Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: http://ees.elsevier.com

Forgetting in Alzheimer's disease: Is it fast? Is it affected by repeated retrieval?

Andreea Stamate [a,b,*], Robert H. Logie [a,c], Alan D. Baddeley [d], Sergio Della Sala [a,c]

[a] Human Cognitive Neuroscience, Psychology Department, University of Edinburgh, UK
[b] University Suor Orsola Benincasa, Naples, Italy
[c] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, UK
[d] Department of Psychology, University of York, Heslington, York, UK

ARTICLE INFO

Keywords
Accelerated forgetting
Repeated testing
Alzheimer's disease

ABSTRACT

Objectives: Whether people with Alzheimer's Disease present with accelerated long term forgetting compared to healthy controls is still debated. Typically, accelerated long term forgetting implies testing the same participants repeatedly over several delays. This testing method raises the issue of confounding repetition effects with forgetting rates. We used a novel procedure to disentangle the two effects.

Methods: Four short stories were presented during an initial in-person assessment of 40 patients with Alzheimer's Disease and 42 age-matched healthy controls. Our aim was for participants to reach a score of 70% correct (9 out of 13 questions) at encoding. If this criterion was not achieved after the first trial, the four stories were presented again (in a different order); participants took the 1 min filler task again and were then retested. We repeated this process until participants reached the 70% criterion or to a maximum of four trials. Cued recall memory tests were completed during follow-up telephone call(s) at different delay intervals. Study material was presented only at encoding, then probed with different question sets on all other delays. Each question set tested different sub-parts of the material. The experiment employed a mixed design. Participants were randomly allocated to either a condition without retrieval practice or a condition with retrieval practice. Participants in the condition without retrieval practice were only tested at two delays: post encoding filled delay and at one month. Participants in the condition with retrieval practice were tested at four delays: post encoding filled delay, one day, one week and one month. Our methodological design allowed us to separate the effects of retesting from the effects of delay.

Results: Alzheimer's Disease patients showed a significant encoding deficit reflected in the higher number of trials required to reach criterion. Using Linear Mixed Models, we found no group by delay interactions between the post encoding filled delay retrieval and one month delays, with Alzheimer's Disease groups having a similar decline in performance to healthy controls, irrespective of testing condition. Significant condition by delay interactions were found for both groups (Alzheimer's Disease and healthy controls), with better performance at one month in the condition with retrieval practice.

Conclusions: Our data showed that Alzheimer's Disease is not characterised by accelerated long term forgetting, patients in our sample forgot at the same rate as healthy controls. Given the additional trials required by Alzheimer's patients to reach the 70% correct criterion, their memory impairment appears to be one of encoding. Moreover, Alzheimer's Disease patients benefited from repeated testing to the same extent as healthy controls. Due to our methodological design, we were also able to show that performance improved under repeated testing conditions, even with partial testing (sampling different features from each narrative on every test session/delay) in both healthy controls and Alzheimer's Disease.

Abbreviations

AD	Alzheimer's disease
HC	Healthy controls
MCI	Mild cognitive impairment
ALF	Accelerated long-term forgetting
MMSE	Mini-Mental State Examination

1. Introduction

Accelerated long-term forgetting (ALF) has been proposed as one of the main reasons for memory deficits in Alzheimer's Disease (AD) (e.g., Vallet et al., 2016). However, studies investigating whether AD patients present with ALF or not, have reported conflicting results (see Table 1). It has been suggested that these differences derive from methodological confounds

* Corresponding author. University of Edinburgh, 7 George Square, Edinburgh, EH8 9JZ, UK.
E-mail address: s1363074@ed.ac.uk (A. Stamate)

https://doi.org/10.1016/j.neuropsychologia.2020.107351
Received 12 August 2019; Received in revised form 11 December 2019; Accepted 15 January 2020
Available online xxx
0028-3932/

Table 1
Summary of studies investigating ALF in AD and MCI.

Authors	Sample	Material	Equated encoding	Delay	Recall, Recognition	Analysis	ALF	Floor effects	Ceiling effects
Kopelman (1985)	8 AD; 14 KS and 16 HC	Words (*BP test*) and Pictures (*HandP method*)	yes	10 min; 24 h; 1w	recognition	percentage correct (from 10 min score)	no	yes	no
Becker et al. (1987)	62 AD and 64 HC	Verbal passage and RCFT	no	I; 30 min	recall	difference in scores between acquisition and delayed-recall trials	no	yes	yes
Larrabee et al. (1993).	80 AD and 80 HC	Name-face associations and grocery list items	no (*but had a subset of participants who were*)	40 min (for name-face associations) and 30 min (for grocery list)	recall	difference between scores on final acquisition and delayed-recall trials	yes	no	no
Carlesimo et al. (1995)	13 AD, 8 MID, 9 Amn and 32 HC	Word list recall (*RCFT's 15 words learning task)* and Pictures (*HandP method)*	yes	I; 15 min (*RCFT)* And 90s; 10 min; 1 h; 24 h (HandP)	recall and recognition	*RCFT's 15 words learning task)* comparison between recall (fifth I) and delayed trial (*RCFT's 15 words learning task)* and whole number of correct responses *(Pictures HandP method*)	yes	no	no
Hart et al., 1987	14 AD; 10 MAD and 14 HC	Line drawings of common objects	no	10 min; 2 h; 48 h	recognition	percentage correct	yes	no	no
Hart et al., 1988	10 AD and 13 HC	Line drawings of common objects	yes	10 min; 2 h; 48 h	recognition	percentage correct	yes at 10 min, but not at 2 h and 48 h	no	no
Greene et al. (1996)	33 AD and 30 HC	Prose, word list, doors and people test	no	I, 30min	recall and recognition	scaled scores	no	yes	yes
Christensen et al. (1998)	15 AD and 15 HC	Picture, forced choice word, forced choice design, a picture recall task and a stem completion task	yes (*only a subset)*	1 min; 10 min; 20 min	recall and recognition	absolute scores and percentage correct	no, for all except picture recall	no	no
Degenszajn et al. (2001)	15 AD and 15 HC	Buschke selective reminding test	no	I, 30 min; 24 h	recall	items recalled at the sixth trial of the learning phase compared to total recall at 30 m and 24 h	no	no	yes
Budson et al., 2007	14 AD, 19 MCI and 22 HC	Real world events	no	initial weeks after the event; three to four months later; one year	recall and recognition	proportion correct	yes, at initial weeks and 3/4-month assessment; not after one year	yes	no
Manes et al. (2008)	10 SMC; 7 SMC with objective memory impairment and 9 HC	Verbal and visual material	no	I; 30 min; 6 weeks	recall and recognition	absolute scores	yes	no	no
Walsh et al. (2014).	15 MCI and 15 HC	Story learning task	yes	I (*last learning trial*); 30 min; 1 week	recall	slope of the linear regression between data points	yes	no	no
Vallet et al. (2016).	16 AD; 16 MCI and 16 HC	DMS-48	no	3 min; 1 h; 1 week	recognition	absolute scores	no	no	no
Lombardi et al. (2018).	16 aMCI and 19 HC	120 words (*HandP*); recollection or familiarity judgements	no	10 min; 1 h; and 24 h	recognition	percentage correct	no	yes	no

(Geurts et al., 2015). Table 1 summarises the literature investigating ALF in AD and prodromal syndromes. Half of the fourteen studies we could glean from the literature found normal long-term forgetting patterns compared to those of healthy controls (HC). We identified several factors that could account for this discrepancy in results.

Firstly, although this is not always acknowledged, a possible confounding factor is whether there are ceiling effects in the performance of HC or floor effects in the patient samples. Four out of the fourteen studies listed in Table 1 are marred by floor effects in the clinical sample (Kopelman, 1985 p. 634; Greene et al., 1996, p. 545; Budson et al., 2007, p. 887; Lombardi et al., 2018, p.8) while three are difficult to interpret given the ceiling effect in the control group (Greene et al., 1996, p. 545; Degenszajn et al., 2001, p.173; Weston et al., 2018, p. 130).

AD: Alzheimer's disease; HC: Healthy controls; KS: Korsakoff's syndrome; MCI: mild cognitive impairment; aMCI: amnestic mild cognitive impairment; MID: multi-infarct demented; SMC: subjective memory complaints; BP test: Brown-Peterson test; HandP: Huppert and Piercy; Amn: amnesics; MAD: major affective disorder; RCFT: Rey complex figure test;

ALF: accelerated long-term forgetting; eFAD: Presymptomatic autosomal dominant familial Alzheimer's disease; I: immediate.

Secondly, many studies failed to equate baseline performance between the clinical and the healthy group, leading to a possible incorrect assessment of the differences in the forgetting rates between the two groups. Greene et al. (1996) evaluated anterograde episodic memory in patients with AD and in HC using immediate and delayed prose recall. They reported that once initial acquisition of new information on the task was equated across groups, patients with AD did not exhibit ALF. Similarly, Kopelman (1985), using the Huppert-Piercy test, found no evidence of ALF at 24 h or 7 days delay, after matching initial learning. On the contrary, Carlesimo et al. (1995) did observe ALF in AD patients at 1 h and 24 h delays on a line drawing recognition task. Recently, Weston et al. (2018) investigating a group of people affected by a gene mutation resulting in a form of presymptomatic autosomal dominant AD found that these people had a performance similar to HC at initial learning and 30-min recall on a series of tests (word lists, stories, and figure recall). However, when assessed again after a week, people carrying the mutation had forgotten more than the non-carriers. These differences in findings cannot be attributed solely to whether initial performance was equated or not, to the type of material or testing method (recall/recognition). An additional influencing factor in investigating forgetting derives from the fact that repeated testing is inherent in the study of forgetting, but repeated testing comes with several caveats. One would be, as Weston et al. (2018) noted, that we cannot control for some participants rehearsing or at least recalling the material between assessments. The authors comment on the difficulties arising with repeated measures and argue for the importance of identifying new methods of assessment. They propose either to embed testing material amongst other unrelated cognitive tests, or to use recognition tests with material that would be difficult to rehearse by participants between test sessions.

Some of the previous studies have discussed the possible implications of repeated testing on patients' performance (Greene et al., 1996; Weston et al., 2018). However, none has directly investigated the effects of such repetitions, and whether the same material or different material is used on each test session. In an attempt to address the difficulties arising with repeated testing, a number of approaches have been identified (for a review see Elliot et al., 2014). Baddeley et al. (2019) propose to use material that once learned can be used to test the same individual over longer delays, repeatedly, without testing the same information on each occasion. From the review of the 14 studies on ALF in AD, listed in Table 1, the issue of whether or not the same material was retested on each delay emerges as one of the differentiating factors between studies that have reported ALF and those which have not. Seven of the 14 studies that investigated ALF in AD patients, used different subsets of the initially encoded material on each testing session. These six studies documented forgetting rates in AD and aMCI similar to that of age-matched controls (Kopelman, 1985; Hart et al., 1987; Hart et al., 1988; Christensen et al., 1998; Vallet et al., 2016; Lombardi et al., 2018).

Lastly, we agree with Weston et al. (2018) in that repeated measures, and more importantly rehearsal raise important methodological issues. Repeated testing of the same material involves (re)learning of that material on each subsequent testing occasion. However, when different subsets of the initially encoded material are tested on each of the following delays, particularly if no feedback is given, relearning is minimised. These two types of testing procedures could lead to large differences in memory performance between individuals with learning deficits and normal groups, with healthy adults benefiting more from the relearning opportunities compared to patients. In a previous study of ours (under review) we have suggested that memory performance benefits from repeated partial testing (testing different subparts of initially taught material) arise as a result of priming, rather than relearning. If this is to be the case, then amnesic patients should benefit to the same extent as HC as a result of repeated partial testing, thus eliminating the difference in forgetting slopes between the two groups. To surmise if repeated testing provides a new learning opportunity, individuals with learning deficits could potentially be mistaken as exhibiting ALF since they benefit from relearning to a lesser extent, compared to healthy individuals. On the other hand, if it represents priming, then patients with amnesia, such as those with AD, should also exhibit relatively preserved long-term memory performance under repeated partial testing, as the act of repetition would serve to strengthen existing representations thus also benefiting AD patients.

In a recent methodological review of ALF studies, Elliot et al. (2014) concluded that several key factors must be considered when assessing longer-term forgetting. Among their recommendations they suggest that when assessing ALF, tests should allow for repeated testing, while avoiding repeated retrieval as much as possible by using distinctive matched tests. Furthermore, standardised tests of ALF should allow for free recall and cued recall testing, or some type of testing with retrieval support. The Crimes Test (Baddeley et al., 2013) meets both these requirements. This prose recall test is composed of four short stories, each based on an incidence of crime that contains five key features (e.g. the crime, the criminal, the location). It does not demand excessive (initial) learning time and allows for different subsamples of questions to be tested via cued recall after a range of delays. In a later study, Baddeley et al. (2019) ran two experiments each comprising a repeated testing condition (testing on: immediate, 24 h, one week and one month) and a condition involving a single test after one month. They found that both the Crimes test and a visual test showed clear evidence of forgetting in the single test condition but little evidence of forgetting in the repeated testing condition. The authors suggested that the testing of individual features (subsamples of questions) enabled participants to remember the entire episode which then acted as a further reminder. This lack of forgetting in healthy individuals could provide an ideal test of ALF by avoiding the danger of floor effects (Baddeley et al., 2019). In the current study, we have addressed the question of whether or not ALF does characterise the memory deficits of AD patients using the procedures devised by Baddeley et al. (2019) and designed material closely following The Crimes Test (Baddeley et al., 2013).

We have also addressed a second question, namely whether the performance of AD patients is enhanced by repeated testing. Several studies have shown the advantage of repeated testing on memory performance (Carpenter and Pashler, 2007; Pilotti et al., 2009; Thomas et al., 2018; Baddeley et al., 2019). This enhancement in performance due to retesting, referred to as the testing effect, has been shown in applied situations, including educational settings (e.g., Roediger and Butler, 2011), in healthy older adults (e.g., Ferrer et al., 2004; Baddeley et al., 2019), and to some extent in individuals with memory impairments (e.g., Yan and Dick, 2006; Duff et al., 2008). While the testing effect emerges when tests probe the entire encoded material, when evaluating the effect of partial testing (probing subparts of that material) different viewpoints emerge on how this influences final memory performance. Some suggest that the benefits that arise as a result of partial testing apply only to material that can be integrated, or reconstructed by participants (e.g. prose, video as opposed to individual words, or pictures). However none of the studies which directly address partial testing effects have investigated these issues in clinical samples. A detailed review of the literature investigating partial testing in healthy samples is beyond the scope of the present article (for a discussion see: e.g. Baddeley et al., 2019; Chan et al., 2015; Chan, 2009).

Some indirect evidence suggesting that repeated testing would prove beneficial to AD patients comes from reports which have shown that increasing the delays between testing when recalling information repeatedly (spaced retrieval) can improve memory performance for dementia patients and amnesiacs (e.g., Cull et al., 1996; Brush and Camp, 1998). Recalling information repeatedly has been shown to improve AD patients' performance on: object–location associations (Camp and Stevens, 1990), names of different objects (Abrahams and Camp, 1993) and prospective memory tasks (Camp et al., 1996). For example, Kinsella et al. (2007) investigated the benefits of spaced retrieval for improving prospective memory performance in patients with early AD compared to healthy older adults and found that the performance of most AD patients improved as a result of spaced-re-

trieval (combined with elaborated encoding of the task). However, experiments aiming at studying retrieval practice in dementia patients have generally focused on simple cognitive tasks such as face-name associations, object-name or object-location associations, and cue-behaviour associations (see Creighton et al., 2013). The current experiment looks at a more complex task, remembering associations between multiple features within stories.

2. Method

2.1. Participants

2.1.1. Patient sample

The patients were recruited from various geriatric institutions in Bucharest (Romania). Participants' eligibility for the AD group was restricted to patients with a diagnosis of probable AD, confirmed at 6 months follow-up, based on international diagnostic criteria (NINCDS-ADRDA: McKhann et al., 1984; DSM-IV-TR: American Psychiatric Association, 2000). Patients included in the study should have a Mini-Mental State Examination (MMSE) score between 26 and 18. They were assessed with a range of standard memory and global cognition tests (see Table 2) and with a paper version of the Temporary Memory Binding test (Della Sala et al., 2018) by the experimenter (first author). Patients also underwent blood screening tests to exclude other potential causes of dementia, all had CT scans, and a few had MRI scans as well. Patients were excluded from the study if they had a past history of stroke, brain traumatic injury, clinical depression or alcoholism. Due to the nature of the testing material, individuals with major hearing impairments were also excluded. Written consent from all patients, or their caregivers was obtained according to the Declaration of Helsinki, as was ethical approval from the relevant ethics committees of each institution involved (Institutul National de Gerontologie si Geriatrie "Ana Aslan" Bucureşti; Spitalul Universitar de Urgenta ELIAS Sectia Geriatrie Gerontologie Bucuresti; Clinica Pro-memoria Bucuresti).

AD: Alzheimer's disease; DS: Digit Span (Blackburn et al., 1957); ADL: Activities of Daily Living (Katz, 1983); IADL: Instrumental Activities of Daily Living (Lawton, and Brody, 1969); CDT: Clock Drawing (Shulman, 2000); GDS: The Geriatric Depression Scale (Yesavage et al., 1983); MoCA: Montreal Cognitive Assessment (Nasreddine et al., 2005); TMB: Temporary Memory Binding test (Della Sala, Kozlova, Stamate, and Parra, 2018).

2.1.2. Healthy controls

The healthy control (HC) sample was recruited in Romania from GP surgeries and from the local communities. The GPs provided a list of older individuals who were registered with their practice whose medical files showed they were in good health. In Romania, GPs perform regular general examinations of their patients, including cognitive assessment. All the participants included in the study were healthy at the time of testing. Exclusion criteria for the HC were: the absence of psychiatric or neurological conditions, including alcohol or drug abuse or head trauma and a MMSE score higher than 28. This latter criterion was documented by GP records. Written consent from all participants was obtained.

Table 2
AD patients' performance on the background Neuropsychological test battery.

Test	AD participants' scores		
	Range	Mean	Std. Deviation
DS (0–10)	3–8	4.6	0.9
ADL (0–10)	3–6	5.2	0.8
IADL (0–8)	2–8	6.2	1.9
CDT (0–10)	2–10	8	1.9
GDS (0–15)	1–14	7.3	2.6
MoCA (0–30)	10–26	19.3	3.8
TMB (0–32)	13–29	20.6	3.9

2.1.3. Comparison between groups

The initial sample included 40 patients with AD (seven men and 33 women) and 44 HC (10 men and 32 women). The HC participants were recruited to match AD patients on age, educational level and when possible gender. The AD participants ranged in age from 55 to 93 years with a mean age of 77.4 years (S.D. = 8.4 years) while HC ranged in age from 56 to 89 years with a mean age of 75.6 years (S.D. = 8.2 years), there was no statistically significant difference between AD and HC on age ($t = -0.990$; $p = .326$). The AD participants ranged from 4 to 16 years with a mean of 12.7 (SD = 3.7) on level of education, and the HC ranged from 7 to 18 years with a mean of 13.5 (SD = 2.8). There was no statistically significant difference between AD and HC on level of education ($t = -0.988$; $p = .326$).

The final sample included 33 AD patients and 42 HC. Four participants refused to take part on following testing delays (two patients and two controls); one patient had a cerebral stroke between the one week and one month testing delay; the performance of one patient in the condition without retrieval practice was excluded as flagged as a significant outlier and 7 patients were not included in the final analysis as they did not reach the 70% encoding criterion.

Table 3 details the demographics of the subgroups (AD & HC) according to experimental conditions.

2.2. Design

All testing was conducted in Romanian, all neuropsychological tests which were carried out had translated and validated Romanian versions. With regard to the Fables test, even though we initially devised it in English, we (the first author) have translated it in Romanian and have used it in a previous experiment on a large (N = 240) Romanian sample of both young and old participants.

The experiment employed a mixed design. Participants were randomly allocated to either a condition without retrieval practice or one with retrieval practice. Participants in the condition without retrieval practice were only tested at two delays: post encoding filled delay and one month. Participants in the condition with retrieval practice were tested at four delays: post encoding filled delay, one day, one week and one month.

During the encoding phase, all participants were presented with four fables read out by the experimenter at a slow and clear pace (2s pause between

Table 3
Demographic variables and MMSE scores of AD and HC groups subdivided by testing condition.

GROUP		Range	Mean	Std. Deviation
AD Repeated Testing (N = 19)	Age	55–88	76.5	8.3
	Education	4–16	12.1	3.9
	MMSE	19–26	23.9	2.4
AD Single Testing (N = 14)	Age	67–93	78.7	8.7
	Education	7–16	13.5	3.3
	MMSE	18–26	22.9	2.9
HC Repeated Testing (N = 21)	Age	56–85	73.6	7.7
	Education	8–16	13.4	2.6
	MMSE	29–30	29.5	0.5
HC Single Testing (N = 21)	Age	62–89	77.4	8.3
	Education	7–18	13.4	3.1
	MMSE	29–30	29.7	0.4

AD: Alzheimer's disease; HC: Healthy controls; MMSE: Mini-Mental State Examination.

each sentence and 5s pause between each fable). To minimise any recency effects, each presentation phase was followed by a written 1 min filler task, involving finding as many words as possible from the letters composing the Romanian word "hippopotam" (see Baddeley et al., 2019). Participants then took the initial post encoding filled delay cued recall test on one subset of questions (there were four subsets in total), which was self-paced. If participants scored less than 70% correct (9 out of 13 questions), the four fables were presented again (in a different order); participants took the 1 min filler task again and were then retested. Our aim was to repeat this process until participants reached the 70% criterion or to a maximum of four trials.

The subsets were randomised both during the encoding phase (in the cases where more trials were needed) and across the various testing delays. In condition without retrieval practice (former single test) one of the subsets not tested at 1-min was randomly selected. In the condition with retrieval practice testing material changed at each delay. The encoding phase and initial test were conducted face to face while all other tests were conducted by telephone. This type of testing, telephone follow-up, has been validated by Baddeley et al. (2013) and used successfully in other studies with similar procedures (Baddeley et al., 2019) as well as studies involving different clinical samples (Walsh et al., 2014).

2.3. Material

The material comprised a simplified version of the Fables test previously devised for another study investigating the effects of partial repeated testing on forgetting in younger and older healthy individuals. After piloting with a small AD group, the Fables test was modified to make it more accessible for clinical use (Supplementary material for details). The material used in this experiment consisted of four fables loosely mimicking Aesop's style. Each was four sentences-long and involved eight main features (e.g., characters, nationality, moral of the fable, etc.; full material in the Supplementary material). This generated 52 questions, which were split across four subsets. Each question in the subsets probed one sentence from each of the four fables, without ever probing the same feature twice (in the same story) within the same subset. All materials were presented in Romanian. The original Aesop's stories are not part of the Romanian culture, not only did we select unrenowned fables, but we also enquired (some participants) at the end of the experiment if any of these were even vaguely familiar to them to ensure they were not.

3. Data availability

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

4. Results

4.1. Initial learning

There was a significant difference between the two groups in the number of trials necessary to reach the criterion performance level set at 70% correct ($t = 7.647$, $P < .001$) with AD groups requiring more trials ($M = 2.64$, $SD = 0.86$) than the HC groups ($M = 1.4$, SD.48). Cohen's effect size value ($d = 1.673$) suggested that the effect of group on the number of trials required to reach the 70% criterion was highly significant. Among the 42 HC, 27 required one trial and the remaining required two trials to reach criterion. Out of the 40 AD patients, two required one trial, 15 required two trials, 13 three trials and 3 four trials. Seven AD patients who did not reach the 70% criterion were excluded from the analysis. Therefore the final AD sample included in the analysis below consisted of the 33 AD patients who had reached criterion at encoding. Even after excluding the AD patients who did not reach criterion, the number of trials to reach this criterion was still not equal between AD and HC.

Mixed effects models were used to examine how groups (AD vs. HC) and testing condition (without retrieval practice, with retrieval practice) may have affected recall performance at different delays. In order to control for individual variability among participants we used a model assuming random intercepts and random slopes for each participant, and a covariance structure to account for heterogenous variances at different delays in each linear mixed-effect. Further information on individual performance can be found in Fig. 1 and in the Supplementary material in the Individual performance data and tables section.

Random intercepts and an unstructured covariance matrix were used to account for within-subject correlations. A random effect of delay was also

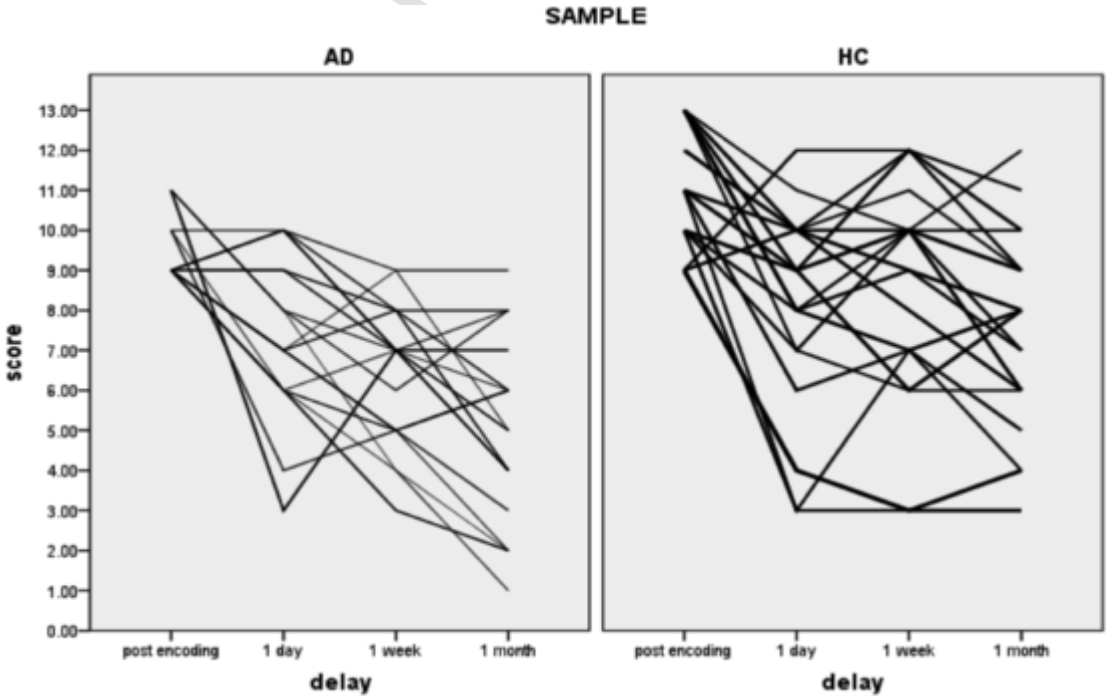

Fig. 1. Individual recall performance on the Fables test in the immediate, one day, one week and one-month tests in the AD and HC groups. AD: Alzheimer's disease; HC: Healthy controls.

included in order to measure the variance in the effects of delay on scores, across participants. The significance of each fixed effect in predicting each behavioural outcome measure was assessed with $\alpha = 0.05$. A total of 230 data points were available for statistical analyses. Mean scores at different time intervals for each of the 4 groups are displayed in Table 4.

Table 4
Mean correct scores on the Fables test at post-encoding retrieval, one day, one week and one month test sessions for AD and HC groups.

GROUP	Delay	Range	Mean	Std. Deviation
AD Repeated Testing	Post-encoding retrieval	9–11	9.5	0.8
	One day	3–10	7.3	2.03
	One week	3–9	6.4	1.7
	One month	1–9	5.1	2.3
AD Single Testing	Post-encoding retrieval	9–10	9.29	0.47
	One month	0–4	2	1.5
HC Repeated Testing	Post-encoding retrieval	9–13	10.8	1.6
	One day	3–12	8	2.6
	One week	3–12	8.3	2.8
	One month	3–12	7.4	2.5
HC Single Testing	Post-encoding retrieval	9–12	9.9	0.9
	One month	1–7	3.4	1.6

AD: Alzheimer's disease; HC: Healthy controls.

4.2. Accelerated long-term forgetting in AD

The first mixed effects model compared recall performance across two delay intervals only (post encoding filled delay retrieval and one month) between AD and HC samples, separately for each condition. The model included correct scores as the dependent variable and 2 factors: delay with two levels (post encoding filled delay retrieval and one month) and sample (AD and HC). Significant main effects were found in each testing condition for delay (Without retrieval practice condition: $F(1, 33) = 491.851$, $P < .001$; With retrieval practice condition: $F(1, 38) = 88.360$, $P < .001$) and sample (Without retrieval practice condition: $F(1, 33) = 12.441$, $P < .001$; With retrieval practice condition: $F(1, 38) = 15.345$, $P < .001$) however there was no significant interaction between delay and sample in any of the experimental conditions (Without retrieval practice condition: $F=(1, 33) = 1.921$, $p = .175$; With retrieval practice condition: $F(1, 38) = 1.546$, $p = .221$).

Pairwise Comparisons showed that HC performed significantly better than AD at post-encoding retrieval test (MD = − 1.28 SE = 0.41, $P < .001 = 0.004$) and at one month test (MD = 2.32 SE = 0.77, $p = .005$) in the condition with retrieval practice as well as in the condition without retrieval practice (post-encoding retrieval test (MD = 0.62 SE = 0.26, $P < .001 = 0.023$) and at one month test (MD = 1.47 SE = 0.55, $p < .001$). Thus, HC participants had a significantly better performance on post-encoding retrieval test and at one month test compared to AD, in both conditions, however there is no evidence of a difference between the rate of forgetting over one month delay in AD group compared to the HC in any testing condition (forgetting rates from post-encoding retrieval to one month were essentially parallel between the groups - Fig. 2).

4.3. The testing effect

We ran a linear mixed effects model with main effects of delay, condition and sample and their interactions including the three-way interaction between all main effects as predictors. All three main effects, and the interaction between delay and condition, reached significance. The three-way interaction between delay, sample and condition was not significant ($F(2, 71.000) = 1.140$, $p. = 0.326$).

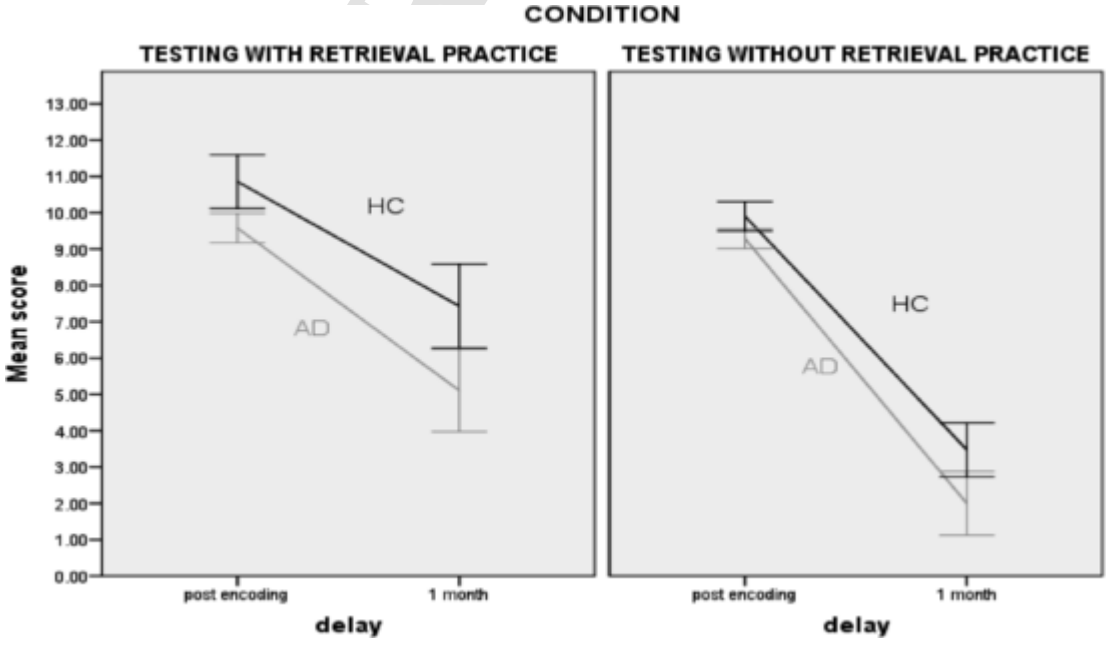

Fig. 2. *Mean recall performance on the Fables test at post-encoding retrieval and one-month delays as a function of group (AD and HC) in both testing conditions (single testing; repeated testing).*AD: Alzheimer's disease; HC: Healthy controls.

The second mixed effects model investigated the change in recall performance (mean correct scores) across 2 delay intervals (post-encoding retrieval; one month) between the 2 conditions (condition without retrieval practice vs. condition with retrieval practice). The analysis was performed separately for each group (AD, HC). Where statistically significant differences between conditions in rate of decline (i.e., a significant condition by delay interaction) were identified, model-based estimates for each delay were computed.

Significant main effects were found in each sample for delay (AD: ($F(1, 27) = 218.408$ $P < .001$); HC: ($F(1,40) = 185.253$ $P < .001$) and condition (AD:$F(1, 17) = 18.621$, $P < 001$; HC: $F(1,40) = 35.926$, $P < .001$). There was also a significant interaction between delay and condition in each group (AD: ($F (1, 27) = 10.515$ $P < .001$); HC: $F (1,40) = 35.926$, $P < .001$). AD participants in the condition with retrieval practice ($M = 5.1$, $SE = 0.47$) performed significantly better at one month ($MD = 3.105$, $SE = 0.721$, $p < .001$) compared to AD participants in the without retrieval practice condition ($M = 2$, $SE = 0.547$) while their performance on post-encoding retrieval test was similar ($MD = 0.293$, $SE = 0.327$, $p = .416$; (AD -With retrieval practice condition: $M = 9.58$, $SE = 0.21$; AD- Without retrieval practice condition: $M = 9.29$, $SE = 0.25$; $MD = 0.29$ $SE = 0.33$, $p = .416$). Three AD participants in the condition without retrieval practice performed at floor at the one month assessment.

HC participants in the with retrieval practice condition ($M = 7.43$, $SE = 0.47$) performed significantly better at one month test ($MD = 0.3.95$, $SE = .66$ $P<. 001$, Cohen's $d = 1.896$) than HC participants in the without retrieval practice condition ($M = 3.48$, $SE = 0.47$), there was also a statistically significant difference in post-encoding retrieval mean scores ($MD = 0.95$ $SE = .40$ $p = . 023$) with higher mean scores in the with retrieval practice condition ($M = 10.88$ $SE = 0.29$) compared to HC in the without retrieval practice condition ($M = 9.91$ $SE = 0.29$). A one-way ANCOVA was conducted with the scores from the HC group to compare the effect of condition on performance at one month test whilst controlling for scores on post-encoding retrieval test. Results showed that the significant effect of condition still holds ($F (14,39) = 28.092$, $P < .001$). Therefore, the HC participants in the with retrieval practice condition performed significantly better at one month test compared to HC participants in the without retrieval practice condition even after controlling for the differences in performance on post-encoding retrieval test (Fig. 3).

4.4. Summary of results

AD patients showed a significant learning deficit (requiring more trials to reach criterion) and significantly impaired recall performance on post-encoding retrieval test, as well as at one month test compared to HC. However, AD patients did not show ALF between post-encoding retrieval and the one month test in any of the testing condition.

In both conditions both groups declined in recall performance at one month test compared to post-encoding retrieval test, but the decline was significantly smaller for the groups in the condition with retrieval practice (See Fig. 3). This suggests that repeated-testing reduces forgetting at one month delay, producing gains in long-term retention in both AD and HC, even when retesting does not involve relearning of the tested material as different features of the initially learnt material were probed at each trial.

5. Discussion

Our study had two aims: (1) to investigate whether people with AD show accelerated long-term forgetting (ALF) relative to HC and (2) to investigate whether people with AD benefit from repeated testing.

5.1. Accelerated long-term forgetting in AD

Some authors have argued that AD memory impairment is characterised predominantly by an acquisition deficit (e.g., Kopelman, 1985; Greene et al., 1996; Grober and Kawas, 1997), whereas others have emphasised forgetting (e.g., Moss et al., 1986; Hart et al., 1988).

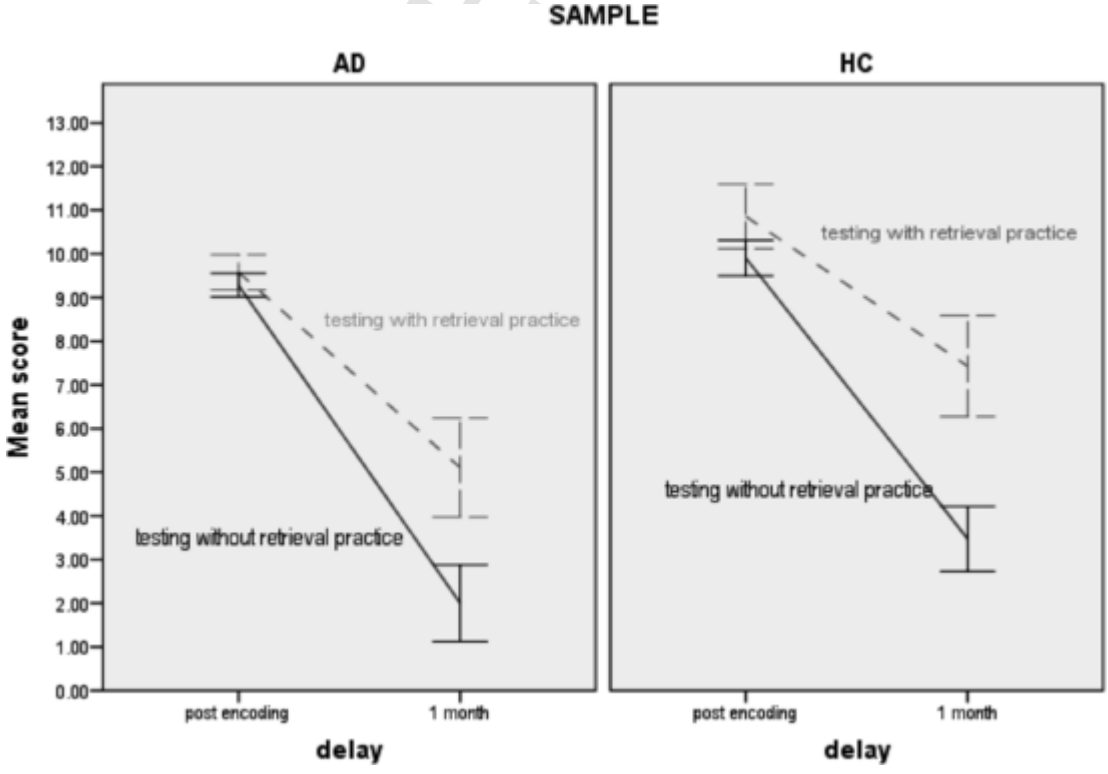

Fig. 3. Mean recall performance on the Fables test at post-encoding retrieval and one-month delays as a function of condition (single testing vs repeated testing) by the AD and HC groups.AD: Alzheimer's disease; HC: Healthy controls.

The AD patients in our study did differ from HC in learning rate and showed impaired performance compared to HC at all testing delays. Patients also needed more trials to reach criterion compared to HC. Loftus (1985) has noted that differences in initial learning ability may confound analyses of forgetting rates. Other authors have also suggested that forgetting rates may be underestimated in a lower-performing group, as they have less material to forget. In the present study, we attempted to avoid this pitfall by training all participants to a preset criterion (70% correct). All participants reached this criterion (after varying encoding trials), apart from seven patients who did not, and were excluded from the statistical analysis. Equating performance between patients and healthy participants can however present with its own limitation. Isaac and Mayes (1999a, 1999b) mention that matching procedures can in turn bias against findings that amnesiacs forget faster than controls. In order to match groups at encoding, patients invariably need longer or multiple exposures to test material compared to controls. Therefore, because the mean item-presentation-to-test delay is longer for patients, this can lead to an underestimate of the patients' forgetting rate (). When two memories are of the same strength, but different ages, the older one will generally decline slower (see Mayes, 1986, 1988). Our design cannot exclude these possible very early consolidation differences between Alzheimer patients and controls.

The results of our study speak against the occurrence of accelerated forgetting of verbal material in AD patients over the course of one month. When comparing performance from post-encoding retrieval to one month test, AD patients did not show ALF in either the condition with retrieval practice or the condition without retrieval practice.

When investigating ALF a combination of recognition and free recall is recommended (Elliot et al., 2014). We acknowledge the lack of a free recall measure as a limitation of the current experiment. A free recall measure could be easily devised for the current test (as in the case of the Crimes test- Baddeley et al., 2013). However, free recall would be affected by disturbances of executive functions and attention that typically characterise dementia, in addition to anxiety or depression (Cerciello et al., 2017). It is also likely to reflect the level of motivation. Recognition is less affected by these variables (Cerciello et al., 2017). The present study was influenced by the Crimes Test study (Baddeley et al., 2013) where unpublished research (Alber, 2014) showed more variability within a normal sample for free than for cued recall, presumably because cuing reduces the influence of strategy and criterion effects.

5.2. The testing effect

We compared the performance of the 33 people with AD with that of the 42 age and education matched HC on the Fables cued recall task. By splitting both samples into two groups based on the testing condition (condition with retrieval practice vs the condition without retrieval practice) we were able to disentangle the effect of repeated testing from that of forgetting, thus accurately measuring the impact of repetition on final performance. Three of the AD patients had reached floor, at one month, in the condition without retrieval practice. However, ceiling and floor effects are considered to be a problem if more than 15–20% of respondents achieved either the best or worst possible score (Garin, 2014). The 3 AD patients do not represent more than 15–20% of our sample. Both AD patients and HC in the condition without retrieval practice showed significantly faster forgetting at one month delay compared to the condition with retrieval practice. Therefore, the condition with retrieval practice benefited both HC and AD participants.

We should however acknowledge that repeated testing is not the only factor which can affect differences in forgetting rates. Several studies have found differences based to type of assessment, e.g. free recall versus recognition (Green and Kopelman, 2002; Kopelman and Stanhope, 1997;), type of material, e.g. verbal versus visuo-spatial material (Lucchelli and Spinnler, 1998; Manes et al., 2005; Davidson et al., 2007) and possibly test difficulty (Freed and Corkin, 1988). ,) found accelerated rates of forgetting for semantically related word lists and normal rates for free recall of lists of unrelated words in amnesics. However, recognition and cued recall of both kinds of word lists appeared to decline at a normal rate. They interpret these differences in forgetting patterns as arising from impairments in long-term memory consolidation for complex associations (between 2 or more items). While our material does examine complex associations (between several features), our results may only apply to material that is integrated (such as narrative) where probing one aspect of an integrated narrative might activate the entire narrative. While in the case of material with lower integration, this might not be the case. Probing subparts of material that is not integrated (such as individual words or images), may fail to prime recall of the other subparts.

Additionally, while the use of truly independent items and test forms would probably produce no benefits in performance with repeated testing, they also raise several issues. These would require more intensive initial learning time and would be more challenging to use with patients (Baddeley et al., 2019). Several approaches to repeated testing have been adopted in previous studies. Cassel et al., 2016 studied memory for verbal and visuo-spatial material over delays between 30 s and a week in temporal lobe epilepsy patients. They initially required participants to learn four separate stories and four routes, then tested retention of one story and one route per delay. Their method has the advantage of testing each item once. However the drawback is a relatively heavy initial learning load, though the encoding criterion was of only six out of a possible ten correct answers. This procedure can limit potential sensitivity to scores between zero and six at each testing occasion, in some participants. A further problem is that of serial order effects during initial learning potentially favouring primacy, recency or both, which may be further complicated by test order and possible between-test interference effects (Baddeley et al., 2019). Similarly, Jansari et al., 2010 tested a single patient with temporal lobe epilepsy using ten stories, testing two at each of five delays, one by recall and one by recognition. Evidence of ALF was observed that was not found when the same story was tested repeatedly. Jansari et al., 2010 study provides important information, however requiring participants to learn ten stories would make this test impracticable with a clinical population.

Nonetheless, the fact that both AD and HC benefit from repeated testing to the same extent can have major practical implications. Repeated testing can thus be employed to avoid floor effects (a frequent methodological confound) in studies comparing forgetting rates between AD and HC, without compromising the validity of the comparison.

6. Conclusion

To the best of our knowledge, this study presents the first assessment of long-term forgetting in AD patients over an interval of one month. It is also the first study to compare forgetting rates in AD under a condition with retrieval practice to a one without retrieval practice. By doing so we were able to uncover the importance of the number of tests and the length of test intervals when comparing forgetting rates in clinical and healthy groups over longer periods of time than have been common in previous studies.

Compared to the majority of studies on practice effects, which use within subjects' design, we employed a between subjects' design that allowed us to separate the effects of retesting from the effects of delay. Therefore, we are able to quantify more accurately the magnitude of this effect and show that performance is improved under repeated testing conditions, even with partial testing (sampling different features from each fable on every test session/delay).

Our results have potential practical implications in designing strategies/interventions for AD, as well as informing methodological design in clinical trials. Firstly, interventions that can be demonstrated to be efficient in aiding patients to remember important information over prolonged periods of time, are increasingly needed. Both patients and carers seek practical advice from professionals on neuropsychological interventions that will engage remaining capabilities of AD patients and are proved to promote and prolong independent functioning (Camp, 2001; Clare et al., 2002; Clare and Woods,

2004). Our results offer supporting evidence that repeated testing can be used to improve AD patients long-term memory performance. Secondly, repeated testing is used in clinical assessment as well as in clinical trials and research, the evidence that repeated testing (even when only subparts of material are being tested) increases performance for both healthy and clinical patients' needs to be carefully taken into account when employing this type of design. Practice effects have been shown to result in type 1 or type 2 errors (Goldberg et al., 2015). Goldberg et al. (2015) have drawn attention to the fact that ignoring practice-effect-related gains in performance produce large sources of errors and increase the likelihood of misinterpretation of the outcomes of clinical trials.

In conclusion, our study adds to the previous literature showing that memory impairment in AD disease is primarily characterised as an encoding, or storage deficits, rather than as accelerated forgetting. Our study also shows that re-testing at multiple delay increases long-term memory performance compared to a single test. The beneficial effect of re-testing holds also in people with AD.

Funding

The work was supported by The University of Edinburgh, United Kingdom, through the Research Support Grants for data collection.

Declaration of competing interest

The authors report no competing interests.

Acknowledgements

We are extremely grateful to Dr. Florina Cristescu, Pro-memoria Clinic Bucharest; Dr. Rozeta Draghici, Ana Aslan Geriatric Institute Bucharest and Prof. Dr. Luiza Spiru, Elias Geriatric Hospital Bucharest, for their help in recruiting patients.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2020.107351.

References

Abrahams, J.P., Camp, C.J., 1993. Maintenance and generalization of object naming training in anomia associated with degenerative dementia. Clin. Gerontol. 12, 57–72.
Alber, J., 2014. Improving Longer-Term Memory via Wakeful Rest in Health and Amnesia: Evidence for Memory Consolidation. The University of Edinburgh, Edinburgh UK. (Unpublished doctoral dissertation).
American Psychiatric Association, 2000. Diagnostic and Statistical Manual of Mental Disorders. fourth ed. Author, Washington (DC). text rev..
Baddeley, A.D., Atkinson, A., Kemp, S., Allen, R., 2019. The problem of detecting long-term forgetting: evidence from the crimes test and the four doors test. Cortex 1–8.
Baddeley, A.D., Rawlings, B., Hayes, A., 2013. Constrained prose recall and the assessment of long-term forgetting: the case of ageing and the Crimes Test. Memory 22 (8), 1–8.
Becker, J.T., Boller, F., Saxton, J., McGonigle-Gibson, K.L., 1987. Normal rates of forgetting of verbal and non-verbal material in Alzheimer's disease. Cortex 23 (1), 59–72.
Budson, A.E., Simons, J.S., Waring, J.D., Sullivan, A.L., Hussoin, T., Schacter, D.L., 2007. Memory for the september 11, 2001, terrorist attacks one year later in patients with Alzheimer's disease, patients with mild cognitive impairment, and healthy older adults. Cortex 43 (7), 875–888.
Blackburn, H., Benton, A., Shaffer, L.F., 1957. Revised administration and scoring of the Digit span test. J. Consult. Psychol. 21 (2), 139–143.
Brush, J.A., Camp, C.J., 1998. Using spaced retrieval as an intervention during speech-language therapy. Clin. Gerontol. 19 (1), 51–64.
Camp, C.J., Stevens, A.B., 1990. Spaced-retrieval: a memory intervention for dementia of the Alzheimer's type. Clin. Gerontol. 10 (1), 58–61.
Camp, C.J., Foss, J.W., Stevens, A.B., O'Hanlon, A.M., 1996. Improving prospective memory task performance in persons with Alzheimer's disease. In: Brandimonte, M., Einstein, G.O., McDaniel, M.A. (Eds.), Prospective Memory: Theory and Applications. Lawrence Erlbaum, Mahwah, NJ, pp. 351–367.
Camp, C.J., 2001. From efficacy to effectiveness to diffusion: making the transitions in dementia intervention research. Neuropsychol. Rehabil. 11, 495–517.
Carpenter, S.K., Pashler, H., 2007. Testing beyond words: using tests to enhance visuospatial map learning. Psychon. Bull. Rev. 14, 474–478.
Carlesimo, G.A., Sabbadini, M., Fadda, L., Caltagirone, C., 1995. Forgetting from long-term memory in dementia and pure amnesia: role of task, delay of assessment and aetiology of cerebral damage. Cortex 31 (2), 285–300.
Cassel, A., Morris, R., Koutroumanidis, M., Kopelman, M., 2016. Forgetting in temporal lobe epilepsy: when does it become accelerated? Cortex 78, 70–84.
Cerciello, M., et al., 2017. Assessment of free and cued recall in Alzheimer's disease and vascular and frontotemporal dementia with 24-item Grober and Buschke test. Neurol. Sci. 38 (1), 115–122.
Chan, J.C.K., 2009. When does retrieval induce forgetting and when does it induce facilitation? Implications for retrieval inhibition, testing effect, and text processing. J. Mem. Lang. 61 (2), 153–170.
Chan, J.C.K., Erdman, M.R., Davis, S.D., 2015. Retrieval induces forgetting, but only when nontested items compete for retrieval: implication for interference, inhibition, and context reinstatement. J. Exp. Psychol. Learn. Mem. Cognit. 41 (5), 1298–1315.
Christensen, H., Kopelman, M.D., Stanhope, N., Lorentz, L., Owen, P., 1998. Rates of forgetting in Alzheimer dementia. Neuropsychologia 36 (6), 547–557.
Clare, L., Wilson, B.A., Carter, G., Roth, I., Hodges, J.R., 2002. Relearning face-name associations in early Alzheimer's disease. Neuropsychology 16 (4), 538–547.
Clare, L., Woods, R.T., 2004. Cognitive training and cognitive rehabilitation for people with early-stage Alzheimer's disease: a review. Neuropsychol. Rehabil. 14 (4), 385–401.
Creighton, A., Van der Ploeg, E., O'Connor, D., 2013. A literature review of spaced-retrieval interventions: a direct memory intervention for people with dementia. Int. Psychogeriatr. 25 (11).
Cull, W.L., Shaughnessy, J.J., Zechmeister, E.B., 1996. Expanding the understanding of the expanding-pattern-of-retrieval mnemonic: toward confidence in applicability. J. Exp. Psychol. 2, 365–378.
Davidson, M., Dorris, L., O'Regan, M., Zuberi, S.M., 2007. Memory consolidation and accelerated forgetting in children with idiopathic generalized epilepsy. Epilepsy Behav. 11 (3), 394–400.
Degenszajn, J., Caramelli, P., Caixeta, L., Nitrini, R., 2001. Encoding process in delayed recall impairment and rate of forgetting in Alzheimer's disease. Arq. Neuro. Psiquiatr. 59 (2A), 171–174.
Della Sala, S., Kozlova, I., Stamate, A., Parra, M., 2018. A transcultural cognitive marker of Alzheimer's Disease. Int. J. Geriatr. Psychiatr. 33 (6), 849–856.
Duff, K., Beglinger, L.J., Van Der Heiden, S., Moser, D.J., Arndt, S., Schultz, S.K., Paulsen, J.S., 2008. Short-term practice effects in amnestic mild cognitive impairment: implications for diagnosis and treatment. Int. Psychogeriatr. 20 (5), 986–999.
Elliot, G., Isaac, C.L., Muhlert, N., 2014. Measuring forgetting: a critical review of accelerated long-term forgetting studies. Cortex 54, 16–32.
Freed, D.M., Corkin, S., 1988. Rate of forgetting in H.M.: 6-month recognition. Behav. Neurosci. 102 (6), 823–827.
Ferrer, E., Salthouse, T., Stewart, W., Schwartz, B., 2004. Modeling age and retest processes in longitudinal studies of cognitive abilities. Psychol. Aging 19 (2), 243–259.
Garin, O., 2014. Ceiling effect. In: Michalos, A.C. (Ed.), Encyclopedia of Quality of Life and Well-Being Research. Springer, Dordrecht.
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. Front. Psychol. 6 (752), 752.
Goldberg, T.E., Harvey, P.D., Wesnes, K.A., Snyder, P.J., Schneider, L.S., 2015. Practice effects due to serial cognitive assessment: implications for preclinical Alzheimer's disease randomized controlled trials. Alzheimer's Dementia 1 (1), 103–111.
Green, R.E., Kopelman, M.D., 2002. Contribution of recollection and familiarity judgements to rate of forgetting in organic amnesia. Cortex 38 (2), 161–178.
Greene, J.D.W., Baddeley, A.D., Hodges, J.R., 1996. Analysis of the episodic memory deficit in early Alzheimer's Disease: evidence from the doors and people test. Neuropsychologia 34, 537–551.
Grober, E., Kawas, C., 1997. Learning and retention in preclinical and early Alzheimer's disease. Psychol. Aging 12 (1), 183–188.
Hart, R.P., Kwentus, J.A., Taylor, J.R., Harkins, S.W., 1987. Rate of forgetting in dementia and depression. J. Consult. Clin. Psychol. 55 (1), 101.
Hart, R.P., Kwentus, J.A., Harkins, S.W., Taylor, J.R., 1988. Rate of forgetting in mild Alzheimer's-type dementia. Brain Cognit. 7 (1), 31–38.
Isaac, C.L., Mayes, A.R., 1999. Rate of forgetting in amnesia: I. Recall and recognition of prose. J. Exp. Psychol. Learn. Mem. Cognit. 25 (4), 942–962.
Isaac, C.L., Mayes, A.R., 1999. Rate of forgetting in amnesia: II. Recall and recognition of word lists at different levels of organization. J. Exp. Psychol. Learn. Mem. Cognit. 25 (4), 963–977.
Jansari, A.S., Davis, K., McGibbon, T., Firminger, S., Kapur, N., 2010. When "long-term memory" no longer means "forever": analysis of accelerated long-term forgetting in a patient with temporal lobe epilepsy. Neuropsychologia 48 (6), 1707–1715.
Katz, S., 1983. Assessing self-maintenance: activities of daily living, mobility and instrumental activities of daily living. J. Am. Geriatr. Soc. 31 (12), 721–726.
Kinsella, G.J., Ong, B., Storey, E., Wallace, J., Hester, R., 2007. Elaborated spaced-retrieval and prospective memory in mild Alzheimer's disease. Neuropsychol. Rehabil. 17 (6), 688–706.
Kopelman, M.D., 1985. Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. Neuropsychologia 23, 623–638.
Kopelman, M.D., Stanhope, N., 1997. Rates of forgetting in organic amnesia following temporal lobe, diencephalic, or frontal lobe lesions. Neuropsychology 11 (3), 343–356.
Larrabee, G., Youngjohn, J., Sudilovsky, A., Crook, T., 1993. Accelerated forgetting in Alzheimer-type dementia. J. Clin. Exp. Neuropsychol. 15 (5), 701–712.
Lawton, M.P., Brody, E.M., 1969. Assessment of older people: self-maintaining and instrumental activities of daily living. Gerontol. 9, 179–186.
Loftus, G.R., 1985. Evaluating forgetting curves. J. Exp. Psychol. 11, 397–406.
Lombardi, M., Perri, R., Fadda, L., Caltagirone, C., Carlesimo, G., 2018. Forgetting of the recollection and familiarity components of recognition in patients with amnestic mild cognitive impairment. J. Neuropsychol. 12 (2), 231–247.
Lucchelli, F., Spinnler, H., 1998. Ephemeral new traces and evaporated remote engrams: a form of neocortical temporal lobe amnesia? A preliminary case report. Neurocase 4 (6), 447–459.

Manes, F., Graham, K.S., Zeman, A., de Lujan Calcagno, M., Hodges, J.R., 2005. Autobiographical amnesia and accelerated forgetting in transient epileptic amnesia. J. Neurol. Neurosurg. Psychiatry 76 (10), 1387–1391.

Manes, F., Serrano, C., Calcagno, M.L., Cardozo, J., Hodges, J., 2008. Accelerated forgetting in subjects with memory complaints. A new form of mild cognitive impairment? J. Neurol. 255, 1067–1070.

Mayes, A.R., 1986. Learning and memory disorders and their assessment. Neuropsychologia 24, 25–39.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA work group under the auspices of department of health and human services task force on Alzheimer's disease. Neurology 34 (7), 939–944.

Moss, M.B., Albert, M.S., Butters, N., Payne, M., 1986. Differential patterns of memory loss among patients with Alzheimer's disease, Huntington's disease and alcoholic Korsakoff s syndrome. Arch. Neurol. 43, 239–246.

Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al., 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J. Am. Geriatr. Soc. 53 (4), 695–699.

Pilotti, M., Chodorow, M., Petrov, R., 2009. The usefulness of retrieval practice and review- only practice for answering conceptually related test questions. J. Gen. Psychol. 136, 179–203.

Roediger, H.L., Butler, A.C., 2011. The critical role of retrieval practice in long-term retention. Trends Cognit. Sci. 15, 20–27.

Shulman, K.I., 2000. Clock-drawing: is it the ideal cognitive screening test? Int. J. Geriatr. Psychiatr. 15, 548–561.

Thomas, R.C., Weywadt, C.R., Anderson, J.L., Martinez-Papponi, B., McDaniel, M.A., 2018. Testing encourages transfer between factual and application questions in an online learning environment. J. Appl. Res. Memory Cognit. 7, 252–260.

Vallet, G., Rouleau, I., Benoit, S., Langlois, R., Barbeau, E., Joubert, S., 2016. Alzheimer's disease and memory strength: gradual decline of memory traces as a function of their strength. J. Clin. Exp. Neuropsychol. 38 (6), 648–660.

Walsh, C.M., Wilkins, S., Bettcher, B.M., Butler, C.R., Miller, B.L., Kramer, J.H., 2014. Memory consolidation in aging and MCI after 1 week. Neuropsychology 28, 273–280.

Weston, P.S.J., Nicholas, J.M., Henley, S.M.D., Liang, Y., Macpherson, K., et al., 2018. Accelerated long-term forgetting in presymptomatic autosomal dominant Alzheimer's disease: a cross-sectional study. Lancet Neurol. 17 (2), 123–132.

Yan, J.H., Dick, M.B., 2006. Practice effects on motor control in healthy seniors and patients with mild cognitive impairment and Alzheimer's disease. Neuropsychol. Dev. Cognition. Section B, Aging, Neuropsychol. cognition 13 (3–4), 385–410.

Yesavage, J.A., Brink, T.L., Rose, T.L., Lum, O., Huang, V., Adey, M.B., Leirer, V.O., 1983. Development and validation of a geriatric depression screening scale: a preliminary report. J. Psychiatr. Res. 17, 37–49.