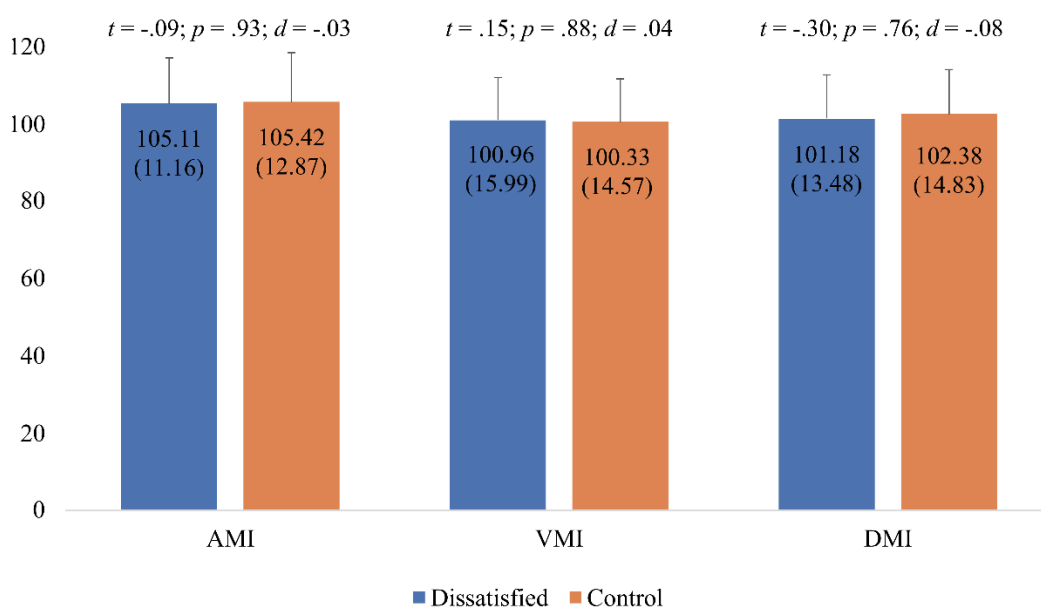
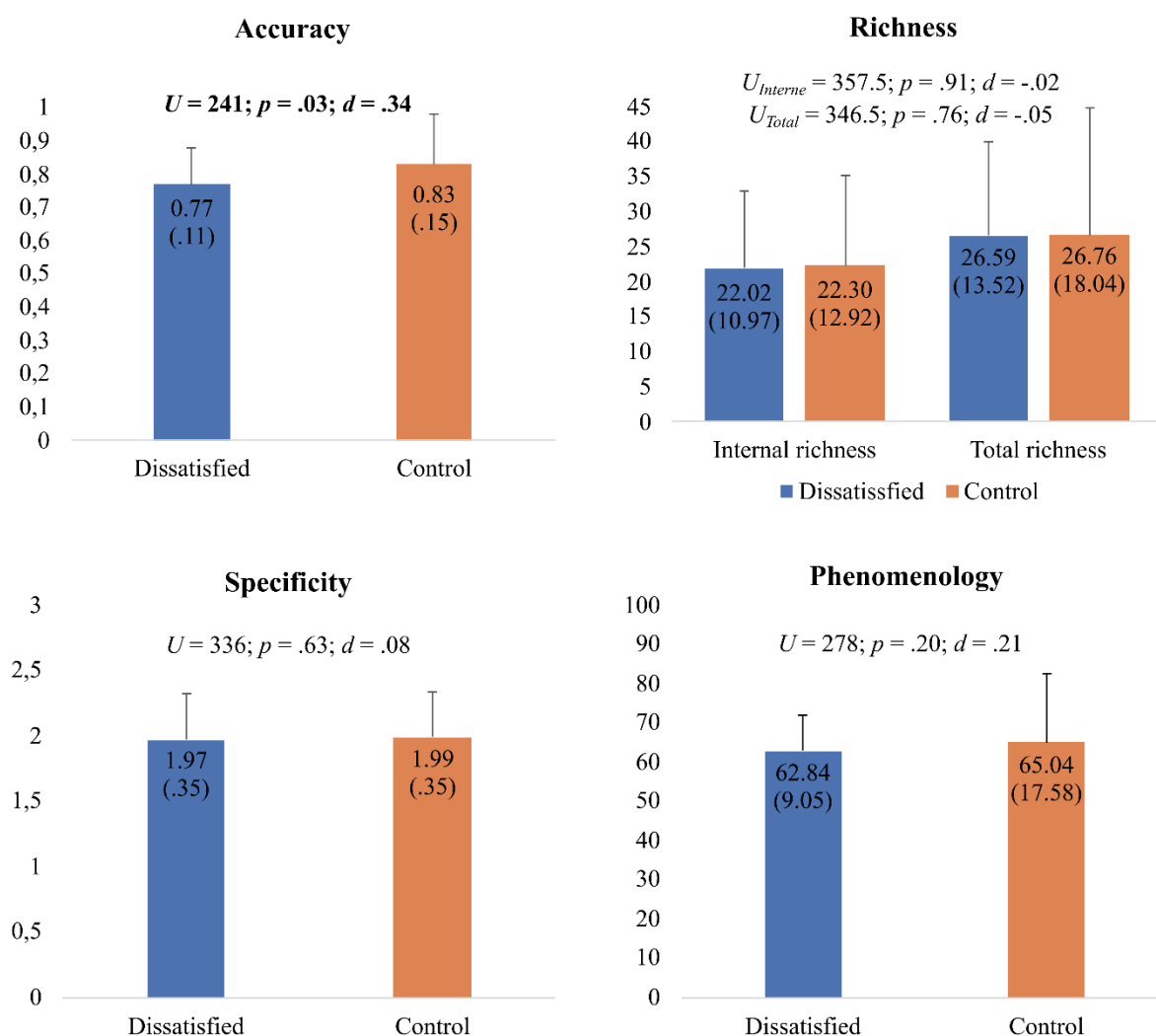


Figure 5. Group Comparisons on Memory Performance on L-RECAP ($n = 54$)



The two groups did not differ on the WMS-IV indices, with performance falling exactly within the normal range for both groups. This absence of group difference may reflect the limited ecological validity of traditional memory tests. Indeed, the WMS-IV assesses memory under structured and optimized memory conditions, whereas subjective memory complaints typically arise from failures occurring in everyday situations involving personally relevant information, competing demands, and less explicit encoding conditions. Consequently, individuals with subjective memory complaints may perform normally on standardized memory measures while still experiencing subtle memory difficulties in real-world. An alternative explanation is that subjective memory complaints stem from broader

psychological factors (e.g., depressive and/or anxiety symptoms; Benito-León et al., 2010; Comijs et al., 2002) . Nevertheless, the ecological validity account appears more consistent with the results of the group comparisons on L-RECAP.

Indeed, the two groups differ significantly on the accuracy score of L-RECAP, suggesting that this measure is sensitive to memory difficulties that are not captured by conventional laboratory-based assessments. This finding supports the discriminant validity of the L-RECAP accuracy score, which distinguishes between healthy individuals who are dissatisfied with their memory functioning and a control group who are satisfied with their memory. However, the richness, specificity, and phenomenology scores were equivalent between the two groups. One possible explanation for the lack of sensitivity of these indices may be that they capture characteristics that are largely dependent on the nature of the events being recalled, which is explored in Study 6. Another possible explanation is that subjective memory complaints are more closely related to concerns about the accuracy and reliability of the remembered information than to the number of details, their specificity, or to the subjective experience of remembering.

Taken together, these findings suggest that L-RECAP is sensitive to subjective memory complaints, showing greater discriminant capacity than the WMS-IV scales. The fact that the difference is only apparent on a single L-RECAP score and is modest in magnitude is not surprising given that both groups consisted of healthy individuals with high education levels and no medical history. There was therefore no reason to expect a larger difference between the groups.

STUDY 6: L-RECAP's Test-Retest Reliability

Objective

The aim of Study 6 was to determine L-RECAP's test-retest reliability by comparing participants' performance at two measurement points. In addition, we wanted to examine

whether L-RECAP performance was influenced by the fact that the events selected at each measurement point were not the same. We hypothesized that the participants' stable memory abilities would be the primary predictors of their L-RECAP performance, although certain situational variables related to event characteristics (importance, frequency, memorability) might also account for some variance. Indeed, each event examined is inherently unique, and its intrinsic memorability may influence recall performance. This variability can reduce measurement reliability, potentially leading to an underestimation of an individual assessed during a period characterized by less memorable events, or conversely to an overestimation during periods rich in salient experiences. Consequently, fluctuations in performance may be observed across different assessment periods but, ideally, such event-related factors would have only a minimal impact on performance and participants' actual memory abilities would remain the main predictor.

Materials and Methods

Participants

The sample was composed of 30 healthy participants. This sample size was determined a priori using G*Power to detect large correlation effects, with an α level of .05 and a statistical power of .80. Twenty-three individuals who had participated in one of the previous studies agreed to perform L-RECAP a second time, and the other seven participants were recruited exclusively for the test-retest study. As in the preceding studies, they had to be over 18 years old, not have uncorrected hearing or visual problems, and not have a history of psychiatric, neurological, or alcohol-related disorders. There were 15 females and 15 males. Their mean age was 35.23 ($SD = 13.25$), ranging from 21 to 61. Their mean years of education were 15.13 ($SD = 1.66$), ranging from 12 to 18.

Procedure

The participants completed L-RECAP at two different times. The two recall phases were separated by approximately six weeks and conducted by the same experimenter. Intervals for test-retest reliability are generally shorter (Streiner & Kottner, 2014), but we extended the delay due to the time commitment required for L-RECAP completion (i.e., the one-week notification period). Adjustments were also made to accommodate participants' availability. Consequently, the intervals between sessions ranged from 32 to 83 days, with a mean of 46.5 days ($SD = 11.48$). The seven participants recruited specifically for the test-retest reliability analysis completed three sessions: the familiarization phase, the first recall phase, and the second recall phase. The 23 participants who had already taken part in one of the previous studies only completed one additional session dedicated to the second recall phase. The sampling of everyday events was completed the week before the recall phase at each time point.

Statistical Analyses

Correlational analyses were conducted to explore L-RECAP's test-retest reliability by comparing performance at Time 1 and Time 2. Then, since the events examined differed across the two points, we explored whether performance was better explained by participants' memory abilities or by the events selected. To do this, we conducted linear mixed models with the different L-RECAP scores included as dependent variables, time as a fixed effect, and participant and event as random effects. However, event could not initially be included as a random factor, since each participant recalled different events. To address this issue, for each participant, we ordered all the events from Time 1 and Time 2 from 1 to 10 based on the composite score used for event selection (i.e., the mean frequency, importance, and memorability). This adjustment made events more comparable across subjects, allowing the variable "event" to be included in the mixed models as a random effect. The dependent

variables were the different L-RECAP scores. Analyses were performed using JAMOVI (The jamovi project, 2025).

Results and Discussion

Table 14 presents descriptive statistics and the correlations between Time 1 and Time 2.

Table 14. Test-Retest Reliability of L-RECAP

	Time 1	Time 2	<i>r</i> Time 1 vs. Time 2
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	(<i>p</i>)
Accuracy	.86 (.10)	.87 (.09)	.76 (<.001)
Internal richness	28.17 (14.62)	22.20 (7.17)	.57 (.001)
Total richness	32.11 (16.11)	25.21 (8.54)	.64 (<.001)
Specificity	2.16 (.32)	2.10 (.30)	.75 (<.001)
Phenomenology	69.44 (16.52)	69.02 (15.99)	.71 (<.001)

Note: Significant results are highlighted in bold.

These results revealed strong significant correlations between the two measures of accuracy, specificity, and phenomenology, which suggest that L-RECAP has satisfactory test-retest reliability. However, correlations between the richness scores were lower ($r = .57$ and $.64$), suggesting that these scores should not be used in clinical practice.

Table 15 presents the results of the linear mixed models exploring the effect of time, participant, and event on the different L-RECAP scores.

Table 15. Linear Mixed Models Exploring the Effect of Time, Participant, and Event on the Different L-RECAP Scores

Outcome variable	Fixed effect	β	<i>SE</i>	<i>t</i>	<i>p</i>	Random effect	Variance	<i>SD</i>	ICC	<i>p</i>
Accuracy	Time	.003	.02	.14	.89	Participant	.004	.07	.10	.002
						Event	.06	.02	.01	.66

Internal richness	Time	-5.99	1.48	-4.04	<.001	Participant	76.27	8.73	.32	<.001
						Event	11.26	3.36	.07	.005
Total richness	Time	-7.01	1.50	-4.68	<.001	Participant	104.61	10.22	.39	<.001
						Event	12.31	3.51	.07	.003
Specificity	Time	-.06	.06	-.96	.34	Participant	.06	.23	.18	<.001
						Event	.001	.09	.03	.14
Phenomenology	Time	-.48	2.07	-.23	.82	Participant	179.90	13.41	.37	<.001
						Event	30.63	5.53	.09	<.001

Note: Significant results are highlighted in bold.

These different models revealed that there was no time effect on the accuracy, specificity and phenomenological scores, confirming their satisfactory test-retest reliability. However, the richness scores demonstrated a significant effect of time, with higher scores at Time 1 than at Time 2, confirming that richness is less reliable. This discrepancy for richness between the two time points may reflect a recall training effect resulting in a progressively more “straight-to-the-point” recall style.

Beyond their stability over time, accuracy and specificity were influenced only by the participants-related variables, and not by event characteristics, which confirms their relevance in clinical use. However, richness and phenomenology scores appeared to be dependent on event characteristics, although the proportion of variance they explained remained weaker than the proportion attributable to the participant. This impact is in line with previous results showing that some characteristics of the event, such as emotional arousal, influence the number of details recalled (Heuer & Reisberg, 1990; Simpson & Sheldon, 2020; St Jacques & Levine, 2007) and a memory’s phenomenological characteristics (Maki et al., 2013; Rubin et al., 2003). Several explanations can be proposed for these memory advantages for highly significant events, such as increased activation of the amygdala and its interactions with the hippocampus, enhanced attention at encoding (Baraly et al., 2016; Kensinger, 2009), or more frequent rehearsals (Luminet et al., 2000; Walker et al., 2009). However, the finding that

these event characteristics did not influence accuracy and specificity may suggest that these scores capture a form of event memory that depends on more stable memory abilities. Through their focus on essential information, these scores may have removed the impact that situational variables (e.g., social sharing) can have on memory performance.

General Discussion

This paper introduces a new memory assessment tool called L-RECAP, which was developed in light of current conceptions of memory of everyday events. Indeed, remembering events from daily life involves multiple processes and many of them are not captured by traditional clinical memory tasks, which makes it hard to detect certain memory difficulties. Several ecological approaches have been proposed, such as memory assessment of events experienced in the real world (e.g., Jeunehomme & D'Argembeau, 2020; Lemesle et al., 2022; Smulders et al., 2017) or in virtual reality (e.g., Plancher & Piolino, 2017), but they tend to focus on a single event selected by the experimenter. Other experience sampling studies explored memories of personal events, but they focused only on limited aspects of memory (e.g., Laliberte et al., 2021; McKenzie et al., 2020). As far as we are aware, L-RECAP is the first tool created to assess the accuracy, richness, specificity, and phenomenology of memories of recent real-life events, chosen according to the personal significance attributed by the participant. The tool was specifically designed to be easily implemented in clinical practice. However, validating this kind of personalized assessment is challenging, and this study aimed to provide first insights concerning its feasibility, validity, and reliability among healthy individuals, a crucial first step before its use can be extended to clinical contexts.

Feasibility, Validity, and Reliability

Study 1 demonstrated the broad acceptability and feasibility of L-RECAP in participants aged from 19 to 80 years and showed that age did not seem to impact the

completion of this task despite the use of a mobile application. However, feasibility and response rate in the general population may be lower, as participation in the present studies required volunteering for a week-long smartphone-based sampling protocol, which may not be acceptable to all individuals.

Regarding L-RECAP's validity, Study 2 showed good convergent and divergent validity with validated neuropsychological tasks, confirming the reliance on strategic retrieval processes for both verbal and visual information in L-RECAP. Study 3 further supports the claim that L-RECAP assesses event memory, as there was a stronger correlation between memories from a previous session and the L-RECAP specificity score, compared to that of a word list test. Study 4, on the other hand, showed L-RECAP's predictive validity with questionnaires assessing memory satisfaction, ability, and the phenomenological qualities of autobiographical memories. These findings provide support for the ecological validity of L-RECAP. Although the concept of ecological validity has been defined in different ways across the literature, it can be conceptualized as comprising two major components: verisimilitude and veridicality (Suchy et al., 2024). The present results suggest that L-RECAP meets both criteria. Regarding verisimilitude, this task not only resembles real-world memory functioning, but also engages a broad range of processes that are typically involved in everyday event memory formation and retrieval (i.e., segmentation of ongoing personal experiences, encoded in a multisensorial way and retrieved across varying retention intervals with the associated subjective experience of remembering). Moreover, L-RECAP also demonstrates veridicality, as performances were associated with self-perceived memory functioning in daily life. The modest size of the correlations with questionnaires likely reflects the fact that the sample consisted of healthy participants. Stronger correlations might be observed in clinical populations that experience significant memory difficulties in everyday life. Ongoing studies in our lab are exploring whether this hypothesis is true in a sample of

patients with traumatic brain injury (TBI). In Study 5, we highlighted the discriminant validity of the accuracy score, confirming that L-RECAP is sensitive to subjective memory complaints even among healthy individuals. This discriminant capacity appeared to be greater than that of the WMS-IV. Although this finding still needs to be confirmed in clinical populations, it is encouraging that the accuracy score appears to be the measure that discriminates best. Indeed, it is probably the most suitable score for use in clinical settings as it does not require recording or detailed analysis of the participants' reported information and can be calculated directly during the recall phase. The discriminant power of L-RECAP is further supported by the absence of a ceiling effect across its different scores (as highlighted in the descriptive statistics from the different studies). Although memory performance was generally high, as is expected in healthy participants, the task nevertheless provides sufficient headroom to capture interindividual variability.

L-RECAP's reliability was investigated with two different measures. In Studies 2 and 5, interrater reliability was calculated on the richness and specificity scores – those that involved a significant degree of subjectivity. The results revealed good interrater reliability (κ from .83 to .86 for richness, and from .79 to .83 for specificity). Moreover, Study 6 examined test-retest reliability, which was a major concern given that the events assessed were not identical across testing sessions. This study's results revealed satisfactory reliability for the accuracy, specificity, and phenomenology scores (r from .71 to .76), but lower reliability for richness (r from .57 to .64). These results are congruent with test-retest reliability coefficients observed with traditional validated memory tasks: r from .59 to .76 for the WMS-IV (Wechsler, 2012); from .57 to .69 for the CVLT (Alioto et al., 2017); and from .63 to .69 for the Rey Auditory Verbal Learning Test (Van der Elst et al., 2008). Importantly, in our study, the variance of the accuracy and specificity scores was not explained by the events selected at the different time points. Although event characteristics influenced performance for richness

and phenomenology, individual differences among participants remained the primary factor explaining their performance. Nevertheless, the significant effect of situational factors highlights the importance of calculating confidence intervals before interpreting a decline or improvement in memory performance, regardless of the memory task.

Limitations and Perspectives

These results are encouraging and demonstrate the good psychometric properties of L-RECAP, which allow the collection of relevant information about everyday memory functioning. This provides fertile ground for the use of L-RECAP in the context not only of memory assessment but also of rehabilitation. Although the studies presented here focused on global scores for accuracy, richness, specificity, and phenomenology, L-RECAP can provide useful insights concerning the specific types of information that may be affected (e.g., temporal information, visual aspects, emotional state, etc.). This can help clinicians to set very precise and personalized goals that align with the specific needs of each patient, which can be difficult based solely on the results of traditional memory tasks. For example, if L-RECAP revealed specific difficulties with the retrieval of temporal information, a targeted rehabilitation program using reminiscence techniques and/or diary-based strategies could be proposed to reduce these difficulties. Given the task's good test-retest reliability, L-RECAP could also be used to assess the effectiveness of such an intervention – at least for the accuracy, specificity, and phenomenology of memories – where traditional word list tasks seem quite unsuitable.

However, the successful implementation of L-RECAP involves certain considerations prior to administration. Firstly, individuals must keep their smartphones accessible and connected (ideally, even at work) throughout the week of notifications, which should represent a typical week and include a variety of events. Consequently, the use of L-RECAP may not be recommended for patients or participants who are very sedentary and engage in

few activities, or for whom it would not be possible to keep their smartphones nearby. Its applicability to hospitalized patients or individuals living in institutions therefore remains uncertain and needs to be explored, as they might not be inclined to respond to notifications for a week. When used for clinical purposes, it is important for the events captured during the notification week to reflect situations in which the patient reports memory complaints.

Another limitation that can be addressed to L-RECAP is that this experience sampling procedure itself may influence memory encoding. Since participants knew that they would be interviewed about events recorded in the app, they may have increased their attention devoted to these events, thereby promoting deeper encoding and facilitating subsequent recall. As a result, memory performance may not fully reflect naturally occurring encoding processes in everyday life. However, this potential bias is probably reduced by the number of notifications sent. Since participants had to answer five notifications per day for seven days, it was unlikely that they could anticipate which specific event would later be selected for recall and they were unaware of the nature of the recall phase. Therefore, they could not strategically focus on specific events or information in anticipation of later memory assessment. Moreover, this potential limitation is not unique to L-RECAP, as many standard episodic memory tests also rely on intentional encoding procedures in which participants are explicitly aware that the information presented will later have to be recalled. Importantly, all participants were assessed under identical encoding conditions, ensuring that any potential facilitation of encoding was equivalent across individuals. This is particularly relevant for future applications of L-RECAP in clinical populations, as group comparisons would be conducted under the same conditions.

L-RECAP was indeed initially created to be used with clinical populations. One may therefore wonder whether it is applicable in neuropsychological practice, since all the studies in this paper were conducted with healthy individuals. The aim of this paper was to establish

some of the test's psychometric properties, which is an essential preliminary step before implementing it in clinical settings. Ongoing research in our lab is now being conducted with individuals with TBI, which will help determine L-RECAP's discriminant validity in patients with brain injuries, as well as its predictive validity with individuals who report higher-level of memory complaints. Future research should also examine its applicability in other clinical populations in which memory impairments are commonly reported, such as multiple sclerosis (Sedighi, 2011), epilepsy (Rayner & Tailby, 2017), or even MCI patients (Clément et al., 2008). In this latter population, tasks requiring the binding and retrieval of multiple contextual features may be particularly sensitive to subtle memory changes, as these processes rely heavily on hippocampal and parahippocampal structures (Konkel & Cohen, 2009), which are among the earliest regions affected in MCI (Bastin & Delhay, 2023). Accordingly, L-RECAP may be useful for detecting memory difficulties that may go undetected by traditional assessment tools. Future studies should, however, first establish its feasibility in older adults with neurocognitive impairment.

Regarding the relevance of L-RECAP scores for future clinical use, this paper revealed that richness scores did not show discriminant power or good test-retest reliability, unlike accuracy, specificity, and phenomenology scores. Although it may be premature to dismiss the potential importance of memory richness, particularly given its important role in maintaining social relationships (Alea & Bluck, 2003), future applications may prioritize alternative indices (e.g., recall length or word count) considering the present findings and the time-intensive nature of richness scoring. In any case, the selection of relevant L-RECAP scores should be guided by the specific purpose for which the task is being used.

Another limitation is the fact that, even though the different scores of L-RECAP may highlight everyday memory difficulties, they cannot identify the specific memory processes that may be altered when performance is low. Complementary assessment tools are still

needed to clarify the mechanisms underlying reduced performance. To determine whether L-RECAP may be sensitive to deficits of specific memory stages (i.e., encoding, consolidation, or retrieval), future studies could examine its associations with traditional measures of these specific stages and assess whether it correlates more strongly with indicators of encoding, consolidation, or retrieval abilities. Beyond the main scores analyzed in this paper, other L-RECAP indicators could also provide information about the potential altered memory stage; for example, the time taken to retrieve events after the three cue words are given could reflect retrieval skills. It would also be interesting to examine the effect of delay on L-RECAP performance, for example, by comparing memories from the most recent days of the week with those from earlier days. This could help reveal potential alterations of the consolidation process, which are often reported in patients with epilepsy (Geurts et al., 2015; Landry et al., 2022). Other clinical populations have also shown alterations of various processes that affect their memory. For instance, individuals with obsessive-compulsive disorder, schizophrenia, and Parkinson's disease have exhibited segmentation impairments (Zacks & Sargent, 2010). Although L-RECAP does not directly assess segmentation ability, such deficits are expected to manifest in reduced performance on the task. To confirm this hypothesis, ongoing studies in the lab are currently exploring associations between segmentation tasks and L-RECAP.

It would also be interesting to investigate whether memory performance in everyday life varies according to ethnic or cultural background. For example, Wang (2009) demonstrated that Asian individuals tend to recall more general, less specific events than Western participants. This difference is thought to reflect distinct cultural constructions of the self (more individualistic in Western cultures vs. more relational in Asian cultures), which in turn influence attentional focus during scene perception (e.g., emphasis on central object versus overall context). Such variations may lead to different performance on memory tasks.

An ongoing study in the laboratory is therefore exploring whether cultural differences are also reflected in an ecological task such as L-RECAP.

Conclusion

This study validated a new tool for assessing memory of everyday events, designed to align with contemporary understanding of how memory functions in real-life contexts. In a series of studies of L-RECAP, we demonstrated its feasibility, as well as its convergent, divergent, discriminant, and predictive validity in healthy individuals. It also showed good interrater and test-retest reliability. While further work is needed to explore its applicability in clinical populations, L-RECAP constitutes the first memory task targeting personal memories of daily life that combines good psychometric properties with ease of use in routine clinical practice, and is not excessively time-consuming or costly. These features thus support the tool's utility as a reliable, ecologically valid measure in both clinical assessment and research contexts.

References

- Alioto, A. G., Kramer, J. H., Borish, S., Neuhaus, J., Saloner, R., Wynn, M., & Foley, J. M. (2017). Long-term test-retest reliability of the California Verbal Learning Test—second edition. *Clinical Neuropsychologist*, 31(8), 1449–1458. <https://doi.org/10.1080/13854046.2017.1310300>
- Allé, M. C., Manning, L., Potheegadoo, J., Coutelle, R., Danion, J. M., & Berna, F. (2017). Wearable Cameras Are Useful Tools to Investigate and Remediate Autobiographical Memory Impairment: A Systematic PRISMA Review. *Neuropsychology Review*, 27(1), 81–99. <https://doi.org/10.1007/s11065-016-9337-x>
- Aytürk, E., Ece, B., Göktaş, N., & Gülgöz, S. (2024). How much trait variance is captured in autobiographical memory ratings? *Applied Cognitive Psychology*, 38(5). <https://doi.org/10.1002/acp.4240>
- Baraly, K. T. A., Hot, P., Davidson, P. S. R., & Talmi, D. (2016). How emotional arousal enhances episodic memory. In *Learning and memory: A comprehensive reference* (pp. 295–324). Elsevier Science. <https://doi.org/10.1016/B978-0-12-809324-5.21051-1>
- Bastin, C., & Delhaye, E. (2023). Targeting the function of the transentorhinal cortex to identify early cognitive markers of Alzheimer’s disease. *Cognitive, Affective and Behavioral Neuroscience*, 23(4), 986–996. <https://doi.org/10.3758/s13415-023-01093-5>
- Bayard, S., Erkes, J., & Moroni, C. (2009). *Test du Stroop Victoria - Adaptation francophone*.
- Bayard, S., Madiouni, C., Niel Carlos, C., Cornu, C., Moulin, A., & Raffard, S. (2023). French adult normative data for a Standard Executive Neuropsychological Test Battery. *Archives of Clinical Neuropsychology*, 40(3), 486–497. <https://doi.org/10.1093/arclin/acad062>
- Beaver, J., & Schmitter-Edgecombe, M. (2017). Multiple Types of Memory and Everyday Functional Assessment in Older Adults. *Archives of Clinical Neuropsychology*, 32(4), 413–426. <https://doi.org/10.1093/arclin/acx016>
- Benedict, R. H. B., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment*, 8(2), 145–153.
- Benito-León, J., Mitchell, A. J., Vega, S., & Bermejo-Pareja, F. (2010). A population-based study of cognitive function in older people with subjective memory complaints. *Journal of Alzheimer’s Disease*, 22(1), 159–170. <https://doi.org/10.3233/JAD-2010-100972>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300. <https://doi.org/https://www.jstor.org/stable/2346101>
- Berntsen, D., Hoyle, R. H., & Rubin, D. C. (2019). The Autobiographical Recollection Test (ART): A measure of individual differences in autobiographical memory. *Journal of Applied Research in Memory and Cognition*, 8(3), 305–318. <https://doi.org/10.1016/j.jarmac.2019.06.005>

- Billet, M., Geurten, M., & Willems, S. (2023). How well do you think you remember your personal past? French validation of the Autobiographical Recollection Test (ART) and exploration of age effect. *Memory*, 31(6), 864–870. <https://doi.org/10.1080/09658211.2023.2207805>
- Burgess, P. W., & Shallice, T. (1996). Response suppression, initiation and strategy use following frontal lobe lesions. *Neuropsychologia*, 34(4), 263–272. [https://doi.org/10.1016/0028-3932\(95\)00104-2](https://doi.org/10.1016/0028-3932(95)00104-2)
- Caserman, P., Garcia-Agundez, A., Gámez Zerban, A., & Göbel, S. (2021). Cybersickness in current-generation virtual reality head-mounted displays: systematic review and outlook. *Virtual Reality*, 25(4), 1153–1170. <https://doi.org/10.1007/s10055-021-00513-6>
- Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology Review*, 13(4), 181–197. <https://doi.org/10.1023/B:NERV.00000009483.91468.fb>
- Clément, F., Belleville, S., & Gauthier, S. (2008). Cognitive complaint in mild cognitive impairment and Alzheimer's disease. *Journal of the International Neuropsychological Society*, 14(2), 222–232. <https://doi.org/10.1017/S1355617708080260>
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1(1). <https://doi.org/https://doi.org/10.1177/001316446002000104>
- Comijs, H. C., Deeg, D. J. H., Dik, M. G., Twisk, J. W. R., & Jonker, C. (2002). Memory complaints; the association with psycho-affective and health problems and the role of personality characteristics A 6-year follow-up study. *Journal of Affective Disorders*, 72, 157–165. [https://doi.org/10.1016/s0165-0327\(01\)00453-0](https://doi.org/10.1016/s0165-0327(01)00453-0)
- Conway, M. A. (2001). Sensory-perceptual episodic memory and its context: Autobiographical memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 356(1413), 1375–1384. <https://doi.org/10.1098/rstb.2001.0940>
- Conway, M. A. (2009). Episodic memories. *Neuropsychologia*, 47(11), 2305–2313. <https://doi.org/10.1016/j.neuropsychologia.2009.02.003>
- Fort, I., Adoul, L., Holl, D., Kaddour, J., Gana, K., & Kamel, G. (2004). Psychometric Properties of the French Version of the Multifactorial Memory Questionnaire for Adults and the Elderly. *Canadian Journal on Aging*, 23(4), 347–357. <https://amu.hal.science/hal-01789129v1>
- Frankland, P. W., & Bontempi, B. (2005). The organization of recent and remote memories. *Nature Reviews Neuroscience*, 6(2), 119–130. <https://doi.org/10.1038/nrn1607>
- Gehrt, T. B., Nielsen, N. P., Hoyle, R. H., Rubin, D. C., & Berntsen, D. (2022). Individual differences in autobiographical memory: The Autobiographical Recollection Test predicts ratings of specific memories across cueing conditions. *Journal of Applied Research in Memory and Cognition*, 11(1), 85–96. <https://doi.org/10.1037/h0101869>

- Geurts, S., van der Werf, S. P., & Kessels, R. P. C. (2015). Accelerated forgetting? An evaluation on the use of long-term forgetting rates in patients with memory problems. *Frontiers in Psychology*, 6, 1–9. <https://doi.org/10.3389/fpsyg.2015.00752>
- Grilli, M. D., & Sheldon, S. (2022). Autobiographical event memory and aging: older adults get the gist. *Trends in Cognitive Sciences*, 26(12), 1079–1089. <https://doi.org/10.1016/j.tics.2022.09.007>
- Harris, M. T., Blocker, K. A., & Rogers, W. A. (2022). Older Adults and Smart Technology: Facilitators and Barriers to Use. *Frontiers in Computer Science*, 4. <https://doi.org/10.3389/fcomp.2022.835927>
- Hassabis, D., Kumaran, D., & Maguire, E. A. (2007). Using imagination to understand the neural basis of episodic memory. *Journal of Neuroscience*, 27(52), 14365–14374. <https://doi.org/10.1523/JNEUROSCI.4549-07.2007>
- Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition*, 18(5), 496–506.
- Janssen, S. M. J., Kristo, G., Rouw, R., & Murre, J. M. J. (2015). The relation between verbal and visuospatial memory and autobiographical memory. *Consciousness and Cognition*, 31, 12–23. <https://doi.org/10.1016/j.concog.2014.10.001>
- Jeunehomme, O., & D'Argembeau, A. (2020). Event segmentation and the temporal compression of experience in episodic memory. *Psychological Research*, 84(2), 481–490. <https://doi.org/10.1007/s00426-018-1047-y>
- Johnson, M. K. (2006). Memory and reality. *American Psychologist*, 61(8), 760–771. <https://doi.org/10.1037/0003-066X.61.8.760>
- Kensinger, E. A. (2009). Remembering the details: Effects of emotion. *Emotion Review*, 1(2), 99–113. <https://doi.org/10.1177/1754073908100432>
- Klein, S. B., & Nichols, S. (2012). Memory and the Sense of Personal Identity. *Mind*, 121(483), 677–702. <https://doi.org/10.1093/mind/fzs080>
- Konkel, A., & Cohen, N. J. (2009). Relational memory and the hippocampus: Representations and methods. *Frontiers in Neuroscience*, 3(2), 166–174. <https://doi.org/10.3389/neuro.01.023.2009>
- Kopelman, M. D., Wilson, B. A., & Baddeley, A. D. (1989). The autobiographical memory interview: A new assessment of autobiographical and personal semantic memory in amnesic patients. *Journal of Clinical and Experimental Neuropsychology*, 11, 724–744. <https://doi.org/10.1080/01688638908400928>
- Laliberte, E., Yim, H., Stone, B., & Dennis, S. J. (2021). The Fallacy of an Airtight Alibi: Understanding Human Memory for “Where” Using Experience Sampling. *Psychological Science*, 32(6), 944–951. <https://doi.org/10.1177/0956797620980752>
- Landry, A., Rouleau, I., Nguyen, D. K., & Boucher, O. (2022). L’oubli accéléré à long terme chez les personnes atteintes d’épilepsie : état des connaissances et implications pour l’évaluation clinique. *Revue de Neuropsychologie*, 14(3), 179–188. <https://doi.org/10.1684/nrp.2022.0717>

- Lemesle, B., Barbeau, E. J., Rigal, E. M., Denuelle, M., Valton, L., Pariente, J., & Curot, J. (2022). Hidden objective memory deficits behind subjective memory complaints in temporal lobe epilepsy. *Neurology*, 98(8). <https://doi.org/10.1212/WNL.00000000000013212>
- Levine, B., Svoboda, E., Hay, J. F., Winocur, G., & Moscovitch, M. (2002). Aging and autobiographical memory: Dissociating episodic from semantic retrieval. *Psychology and Aging*, 17(4), 677–689. <https://doi.org/10.1037/0882-7974.17.4.677>
- Luck, T., Roehr, S., Rodriguez, F. S., Schroeter, M. L., Witte, A. V., Hinz, A., Mehnert, A., Engel, C., Loeffler, M., Thiery, J., Villringer, A., & Riedel-Heller, S. G. (2018). Memory-related subjective cognitive symptoms in the adult population: Prevalence and associated factors - results of the LIFE-Adult-Study. *BMC Psychology*, 6(23). <https://doi.org/10.1186/s40359-018-0236-1>
- Luminet, O., Bouts, P., Delie, F., Manstead, A. S. R., & Rime, B. (2000). Social sharing of emotion following exposure to a negatively valenced situation. *Cognition and Emotion*, 14(5), 661–688. <https://doi.org/10.1080/02699930050117666>
- Maki, Y., Janssen, S. M. J., Uemiya, A., & Naka, M. (2013). The phenomenology and temporal distributions of autobiographical memories elicited with emotional and neutral cue words. *Memory*, 21(3), 286–300. <https://doi.org/10.1080/09658211.2012.725739>
- Mameniskiene, R., Jatuzis, D., Kaubrys, G., & Budrys, V. (2006). The decay of memory between delayed and long-term recall in patients with temporal lobe epilepsy. *Epilepsy and Behavior*, 8(1), 278–288. <https://doi.org/10.1016/j.yebeh.2005.11.003>
- Mckenzie, A., Yim, H., Stone, B., & Dennis, S. J. (2020). Using Experience Sampling to Investigate Affect at Encoding and Episodic Memory. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 42.
- Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Kilani, G., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2023). m-Path: an easy-to-use and highly tailorable platform for ecological momentary assessment and intervention in behavioral research and clinical practice. *Frontiers in Digital Health*, 5. <https://doi.org/10.3389/fdgth.2023.1182175>
- Milton, F., Fulford, J., Dance, C., Gaddum, J., Heuerman-Williamson, B., Jones, K., Knight, K. F., MacKisack, M., Winlove, C., & Zeman, A. (2021). Behavioral and Neural Signatures of Visual Imagery Vividness Extremes: Aphantasia versus Hyperphantasia. *Cerebral Cortex Communications*, 2(2). <https://doi.org/10.1093/texcom/tgab035>
- Moscovitch, M., Cabeza, R., Winocur, G., & Nadel, L. (2016). Episodic memory and beyond: The hippocampus and neocortex in transformation. *Annual Review of Psychology*, 67, 105–134. <https://doi.org/10.1146/annurev-psych-113011-143733>
- Palombo, D. J., Alain, C., Soderlund, H., Khuu, W., & Levine, B. (2015). Severely deficient autobiographical memory (SDAM) in healthy adults: A new mnemonic syndrome. *Neuropsychologia*, 72, 105–118. <https://doi.org/10.1016/j>
- Palombo, D. J., Sheldon, S., & Levine, B. (2018). Individual differences in autobiographical memory. *Trends in Cognitive Sciences*, 22(7), 583–597. <https://doi.org/10.1016/j.tics.2018.04.007>

- Pinho, P. J. M. R., Bertola, L., Ramos, A. A., Ghossain Barbosa, M., Rabelo, W., Castro-Costa, É., Lima-Costa, M. F., & Ferri, C. P. (2023). Subjective memory complaints: Prevalence, associated factors and sex differences in the ELSI-Brazil study. *International Journal of Geriatric Psychiatry*, 38(11). <https://doi.org/10.1002/gps.6026>
- Piolino, P., Desgrandes, B., Belliard, S., Matuszewski, V., Lavelée, C., De La Sayette, V., & Eustache, F. (2003). Autobiographical memory and autonoetic consciousness: triple dissociation in neurodegenerative diseases. *Brain*, 126(10), 2203–2219. <https://doi.org/10.1093/brain/awg222>
- Plancher, G., & Piolino, P. (2017). Virtual Reality for Assessment of Episodic Memory in Normal and Pathological Aging. In *The Role of Technology in Clinical Neuropsychology*. Oxford University Press. <https://doi.org/10.1093/oso/9780190234737.003.0015>
- Poitrenaud, J., Deweer, B., Kalafat, M., & Van der Linden, M. (2007). *CVLT test d'apprentissage et de mémoire verbale*. ECPA.
- Ramaseri Chandra, A. N., El Jamiy, F., & Reza, H. (2022). A Systematic Survey on Cybersickness in Virtual Environments. *Computers*, 11(51), 1–15. <https://doi.org/10.3390/computers11040051>
- Rayner, G., & Tailby, C. (2017). Current Concepts of Memory Disorder in Epilepsy: Edging Towards a Network Account. *Current Neurology and Neuroscience Reports*, 17(8). <https://doi.org/10.1007/s11910-017-0765-7>
- Reitan, R. M. (1971). Trail Making Test Results for Normal and Brain-Damaged Children. *Perceptual and Motor Skills*, 33(2), 575–581. <https://doi.org/10.2466/pms.1971.33.2.575>
- Rubin, D. C., Schrauf, R. W., & Greenberg, D. L. (2003). Belief and recollection of autobiographical memories. *Memory & Cognition*, 31(6), 887–901. <https://doi.org/10.3758/BF03196443>
- Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1997). The Psychological Construct of Word Fluency. *Brain and Language*, 57, 394–405. <https://doi.org/10.1006/brln.1997.1755>
- Sedighi, B. (2011). Memory impairment in multiple sclerosis and its determinant factors. *Neurosciences*, 16(1), 24–28.
- Sheldon, S., Amaral, R., & Levine, B. (2017). Individual differences in visual imagery determine how event information is remembered. *Memory*, 25(3), 360–369. <https://doi.org/10.1080/09658211.2016.1178777>
- Simpson, S., & Sheldon, S. (2020). Testing the Impact of Emotional Mood and Cue Characteristics on Detailed Autobiographical Memory Retrieval. *Emotion*, 20(6), 965–979. <https://doi.org/10.1037/emo0000603>
- Smulders, T. V., Black-Dominique, A., Choudhury, T. S., Constantinescu, S. E., Foka, K., Walker, T. J., Dick, K., Bradwel, S., McAllister-Williams, H., & Gallagher, P. (2017). A real-world what-where-when memory test. *Journal of Visualized Experiments*, 2017(123). <https://doi.org/10.3791/55646>
- Spooner, D. M., & Pachana, N. A. (2006). Ecological validity in neuropsychological assessment: A case for greater consideration in research with neurologically intact

- populations. *Archives of Clinical Neuropsychology*, 21(4), 327–337. <https://doi.org/10.1016/j.acn.2006.04.004>
- Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. *Cold Spring Harbor Perspectives in Biology*, 7(8). <https://doi.org/10.1101/cshperspect.a021766>
- St Jacques, P. L., & Levine, B. (2007). Ageing and autobiographical memory for emotional and neutral events. *Memory*, 2(15), 129–144. <https://doi.org/10.1080/09658210601119762>
- Streiner, D. L., & Kottner, J. (2014). Recommendations for reporting the results of studies of instrument and scale development and testing. *Journal of Advanced Nursing*, 70(9), 1970–1979. <https://doi.org/10.1111/jan.12402>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/https://doi.org/10.1037/h0054651>
- Suchy, Y., Desruisseaux, L. A., Gereau Mora, M., Brothers, S. L., & Niermeyer, M. A. (2024). Conceptualization of the term ecological validity in neuropsychological research on executive function assessment: A systematic review and call to action. *Journal of the International Neuropsychological Society*, 30(5), 499–522. <https://doi.org/10.1017/S1355617723000735>
- The jamovi project. (2025). *jamovi* (2.6). <https://www.jamovi.org>
- Troyer, A. K., & Rich, J. B. (2002). Psychometric Properties of a New Metamemory Questionnaire for Older Adults. *Journal of Gerontology*, 57(1), 19–27. <https://doi.org/https://doi.org/10.1093/geronb/57.1.P19>
- Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, 26(1), 1–12. <https://doi.org/10.1037/h0080017>
- Van der Elst, W., Van Boxtel, M. P. J., Van Breukelen, G. J. P., & Jolles, J. (2008). Detecting the significance of changes in performance on the Stroop Color-Word Test, Rey's Verbal Learning Test, and the Letter Digit Substitution Test: The regression-based change approach. *Journal of the International Neuropsychological Society*, 14(1), 71–80. <https://doi.org/10.1017/S1355617708080028>
- Visser, M., Forn, C., Gómez-Ibáñez, A., Rosell-Negre, P., Villanueva, V., & Ávila, C. (2019). Accelerated long-term forgetting in resected and seizure-free temporal lobe epilepsy patients. *Cortex*, 110, 80–91. <https://doi.org/10.1016/j.cortex.2018.02.017>
- Walker, W. R., Skowronski, J. J., Gibbons, J. A., Vogl, R. J., & Ritchie, T. D. (2009). Why people rehearse their memories: Frequency of use and relations to the intensity of emotions associated with autobiographical memories. *Memory*, 17(7), 760–773. <https://doi.org/10.1080/09658210903107846>
- Wang, Q. (2009). Once upon a Time: Explaining Cultural Differences in Episodic Specificity. *Social and Personality Psychology Compass*, 3(4), 413–432. <https://doi.org/10.1111/j.1751-9004.2009.00182.x>
- Wechsler, D. (2011). *WAIS-IV - Échelle d'intelligence de Wechsler pour adultes - 4ème édition*. Pearson.

- Wechsler, D. (2012). *MEM-IV - Echelle clinique de mémoire de Wechsler - 4ème édition*.
- Werner-Seidler, A., & Moulds, M. L. (2011). Autobiographical memory characteristics in depression vulnerability: Formerly depressed individuals recall less vivid positive memories. *Cognition and Emotion*, 25(6), 1087–1103. <https://doi.org/10.1080/02699931.2010.531007>
- Yeh, T. F., Chang, Y. C., Yang, S. Y., & Yang, C. C. (2025). Effect of Successful Aging on Technology Acceptance: The Moderating Role of Selection, Optimization, and Compensation Strategies. *Social Science Computer Review*, 43(3), 649–671. <https://doi.org/10.1177/08944393241258218>
- Yim, H., Garrett, P. M., & Dennis, S. J. (2018). Examining the Independence of Scales in Episodic Memory using Experience Sampling Data. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 40. <https://escholarship.org/uc/item/85h082b1>
- Zacks, J. M. (2020). Event Perception and Memory. *Annual Review of Psychology*, 71, 165–191. <https://doi.org/10.1146/annurev-psych-010419-051101>
- Zacks, J. M., & Sargent, J. Q. (2010). Event Perception: A Theory and Its Application to Clinical Neuroscience. In *Psychology of Learning and Motivation* (Vol. 53, Number C, pp. 253–299). [https://doi.org/10.1016/S0079-7421\(10\)53007-X](https://doi.org/10.1016/S0079-7421(10)53007-X)

