This version of the manuscript has been submitted for publication on August 26, 2022

Drawing as an efficient encoding tool in younger but not always older adults: The case of associative memory

Rebecca Ovalle-Fresa and Corinna S. Martarelli

Faculty of Psychology, UniDistance Suisse, Brig, Switzerland

Author Note

Correspondence concerning this article should be addressed to Rebecca Ovalle-Fresa, Schinerstrasse 18, 3900 Brig, Switzerland. Email: rebecca.ovallefresa@gmail.com

Abstract

Episodic memory strongly declines in healthy aging, at least partly because of reduced abilities to create and remember associations (associative memory) and to use efficient memory strategies. Several studies have shown that drawing is a reliable encoding tool to enhance memory of individual items (item memory) because of the simultaneous integration of elaborative, pictorial, and motoric processes. These processes in isolation can enhance associative memory in older adults. Nevertheless, their impact on associative memory has never been investigated as simultaneous processes in drawing as an encoding tool. We thus aimed to investigate whether drawing as an encoding tool not only enhances item memory, but whether its benefit could also be extended to associative memory in younger and especially older adults. Therefore, we tested 70 older and 67 younger participants in two online experiments. All participants took part in a memory task involving unrelated word pairs. We compared drawing and writing as encoding tools and assessed immediate recognition memory of items and associations. In Experiment 2, we additionally assessed their recognition memory after one week. While younger participants benefited from drawing over writing in terms of both item and associative memory, older participants benefited in item but not in associative memory. All the beneficial effects of drawing remained after one week. Thus, while we could extend the benefit of drawing to associative memory in younger adults, the lack of benefit in older adults' associative memory might be explained by the existing age-related difficulties in benefiting from memory strategies and creating associations.

Keywords: Episodic memory, associative memory, aging, drawing, strategy.

Drawing as an efficient encoding tool in younger but not always older adults: The case of associative memory

Cognitive decline negatively influences the wellbeing and everyday functioning of elderly people (Wilson et al., 2013). A central aspect of cognition in healthy aging is a pronounced decline in episodic memory (Grady & Craik, 2000), which refers to the ability to remember past events in the context of time and space (Tulving, 1972). According to the associative deficit hypothesis, the age-related decline in episodic memory can essentially be explained by difficulties in building and retrieving associations (i.e., associative memory) rather than by difficulties in encoding and retrieving individual items (i.e., item memory; Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). Previous studies have revealed that drawing during encoding enhances subsequent memory performance on individual items (for a review, see Fernandes, Wammes, & Meade, 2018), with similar results being demonstrated in elderly people (Meade et al., 2018, 2019). In the present study, we aimed to extend these findings to associative memory and thus investigated, whether drawing as an encoding tool promotes immediate and delayed associative memory performance on word pairs in younger and older adults.

Several strategies have been shown to enhance memory performance. For example, *elaborative processing* during encoding has been found to benefit subsequent memory. More elaborative processes are supposed to lead to richer memory representations and are thus more strongly embedded in preexisting semantic networks when compared to less elaborative processes (Craik & Watkins, 1973; Greene, 1987). This idea has also been described as the levels of processing account (LOP), which predicts a memory advantage for high elaborative (i.e., including semantic meaning) compared to low elaborative (i.e., merely perceptual) encoding processes (Craik & Lockhart, 1972). For instance, in a study investigating associative memory in older and younger adults, word pairs were learned by combining the two words in a meaningful sentence (high elaborative) or by counting syllables (low elaborative; O'Hanlon et al., 2001). Both age groups remembered more word pairs after high rather than low elaboration, indicating that high elaborative encoding processes were beneficial for older and younger adults alike.

The association of pictorials with verbal material can also be used as a strategy for improving subsequent memory. This strategy is described in the *dual coding theory,* which predicts enhanced memory for stimuli simultaneously encoded in both pictorial and verbal form, as compared to stimuli only encoded in verbal form (Paivio, 1969). Studies investigating dual coding in the associative memory of elderly people often use mental imagery (i.e., the creation of a picture in the mind's eye) to bind the elements that are to be associated during encoding. The task of creating one picture in the mind's eye that relates both words of a pair (as opposed to creating two individual pictures for two words) is often referred to as relational imagery. Older adults have been shown to benefit from relational imagery when compared to verbal strategies, such as mere repeating, to a similar extent as younger adults (Bender & Raz, 2012; Dunlosky et al., 2005; Fox et al., 2016; Hertzog et al., 2013; Naveh-Benjamin et al., 2007).

Another beneficial effect on memory was observed after the use of *motoric actions* associated with to-be-encoded stimuli. Examples of such actions are stimuli that are gestured by participants (i.e., enactment effect; Engelkamp & Zimmer, 1997) or stimuli that are spoken out by participants (i.e., production effect; MacLeod et al., 2010) during encoding. When it comes to associative memory in elderly people, the enhancing effects of speaking out sentences compared to reading sentences have been found both in older and younger adults alike (e.g., Bäckman & Nilsson, 1984; Feyereisen, 2009; Silva et al., 2015).

According to the *integrated-components model*, the efficacy of drawing as an encoding tool is based on the integrated combination of the processes involved in the three described strategies, namely, *elaborative*, *pictorial*, and *motoric* processes (for a review, see Fernandes et al., 2018). The effect of simultaneously combining the three processes is assumed to be more than additive, as it leads to detailed mental representations of the to-be-remembered material and thus to strong memory traces and enhanced memory performance (compared to encoding strategies that do not integrate all three processes; Wammes et al., 2019).

To sum up, it has been shown that each of the three processes (elaborative, pictorial, and motoric) outlined above facilitates *associative memory* in elderly people - when studied in isolation - and that they have a more than additive effect on *item memory* when simultaneously combined during drawing (Meade et al., 2018). However, despite the disproportionally large age-related deficit in associative memory (e.g., Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008), research assessing whether drawing as an encoding tool can help older adults to build and remember associations is currently lacking. As the associative deficit is partly due to the diminished ability to *spontaneously* use appropriate strategies (Craik, 1983; Shing et al., 2010), elderly people could especially benefit from drawing as an encoding tool that, by its nature, automatically integrates the use of elaborative, pictorial, and motoric processes.

Further insights into the efficacy of a particular encoding strategy can be gained by investigating the evolution of memories over time. This is because successful memory consolidation mainly depends on how the information has been encoded (Craik, 1983). Memory consolidation refers to processes that gradually stabilize a memory trace after its initial acquisition, resulting in resistance against interferences (e.g., Dudai, 2004). Here, we are interested in *system memory consolidation*, which starts approximately one week after the initial memory acquisition (e.g., Dudai, 2004). Previous research comparing the consolidation of associative memory and item memory found that item memory benefits to a larger degree from consolidation processes than associative memory when younger adults were tested after synaptic consolidation in short naps (e.g., Sopp et al., 2018; Studte et al., 2015). Another study reported less efficient system memory consolidation for associative memory in older adults than in younger adults (e.g., Mary et al., 2013). Taken together, these findings indicate weaker consolidation (a) of associative memory than item memory and (b) in older adults compared to younger adults when no effective encoding strategy is used.

We addressed the following three goals in two experiments: (a) replicating the existing findings of drawing as an encoding tool for individual items in older and younger adults, (b) extending these findings to associative memory in both older and younger adults, and (c) investigating the impact of memory consolidation on the potential benefit of drawing after a delay of one week. Because of the COVID-19 pandemic, we conducted both experiments online with the accompaniment of an experimenter by phone. In both experiments, we tested older and younger adults in a memory task adapted from Kamp and Zimmer (2015) to obtain recognition memory scores for items and associations. We used two versions of the memory task, with both versions consisting of an encoding and a recognition phase. Participants completed both task versions both using a different tool to encode word pairs: either drawing one image that integrated both words of a pair or repeatedly writing the two words of a pair. During recognition, participants were presented with word pairs again and had to indicate whether they were old (word pair presented during encoding), recombined (both words presented during encoding but in different pairs), or new (both words never presented during encoding). In Experiment 1, the recognition phase was conducted immediately after encoding. In Experiment 2, we also conducted a delayed recognition phase after one week. To assess potential age differences regarding pictorial processing, we further assessed preferences for visual and verbal processing, vividness of visual mental imagery, and drawing habits with three self-report questionnaires at the end of each experiment.

Based on the consistent benefit of drawing as an encoding tool for item memory - also in older adults (Meade et al., 2019, 2018) - we expected better item memory performance in both age groups after drawing compared to writing. Moreover, as previous studies investigating strategies that involve *elaborative*, *pictorial*, and *motoric* processes in isolation have revealed benefits for associative memory, we expected that drawing would generally enhance associative memory because it integrates these processes into one action. We thus hypothesized that drawing would provide the environmental support needed by older adults to successfully encode associations, and therefore expected the largest benefit of drawing in older adults' associative memory.

Concerning the delayed recognition phase, we further hypothesized that drawing as an encoding tool would lead to stronger consolidated memory traces compared to writing as an encoding tool. Therefore, we expected that the benefits after drawing would remain over an interval of one week. If drawing also provides environmental support to apply imagery strategies for older adults, then we anticipated that it might specifically diminish the age-related associative memory deficit after one week.

Experiment 1

Methods

Design

We conducted the experiment online with the accompaniment of an experimenter by phone. Participants completed two versions of a memory task, each consisting of an encoding and a recognition phase. In one version, we used drawing as an encoding tool (drawing one image that integrated both words of a pair); in the other version, the encoding tool was writing (repeatedly writing the two words of a pair). During the recognition phases, the presented word pairs could be old (word pair presented during encoding), recombined (both words of the pair presented during encoding but in different pairs), or new (both words not presented during encoding). This paradigm allowed us to calculate recognition memory performance (Pr) individually for item and associative memory (cf. Kamp & Zimmer, 2015). Thus, Experiment 1 followed a 2 x 2 x 2 mixed design, with the between-participant factor of *Age Group* (older, younger), and the within-participant factors of *Encoding Tool* (drawing, writing) and *Memory Type* (associative, item).

Apparatus and Materials

Memory Task. We programmed and presented the memory task in lab.js, an open-source tool based on Java script for creating and running online or lab-based experiments **(Henninger et al., 2022)**. The task consisted of an encoding phase, a short retention phase, and a recognition phase. We used the same task in two versions with different word-pair lists: one version for drawing and one for writing as an encoding tool.

Stimuli consisted of German nouns, represented both in the Berlin Affective Word List-Reloaded (BAWL-R; Võ et al., 2009) and in the Leipzig corpora (Goldhahn et al., 2012). We preselected a pool of 1531 words with an imageability value larger than 3 (BAWL-R, ranging from 1: "low imageability" to 7: "high imageability") and an emotional value larger than -2 (BAWL-R, ranging from -3: "very negative," through 0: "neutral," to 3: "very positive"). From this pool, we randomly selected 312 words. They had a mean imageability value of 6.06 (SD = 0.36), a mean emotional value of 0.37 (SD = 1.05), and a mean word frequency of 13.23 (SD = 2.45)[1]. From the 312 words, we randomly created six word pair lists (A to F) of 26 pairs to have one list per association type (old, recombined, and new) and encoding tool (drawing, writing). The word pair lists did not significantly differ from each other in imageability, $F_{\text{imageability}}(5, 306) = 1.14$, p = . 341, emotional value, $F_{\text{emotional}}(5, 306) = 0.60$, p = .698, and frequency class, $F_{\text{frequency}}(5, 306) = 0.78$, p = .565. We also created an example list with six word pairs, consisting of words only used for the examples but not otherwise in the task.

Cognitive Screening. To screen for potential mild cognitive impairment, we applied the German version of the Montreal Cognitive Assessment for visually impaired

[1] According to the Leipzig corpora word frequency class, the most frequent word, here "und" [the German word for "and"], has a frequency class of zero, and extremely rare words have frequency classes of 20 or higher. "deu-com_web_2018_1M-words", retrieved from https://wortschatz.uni-leipzig.de/en/download

people (MoCA-B; Wittich et al., 2010). The MoCA-B corresponds to the original MoCA (Nasreddine et al., 2005) without visual tasks and is thus applicable by phone. The maximum score for the MoCA-B was 22, with a cut-off value of 17 (Wittich et al., 2010).

Questionnaires. To detect potential age differences in pictorial processing habits and abilities, which may possibly influence the use of drawing as an encoding tool, we assessed the preference for pictorial and verbal processing, the ability to create vivid mental imagery, and drawing habits in self-report questionnaires. All questionnaires were programmed and presented in Limesurvey (Limesurvey GmbH, 2018) and are described in more detail in the following paragraphs.

Verbalizer-Visualizer Questionnaire (VVQ2). We assessed individual preferences for different cognitive processing styles with the VVQ2 (Kirby et al., 1988), a revision of the original VVQ (Richardson, 1977) in its German form (Meier & Rothen, 2013). This questionnaire is used to investigate cognitive preferences for verbal processes (10 items, e.g., “I enjoy doing work that requires the use of words”), visual processes (10 items, e.g., “When I read books with maps in them, I refer to the maps a lot”), and imagery processes (10 items, e.g., “My dreams are sometimes so vivid I feel as though I actually experience the scene”). The 30 statements all consist of five-point Likert scales ranging from 1 (“I strongly disagree”) to 5 (“I strongly agree”). Cronbach’s alpha was .71 for the verbal scale, .60 for the visual scale, and .84 for the imagery scale.

Vividness of Visual Imagery Questionnaire (VVIQ). We further assessed the ability to create mental images of scenes with the VVIQ (Marks, 1973), using a German translation of the adapted version described by Cui et al. (2007). The questionnaire consists of 16 scenes, with the instruction to visualize the scene and rate the resulting mental image on a scale ranging from 1 (“no image”) to 5 (“a perfect image, as clear as in vision”). The same scenes were rated once with eyes open and once with eyes closed. An example description of a scene is “Visualize the rising sun. Carefully consider the picture that comes before your mind’s eye. The sun is rising above the horizon into a hazy sky.” Cronbach’s alpha was .95.

Drawing Questionnaire. We further created 20 statements to assess the usage of drawing in everyday life. Each statement was rated on a five-point Likert scale ranging from 1 (“I strongly disagree”) to 5 (“I strongly agree”). The statements addressed how often and how well someone draws, and how much someone likes to draw (e.g., “This week I drew something on a piece of paper at least once,” or “I draw better than most people of my age”). Cronbach’s alpha for the drawing score was .89. The drawing questionnaire (in German) is provided on the open science platform (OSF; https://osf.io/3y8eb/).

Procedure

Due to the pandemic situation when we started the data acquisition (spring 2020), we decided to conduct the experiment in individual online sessions with phone support (i.e., an experimenter accompanied the participant by phone). Before each session, participants received an email with brief information and individualized links to the tasks. On the test date, the experimenter contacted the participant by phone. At the beginning of the session, the participants gave informed consent, and demographic data were assessed. Then, the experimenter guided the participants through the questions of the MoCA-B (Wittich et al., 2010).

Next, the participants completed the memory task in both encoding tool versions (drawing and writing), which were presented in counterbalanced order. In both versions, the memory task started with a *short example*, during which participants were still accompanied by phone. The examples consisted of an encoding, a retention, and a recognition phase, each phase starting with an instruction screen. The three phases were equal, as described for the proper memory task in the next paragraphs, except for the number of trials. In the encoding example, we presented two old trials (i.e., the word pairs were presented again during recognition) and two recombined trials (i.e., the word pairs were presented again during recognition but in different combinations). In the recognition example, we then presented two old trials, two recombined trials, and two new trials (i.e., not presented during encoding). All trials were presented in an intermixed and random order. To ensure that the participants understood the task correctly, they were asked to provide each instruction in their own words after reading. After they completed the example task, the participants could ask clarifying questions before starting the proper memory task.

The *encoding phase* of the proper memory task started with the same instruction page, as shown in the example task, and the participants were allowed to read it again. The instruction was to memorize the presented word pairs by either drawing a relational picture (drawing version) or by repeatedly writing the words of a pair (writing version). Participants used pencils on paper for both encoding tools. Phone calls ended after the instructions for the proper task and started again before beginning a new phase. Then, 52 encoding trials (which served as 26 old and 26 recombined trials in the subsequent recognition phase) were presented in an intermixed and random order. Each encoding trial started with a centered fixation cross for 3 seconds (s), followed by the word pair presented for 1 s. Next, the prompt for the encoding tool, either drawing or writing, appeared on the screen. The encoding duration was 45 s for older and 25 s for younger participants. The difference in encoding times was chosen because older adults need more processing time for encoding strategies (cf. Naveh-Benjamin et al., 2007; Salthouse, 1995, and our pilot experiment in the Supplemental Material). During the last 500 ms of the encoding duration, a tone indicated the beginning of the next trial. See Figure 1A for examples of trials in the encoding phase.

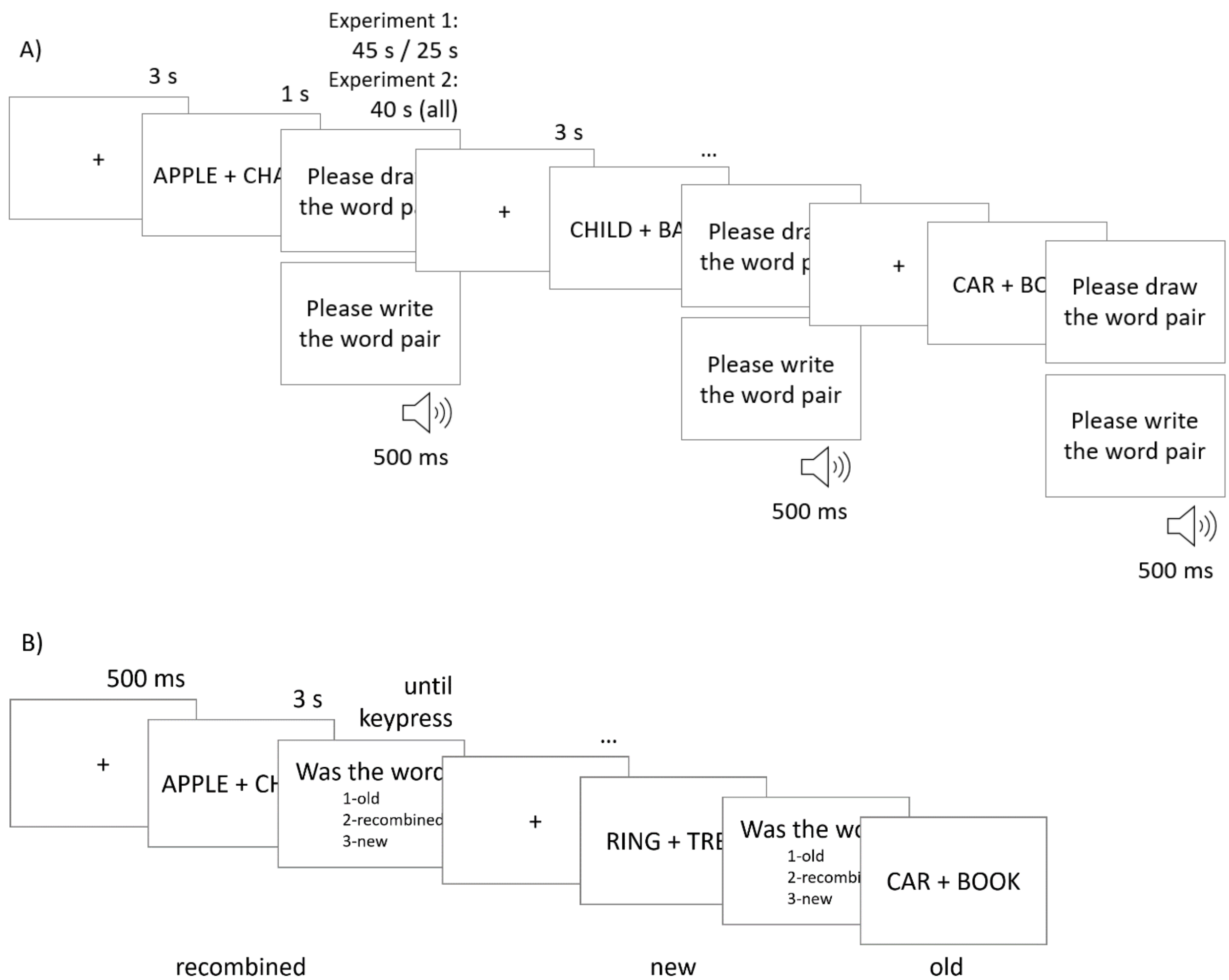

Figure 1. Example trials for the encoding (A) and recognition phases (B) of the memory task used in Experiments 1 and 2. The encoding time was 45 s for older adults and 25 s for younger adults in Experiment 1, and 40 s for all participants in Experiment 2. For the last 500 ms of the encoding duration, a tone indicated the start of the next trial.

During the *retention phase* (approximately 1 minute), participants solved simple arithmetic problems as distractor task. Five simple equations were subsequently presented on the screen, and participants had to indicate by keypress whether an equation was true (i.e., 2 * 4 = 8) or false (i.e., 7 + 4 = 10). The key "1" was used for true equations, and the key "2" for false equations. After the response to the last equation, the recognition phase was initiated.

The *recognition phase* started with an instruction screen. Participants were instructed to indicate if a word pair was old (i.e., word pair presented during encoding), recombined (i.e., words presented during encoding but in different combinations), or new (i.e., both words not presented during encoding). For each association type (old, recombined, and new), 26 word pairs were presented in intermixed and randomized order, resulting in 78 recognition trials. A recognition trial started with a centered fixation cross for 500 ms, followed by a word pair for 3 s. Next, the screen with the recognition question and the corresponding response keys appeared: "Was the word pair 1: old, 2: recombined, 3: new?" The keypress (self-paced) initiated the next trial. See Figure 1B for examples of trials in the recognition phase.

After the last trial of the recognition phase of the first memory task version, the second memory task version (using the second encoding tool) started with the instruction screen for the encoding phase. At this point, the participants had the opportunity to take a short break. The order of the task version was counterbalanced between participants, such that half of the participants started with drawing and the other half with writing as an encoding tool. Furthermore, the six word-pair lists for the association types (old, recombined, and new) per condition (drawing, writing) were counterbalanced between participants. That is, half of the participants were presented with lists A, B, and C for the drawing version and lists D, E, and F for the writing version. The other half of the participants were presented with lists D, E, and F during drawing and A, B, and C during writing. The combination of lists with association type was counterbalanced between participants by rotating the three lists per condition between the three association types.

After completing the memory task in both encoding tool versions, the participants reported the strategies they used during the memory task and whether they looked at their notes and drawings during the test phase. Participants with "yes" responses to the latter question were excluded from the analyses. At the end of the session, participants responded to the drawing questionnaire, then to the VVQ2 (Kirby et al., 1988; Meier & Rothen, 2013; Richardson, 1977), and finally to the VVIQ (Cui et al., 2007; Marks, 1973). Afterwards, participants could ask questions about the project by phone, and we sent them an email with detailed study information.

Analytical Approach

We conducted all analyses in R-Studio 4.0.3 (R Core Team, 2020); the scripts are provided on OSF (https://osf.io/3y8eb/). The alpha level was set to .05. As measures of effect size, we report partial eta squared (η_p^2) for the analyses of variance (ANOVAs) and Cohen's *d* (*d*) for the *t*-tests. All tests were two-tailed, and we report Holm-corrected *p*-values for the *t*-tests.

Memory Task. We aimed to compare recognition memory for items and associations. To this end, we followed the approach used in Kamp and Zimmer (2015). The rationale behind this approach is that item memory requires the recognition of a word, regardless of the second word it was presented with. In contrast, associative memory requires the recognition of the word in the association with which it was presented. Accordingly, we calculated the recognition memory performance scores as proportions of hits minus false alarms (i.e., Pr score). For item memory, we classified "old" and "recombined" responses to old and recombined word pairs as hits. "Old" and "recombined" responses to new word pairs were classified as false alarms. For associative memory, we classified "recombined" responses to recombined pairs as hits and "recombined" responses to old pairs as false alarms. It was also possible to classify "old" responses to old pairs as hits and "old" responses to recombined pairs as false alarms to calculate the Pr score for associative memory. The pattern of results led to the same conclusions in both versions. For the purpose of readability, we report the results for the first and second versions in the Supplemental Material.

Questionnaires. We created the following questionnaire scales per participant: for the VVQ2, we built the three scales reflecting cognitive preferences for verbal, visual, and imagery processes as sum scores ranging from -20 to 20 (negative items reversed as 1 to -1, 2 to -2, … 5 to -5). Higher scores indicated a stronger preference for the representation format and cognitive style represented by the scale (Meier & Rothen, 2013). Next, we calculated the VVIQ score as an average across all ratings per participant, with higher values pointing to a better ability to create vivid mental images (Cui et al., 2007; Marks, 1973). For the drawing questionnaire, we created an average score across all items (negative items reversed as 1 to 5, 2 to 4, 4 to 2, and 5 to 1). Higher values represented higher enjoyment and engagement in drawing.

Participants

In the pre-registration for Experiment 1 (https://osf.io/3y8eb/) we reported the a priori power analysis and the procedure for exclusions. Based on the interaction effects reported by Kamp and Zimmer (2015), *Age Group* x *Memory Type*, $f = 0.54$, and by Meade et al. (2018), *Age Group* x *Encoding Tool*, $f = 0.38$, we conservatively assumed a medium effect size, $f = 0.25$, for our power analysis. We further considered an error level $\alpha = 0.05$ and power $1-\beta = 0.95$. The a priori power analysis for an ANOVA with repeated measures and within-between interactions was computed in G*power (Faul et al., 2007) and estimated a required sample of 54 participants. Thus, we aimed for complete datasets of 27 older and 27 younger participants after eventual exclusion.

The exclusion criteria were (1) incomplete datasets, (2) cognitive impairment indicated by a score below 17 in the MoCA-B (Wittich et al., 2010), (3) use of the encoding notes or drawings during recognition, and (4) mean memory performance in any of the conditions that deviated more than 2.5 *SD* from the condition mean (cf. https://osf.io/3y8eb/). Anticipating potential exclusions, we tested 37 older and 36 younger participants. All participants scored 18 or higher on the MoCA-B, indicating no cognitive impairment (Wittich et al., 2010). Five older and four younger participants were excluded because of incomplete datasets (due to technical errors and dropouts) or not following the instructions. Additionally, three younger participants were excluded due to their self-reported use of encoding notes or drawings during recognition. This resulted in a dataset of 29 older and 32 younger participants before the outlier exclusions. The outlier analysis revealed memory performance (Pr scores) of more than 2.5 *SD* below the condition mean in at least one condition in three younger participants. They were removed from the dataset before analysis.

The remaining sample is described in Table 1. It consisted of 29 older participants with a mean age of 70.3 years (*SD* = 4.4, min = 65.1, max = 79.7) and 29 younger participants with a mean age of 27.1 years (*SD* = 5.2, min = 18.7, max = 39.5). The median educational level according to the International Standard Classification of

Education 2011 (ISCED; UNESCO Institute for Statistics, 2012), was 4 in both age groups. A value of 4 reflects a “post-secondary non-tertiary education,” and a value of 5 reflects a “short-cycle tertiary education.” The age groups did not significantly differ in their questionnaire scores, reflecting a preference for verbal processes (Kirby et al., 1988; Richardson, 1977), the ability to create vivid mental images (Cui et al., 2007; Marks, 1973), and general drawing habits (cf. Table 1. for the exact values). However, older participants, compared to younger participants, showed a significantly lower preference for visual processes, $t(56) = 2.27$, $p = .027$, $d = 0.60$, and for imagery processes, $t(56) = 4.77$, $p < .001$, $d = 1.25$. These are the two scales of the VVQ2 related to visual processing (Kirby et al., 1988; Richardson, 1977).

All participants were fluent in German and recruited online via social media, the participant pool of UniDistance Suisse (including students and non-students), and the acquaintances of the experimenters. They received either 25 Swiss Francs or course credits for participation. The study was approved by the local ethics committee of UniDistance Suisse. Participants were informed that they could withdraw at any time during the experiment before they consented to participate.

Table 1. Demographic variables, MoCA-B, and questionnaire scores of the final sample in Experiment 1.

	Older	Younger		
Parameter	*M (SD)*	*M (SD)*	statistics	*p*
n	29	29		
Age	70.30 (4.42)	27.10 (5.15)		
Gender distribution m/f	16 / 13	9 / 20	2.53	.112
Educational level	4 (7)	4 (5)	392	.658
MoCA-B	20.80 (1.30)	21.40 (0.83)	302	.046
Verbal preferences (VVQ2)	2.07 (4.36)	3.72 (7.24)	1.05	.296
Visual preferences (VVQ2)	3.55 (5.22)	6.62 (5.08)	2.27	.027
Imagery preferences (VVQ2)	0.86 (5.44)	8.69 (6.97)	4.77	< .001
Vividness of mental imagery (VVIQ)	3.79 (0.48)	3.87 (0.65)	0.49	.626
Drawing questionnaire	2.59 (0.71)	2.72 (0.82)	0.62	.541

Notes: We used the MoCA-B *(Wittich et al., 2010)* to screen for cognitive impairment, the VVQ2 to assess preferences in cognitive processing *(Kirby et al., 1988; Richardson, 1977)*, the VVIQ to assess vividness of visual imagery *(Cui et al., 2007; Marks, 1973)*, and our own drawing questionnaire to assess drawing habits. Means are reported with standard deviations in parentheses for age, MoCA-B, and questionnaire scores. The median with range is reported for educational level. The statistics used to compare the two age groups refer to Wilcoxon tests for independent samples (educational level and MoCA-B score), a χ^2 test (gender distribution), and *t*-tests for independent samples for the questionnaire scores. As revealed by the *p*-values, the age groups significantly differed in the MoCA-B scores and the visual scores of the VVQ2 (i.e., preferences for visual and imagery processes).

Results

We aimed to investigate the impact of drawing as an encoding tool on recognition memory performance, depending on age group and memory type. Descriptive statistics of memory performance represented by Pr scores are depicted in Figure 2. Following our pre-registration plan, we conducted a 2 x 2 x 2 ANOVA for mixed designs with the between participant factor *Age Group* (older, younger) and the within participant factors *Encoding Tool* (drawing, writing) and *Memory Type* (associative, item). The ANOVA revealed significant main effects of *Age Group*, $F(1, 56) = 31.10$, $p < .001$, $\eta_p^2 = .36$, due to lower Pr scores in the older than the younger participants, *Encoding Tool*, $F(1, 56) = 5.845$, $p = .019$, $\eta_p^2 = .10$, due to higher Pr scores after drawing than writing, and *Memory Type*, $F(1, 56) = 23.94$, $p < .001$, $\eta_p^2 = .30$, due to higher Pr scores for items than associations.

Moreover, the interaction *Age Group* x *Encoding Tool* was significant, $F(1, 56) = 4.25$, $p = .044$ $\eta_p^2 = .07$. Pairwise *t*-tests revealed that this was because no significant difference between encoding tools was observed in the older age group, $t(28) = 0.23$, $p = .818$, $d = 0.04$ ($M_{drawing} = 0.57$, $SD_{drawing} = 0.30$, $M_{writing} = 0.56$, $SD_{writing} = 0.29$), while in the younger age group, Pr scores were significantly higher after drawing ($M_{drawing} = 0.92$, $SD_{drawing} = 0.07$) than after writing ($M_{writing} = 0.82$, $SD_{writing} = 0.17$), $t(28) = 3.52$, $p = .001$, $d = 0.65$. These results indicated that across both memory types (i.e., item and associative), younger adults benefited from drawing over writing, while older adults did not.

Also, the triple interaction *Age Group* x *Encoding Tool* x *Memory Type* was significant, $F(1, 56) = 6.39$, $p = .014$, $\eta_p^2 = .10$. Pairwise *t*-tests revealed that younger participants performed significantly better after drawing than writing in item memory ($M_{drawing} = 0.94$, $SD_{drawing} = 0.05$, $M_{writing} = 0.88$, $SD_{writing} = 0.12$), $t(28) = 3.31$, $p = .012$, $d = 0.61$, and in associative memory ($M_{drawing} = 0.90$, $SD_{drawing} = 0.11$, $M_{writing} = 0.76$, $SD_{writing} = 0.23$), $t(28) = 3.31$, $p = .003$, $d = 0.61$. In contrast, older participants showed no difference in memory performance between drawing and writing in item memory ($M_{drawing} = 0.63$, $SD_{drawing} = 0.35$, $M_{writing} = 0.53$, $SD_{writing} = 0.29$), $t(28) = 1.15$, $p = .520$, $d = 0.21$, nor in associative memory ($M_{drawing} = 0.50$, $SD_{drawing} = 0.30$, $M_{writing} = 0.53$, $SD_{writing} = 0.29$), $t(28) = 0.52$, $p = .608$, $d = 0.10$. These results indicated that drawing provided an efficient encoding tool not only for item but also for associative memory. However, it seemed that only the younger participants benefited from drawing over writing in both memory types, while the older ones did not. The other interactions were not significant, with all Fs$(1, 56) \leq 0.226$, $p \geq .636$, $\eta_p^2 < .01$.

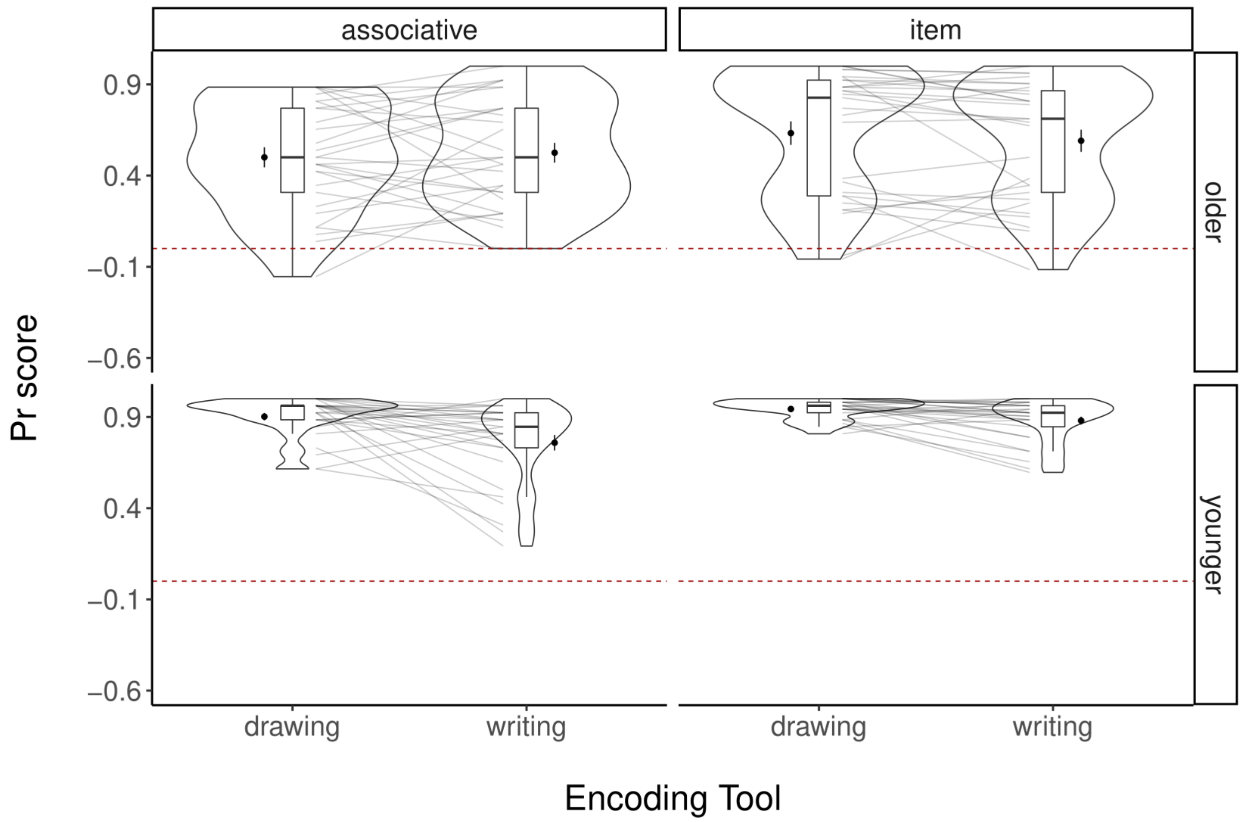

Figure 2. Recognition memory performance (Pr scores) as a function of age group (older, younger), encoding tool (drawing, writing), and memory type (associative, item). The ends of the gray lines represent the averaged Pr scores per participant in one condition. Points and error bars represent condition means and standard errors, respectively. The boxplots entail a horizontal line for the median (i.e., 50th percentile), with the lower and upper ends representing the 25th and 75th percentiles (i.e., entailing the interquartile range), and the whiskers extending 1.5 times past the interquartile range from either end of the box. The violin plots indicate the density of the data. The dashed line represents the level of guessing at Pr scores of zero.

Discussion

In Experiment 1, we confirmed the previous observations that drawing during encoding enhances subsequent item memory in younger adults (e.g., Fernandes et al., 2018, for a review). In younger adults, we could further extend the benefit of drawing over writing to associative memory. However, the beneficial effect of drawing compared to writing was only evident in the younger but not in the older age group. This lack of a clear benefit of drawing in older adults, even in item memory, was surprising, as it conflicted with our expectations, based on the results of previous research (Meade et al., 2018, 2019). A plausible explanation for this result in the older age group might lie in the distribution of older adults' memory performance, which revealed two peaks (cf. the upper right of Figure 2). Therefore, the data might have been too noisy to allow for the detection of statistically significant differences between the encoding tools. Another possible explanation might be the longer encoding times for older participants (45 s) compared to younger participants (25 s), resulting in overall longer encoding phases for older than younger participants. With the longer encoding phases, we intended to compensate for generally longer processing times in older adults (cf. Naveh-Benjamin et al., 2007; Salthouse, 1995). Other than intended, the longer encoding phases might have also led to more forgetting in the older than younger participants. To rule out (or confirm) these explanations, we decided to test our hypotheses in a second experiment with a new sample and equal encoding times for both age groups.

Experiment 2

Experiment 1 revealed that younger adults benefited from drawing in item and associative memory, whereas older adults did not benefit in any of the memory types. As previous research has observed remarkable benefits of drawing in item memory for older adults (Meade et al., 2018, 2019), we aimed to replicate the findings of Experiment 1 in a second experiment. To rule out the possibility that the observed differences in memory performance between older and younger adults in Experiment 1 were based on the difference in encoding durations (i.e., 40 s for older and 25 s for younger participants), we used equal encoding times for both age groups in Experiment 2 (i.e., 40 s).

Moreover, we aimed to investigate the long-term benefits of drawing as an encoding tool. An additional delayed recognition phase after one week allowed us to assess the system-consolidated recognition memory. System consolidation refers to processes that solidify memory traces approximately one week after initial acquisition, resulting in resistance to interference (e.g., Dudai, 2004). Because successful memory consolidation mainly depends on how the information has been encoded (Craik, 1983), drawing as a successful encoding strategy should lead to better memory performance after a delay of one week, when compared to less appropriate strategies.

Methods

The methods were identical to those of Experiment 1. Exceptions are reported below.

Design

In addition to the immediate recognition phase (i.e., immediate recognition), participants were tested in an additional recognition phase after one week (i.e., delayed recognition). The design for the main analyses of Experiment 2 thus consisted of a 2 x 2 x 2 x 2 mixed design with the between-participants factor of *Age Group* (older, younger), and the within-participants factors *Encoding Tool* (drawing, writing) and *Memory Type* (associative, item), and *Session* (immediate, after one week).

Materials

For the recognition memory task, we randomly selected 384 words from the pool of 1531 nouns described in Experiment 1. Average imageability was 5.33 (SD = 0.78), average word frequency was 13.09 (SD = 2.48), and average emotion valence was 0.63 (SD = 0.84) for all 384 words. With these words, we randomly created twelve lists of 16 word pairs each, to have one list for each association type (old, recombined, and new) per encoding tool (drawing, writing) and session (immediate, after one week). The word-pair lists did not significantly differ in imageability, $F_{\text{imageability}}(11, 372) = 0.51$, $p = .90$, emotional valence, $F_{\text{emotion}}(11, 372) = 0.69$, $p = .750$ (both BAWL-R), and frequency class, $F_{\text{frequency}}(11, 372) = 0.93$, $p = .512$ (Leipzig corpora, “deu-com_web_2018_1M-words,” retrieved from https://wortschatz.uni-leipzig.de/en/download; i.e., Goldhahn et al., 2012) and entailed no outlier (i.e., ± 3 standard deviations from the list mean of the three parameters of imageability, emotional valence, and frequency class). Additionally, we created an example list with six word pairs, which consisted of words only used for the examples but not otherwise in the task.

We used the same questionnaires as in Experiment 1. The VVQ2 revealed Cronbach’s alphas of .73 for the verbalizer scale, .73 for the visualizer scale, and .82 for the imagery scale. Cronbach’s alpha for the VVIQ score was .97, and .88 for the drawing questionnaire score.

Procedure

Experiment 2 was conducted in two sessions, separated by one week. The delayed session was held exactly one week after the first session, on the same day of the week and at the same time. We used similar counterbalancing as in Experiment 1 but considered the additional factor *Session*.

The first session included the memory task, with encoding phases and immediate recognition phases. The overall procedure of the first session was equal to that of Experiment 1, and was similarly accompanied by a phone call. At the end of the first session, participants responded to the same questions concerning the memory task as in Experiment 1, such as if they looked at their notes and drawings (created during encoding) during the recognition phases. The delayed session included the delayed recognition phases (no encoding phase), followed by the questionnaires. After ensuring that the participants understood the instructions for the memory task, they immediately started with the delayed recognition phase of the first encoding tool, followed by the delayed recognition phase of the second encoding tool (i.e., drawing and writing). Afterwards, the participants responded to questions concerning the memory task, such as if they looked at their notes and drawings specifically during the delay and at the second session. Then, the participants were instructed by phone on how to complete the drawing questionnaire, the VVQ2 (Kirby et al., 1988; Meier & Rothen, 2013; Richardson, 1977), and the VVIQ (Cui et al., 2007; Marks, 1973).

The procedure of an encoding trial is shown in Figure 1. In contrast to Experiment 1, all participants (older and younger adults) were presented with equal encoding durations of 40 s. Each encoding phase (drawing and writing) consisted of 64 trials.

The recognition phases were presented with the same trial procedure as in Experiment 1 (see Figure 1). Each immediate recognition phase consisted of 48 trials, whereas 16 old, 16 recombined, and 16 new word pairs were presented in random and intermixed order. This means that half of the word pairs presented in the encoding phase were used as old and recombined word pairs during the immediate recognition phase. The other half of the word pairs were presented during the delayed recognition phase, again consisting of 48 trials with 16 old, 16 recombined, and 16 new word pairs. New word pairs were different for the immediate and delayed recall phases.

Participants

The sample size estimation for Experiment 2 was preregistered (https://osf.io/3y8eb/) and conducted using Superpower (https://arcstats.io/shiny/anova-exact/; Lakens & Caldwell, 2021). The a priori power analysis based on the results of Experiment 1 (calculated with the condition means of Experiment 1, a common SD = 0.1945, and an r = 0.6498) revealed that a sample of 78 participants was needed to detect a triple interaction (Age x Encoding Tool x Memory Type) with a power of 0.80 and an alpha set to 0.05. Thus, we aimed for complete datasets of 39 younger (18 – 30 years) and 39 older (65 – 80 years) healthy adults after eventual exclusions.

A total of 104 participants agreed to participate in Experiment 2. None of the participants had already participated in Experiment 1. Two older and one younger participant clicked on the link and decided not to participate before or during the phone call at the beginning of the testing session. We thus tested 101 participants. Technical errors (two older and four younger) and the dropout of an older participant led to six incomplete datasets, which we excluded in the first step. Next, we excluded two older participants because of a MoCA-B score lower than 17 (Wittich et al., 2010), and then seven participants after the self-reported use of drawings or notes during one of the recognition phases (six older and one younger participant). With the remaining data, including 43 older and 42 younger participants, we conducted an outlier analysis. Participants with Pr scores of more than 2.5 *SD* below or above the condition mean in at least one condition were considered outliers and removed from the dataset for the analyses (two older and four younger participants).

The final sample is described in Table 2. It consisted of 41 older participants with a mean age of 72.5 years (*SD* = 6.2, min = 64.8, max = 88.4) and 38 younger participants with a mean age of 29.1 years (*SD* = 4.5, min = 19.8, max = 36.2). We used independent *t*-tests to compare the questionnaire scales between the two age groups. As already observed in Experiment 1, older participants reported a significantly lower preference for visual processes, $t(77) = 2.94$, $p = .004$, $d = 0.66$, and for imagery processes, $t(77) = 3.07$, $p = .003$, $d = 0.69$, when compared to the younger participants. In the other questionnaire scales, we observed no differences between age groups (cf. Table 2).

Table 2. Demographic variables, MoCA-B, and questionnaire scores per age group for Experiment 2.

	Older	Younger		
Parameter	*M (SD)*	*M (SD)*	statistics	*p*
n	41	38		
age	72.5 (6.2)	29.1 (4.5)		
gender distribution m/f	20 / 21	10 / 28	3.33	.068
educational level	4 (6)	6 (4)	650	.198
MoCA-B score	20.34 (1.44)	20.68 (1.28)	674	.284
Verbal preference (VVQ2)	3.05 (7.01)	3.13 (5.24)	0.06	.953
Visual preference (VVQ2)	4.61 (6.31)	8.58 (5.62)	2.94	.004

Imagery preference (VVQ2)	1.15 (7.45)	6.32 (7.51)	3.07	.003
Vividness of mental imagery (VVIQ)	3.84 (0.81)	3.83 (0.53)	0.08	.935
Drawing questionnaire	2.80 (0.78)	2.56 (0.70)	1.39	.169

Notes: Age, MoCA-B, and questionnaire scores are reported with the mean (and the standard deviation in parentheses). Educational level is reported with the median (and the range in parentheses). The statistics refer to Wilcoxon tests for independent samples (educational level and MoCA-B score), a χ^2 test (gender distribution), and *t*-tests for independent samples for the questionnaire scores. Preferences in cognitive processing were assessed with the VVQ2 *(Kirby et al., 1988; Richardson, 1977)*, vividness of mental imagery with the VVIQ *(Cui et al., 2007; Marks, 1973)*, and drawing habits using our own questionnaire.

Results

Descriptive statistics of memory performance represented by Pr scores are depicted in Figure 3. We computed a 2 x 2 x 2 x 2 ANOVA for mixed designs with the between-participant factor *Age Group* (older, younger), and the within-participant factors *Encoding Tool* (drawing, writing), *Memory Type* (associative, item), and *Session* (immediate, after one week). As in Experiment 1, all main effects were significant: *Age Group*, $F(1, 77) = 21.57$, $p < .001$, $\eta_p^2 = .22$, because younger adults generally performed better than those who were older, *Encoding Tool*, $F(1, 77) = 18.38$, $p < .001$, $\eta_p^2 = .19$, due to an overall benefit in the performance of drawing over writing, *Memory Type*, $F(1, 77) = 172.~62$, $p < .001$, $\eta_p^2 = .69$, because of better performance in item than associative memory, and *Session*, $F(1, 77) = 580.68$, $p < .001$, $\eta_p^2 = .88$, because immediate recognition memory was generally better than performance after one week.

As in Experiment 1, the analysis further revealed a significant *Age Group x Encoding Tool x Memory Type* interaction, $F(1, 77) = 4.56$, $p = .036$, $\eta_p^2 = .06$. To resolve this, we investigated the *Encoding Tool x Memory Type* interaction in a separate ANOVA per age group. This revealed a trend toward an interaction for older participants, $F(1, 40) = 3.52$, $p = .068$, $\eta_p^2 = .08$, but not for younger participants, $F(1, 37) = 1.41$, $p = .243$, $\eta_p^2 = .04$, indicating that the drawing benefit differed between memory types in older but not in younger adults. Pairwise *t*-tests for dependent measures showed that older participants performed better after drawing compared to writing in item memory ($M_{\text{drawing}} = 0.60$, $SD_{\text{drawing}} = 0.15$, $M_{\text{writing}} = 0.51$, $SD_{\text{writing}} = 0.16$), $t(40) = 3.24$, $p = .006$, $d = 0.51$, but not in associative memory ($M_{\text{drawing}} = 0.39$, $SD_{\text{drawing}} = 0.21$, $M_{\text{writing}} = 0.36$, $SD_{\text{writing}} = 0.19$), $t(40) = 0.77$, $p = .444$, $d = 0.12$. In contrast, younger adults showed better performance after drawing than after writing in both memory types (associative: $M_{\text{drawing}} = 0.58$, $SD_{\text{drawing}} = 0.17$, $M_{\text{writing}} = 0.43$, $SD_{\text{writing}} = 0.25$; item: $M_{\text{drawing}} = 0.74$, $SD_{\text{drawing}} = 0.12$, $M_{\text{writing}} = 0.64$, $SD_{\text{writing}} = 0.17$), all ts$(37) \geq 3.12$, all ps $\leq .003$, all ds ≥ 0.51. These results indicated that across both sessions, younger adults benefited from drawing over writing in both memory types, while older participants benefited in item but not associative memory.

Next, we were interested in the interactions involving both the *Encoding Tool* and *Session* factors, because they would inform us of the differential consolidation effects on the two encoding tools. None of these interactions revealed significance, with all Fs$(1, 77) \leq 1.20$, all ps $\geq .278$, all η_p^2s $\leq .02$, thus not confirming that drawing led to stronger consolidation than writing, but suggesting instead that the beneficial effects of drawing remained stable over the interval of one week. We aimed to confirm this suggestion and computed further pairwise *t*-tests to compare the two encoding tools per age group, memory type, and session. The analyses revealed a beneficial effect of drawing compared to writing in younger adults' item and associative memory at both sessions, with all ts(38) ≥ 2.65, all ps $\leq .048$, all ds ≥ 0.43. In older adults, the beneficial effect on item memory at the immediate session, $t(41) = 3.01$, $p = .005$, $d = 0.47$, tendentially also remained after one week, $t(41) = 2.32$, $p = .050$, $d = 0.36$, whereas on associative memory, there was no such effect at any of the sessions, with all ts(41) ≤ 0.61, all ps $\geq .542$, all ds ≤ 0.10. The pairwise *t*-tests separated per session thus confirmed that the beneficial effects of drawing observed at the immediate session remained over the interval of one week.

We then resolved the significant *Age Group x Memory Type x Session* interaction, $F(1, 77) = 9.95$, $p = .002$, $\eta_p^2 = .11$. To this end, we investigated the interaction (*Age Group x Memory Type*) in a separate ANOVA per session. The interaction was significant in both ANOVAs, with $F(1, 77) = 4.88$, $p = .030$, $\eta_p^2 = .06$ at the immediate session, and $F(1, 77) = 4.71$, $p = .033$, $\eta_p^2 = .06$ at the delayed session. Pairwise *t*-tests to compare older vs. younger participants revealed that the age difference (i.e., younger adults performed better than younger) was slightly larger for associative memory, $t(77) = 3.54$, $p = .002$, $d = 0.80$, than item memory, $t(77) = 3.37$, $p = .002$, $d = 0.77$, at the immediate session. In contrast, the age difference was slightly larger for item memory, $t(77) = 4.82$, $p < .001$, $d = 1.08$, than for associative memory, $t(77) = 2.11$, $p = .038$, $d = 0.47$, at the second session. These results might indicate that the age difference in associative memory seemed to decrease after the delay of one week, while it instead increased in item memory over the same delay.

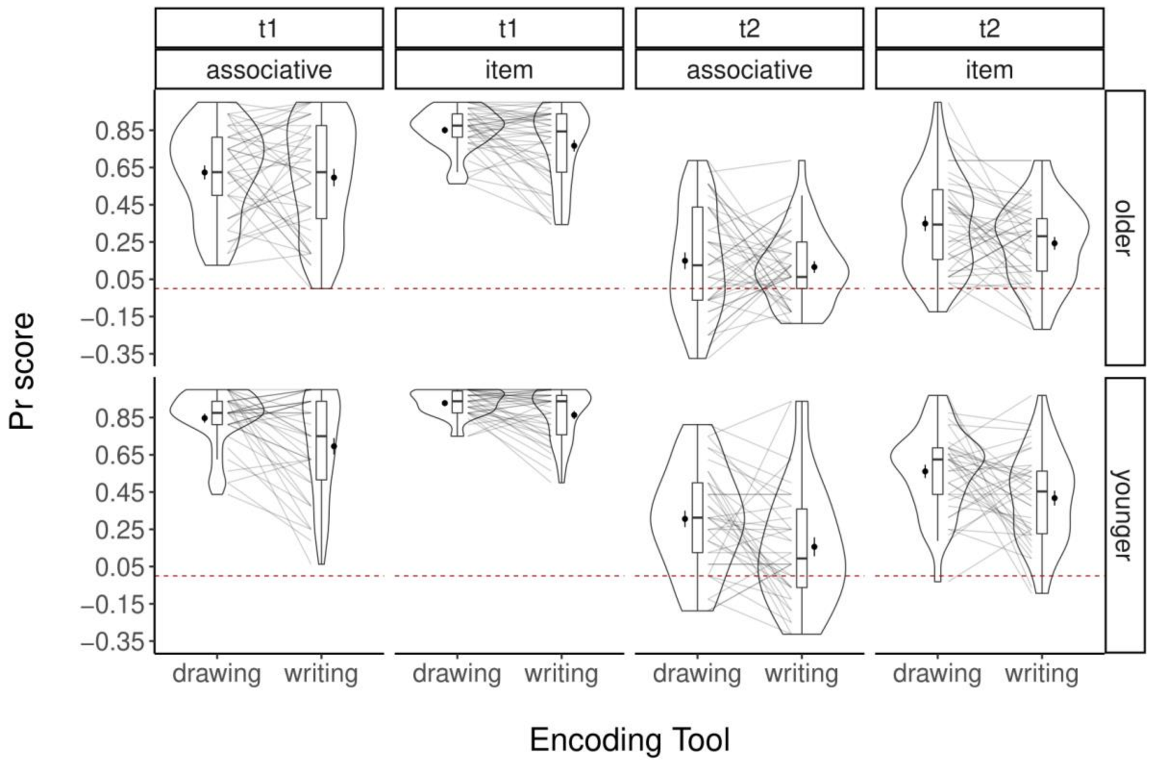

Figure 3. Recognition memory performance (Pr scores) as a function of age group, encoding tool, and memory type. T1 shows Pr scores from the immediate recognition phase and T2 for the recognition phase after one week. The ends of the gray lines represent the averaged Pr scores per participant in one condition. Points and error bars represent condition means and standard errors, respectively. The boxplots entail a horizontal line for the median (i.e., 50th percentile), with the lower and upper ends representing the 25th and 75th percentiles, and the whiskers extending to 1.5 times the interquartile range from either end of the box. The violin plots indicate the density of the data. The dashed line shows the level of guessing at a Pr score of zero.

Discussion

Overall, the results of Experiment 2 confirmed the results of Experiment 1. We observed a clear benefit of drawing in item and associative memory for younger adults. Importantly, we also observed a clear benefit of drawing in older participants' item memory, as we would have expected in Experiment 1. However, in contrast to the younger participants, the older participants did not benefit from drawing in associative memory, confirming the pattern of results observed in Experiment 1. Additionally, the observed benefit of drawing against writing remained over the interval of one week: younger participants still benefited in both memory types, and the benefit in older adults' item memory tendentially remained (p = .05). Compared to Experiment 1, we extended the encoding duration of the younger participants and matched it with that of the older ones (40 s). This adjustment allowed us to rule out longer encoding durations as an explanation for the lack of benefit from drawing in older adults' item memory, as observed in Experiment 1. Potential explanations for the persisting lack of benefit from drawing in older adults' associative memory are expanded on in the general discussion.

General Discussion

The main goal of the present study was to extend the reliably observed benefit of drawing as an encoding tool on item memory (e.g., Fernandes et al., 2018) to associative memory. As older adults are known to have difficulties, especially in associative memory (cf. Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008), we further aimed to investigate whether they might benefit to a greater extent from drawing over writing when compared to younger participants. In two experiments, we tested associative and item memory in a group of younger and a group of older participants and compared drawing (i.e., drawing of both words of a pair in one image) to repeatedly writing the words during encoding. The benefit of drawing on subsequent item memory was evident in both age groups (for older participants in Experiment 2), replicating the beneficial effect of drawing as an encoding tool on memory performance for individual items in older and younger adults (e.g., Meade et al., 2019, 2018).

In younger adults, we further showed that drawing relational images during encoding enhances subsequent recognition memory for associations. This result was consistent across the two experiments with immediate recognition tasks and in an additional delayed recognition task after one week (Experiment 2). Herewith, we provide reliable evidence that the benefit of drawing in item memory (Fernandes et al., 2018) extends to associative memory in younger adults. The integrated combination of the elaborative, pictorial, and motoric processes (for a review, see Fernandes et al., 2018) is thus effective not only for item memory but also for binding and remembering associations. As the control condition of repeatedly writing the words entails motoric processes to a similar extent as the experimental condition, we assumed that pictorial and elaborative processes were the predominant driving mechanisms behind the benefit of drawing that we found. This is in line with the LOP account (Craik & Lockhart, 1972), and in particular with findings of studies showing enhanced associative memory performance after more elaborative processing or mental imagery in younger adults (O'Hanlon et al., 2001).

The benefits of drawing on subsequent memory (i.e., on item and associative memory among younger adults and on item memory among older adults) persisted over a one-week delay. Thus, the drawn word pairs not only led to stronger memory traces over the short delay when compared to the memory traces resulting from written word pairs, but they presumably also underwent strong processes of system consolidation (cf. Dudai, 2004). As it has been suggested that the processes of system consolidation depend on the efficacy of the encoding strategy (Craik, 1983), these results further confirm the efficacy of drawing to encode items in both age groups and associations in younger adults.

Against our prediction, we could not extend the beneficial effect of drawing to associative memory in older adults. This result is in conflict with previous research investigating elaborative, pictorial, and motoric processes in isolation, which shows a positive impact of elaborative strategies (O'Hanlon et al., 2001), relational imagery strategies (e.g., Bender & Raz, 2012; Dunlosky et al., 2005; Fox et al., 2016; Hertzog et al., 2013; Naveh-Benjamin et al., 2007), or motoric actions (Bäckman & Nilsson, 1984; Feyereisen, 2009; Silva et al., 2015) on older adults' associative memory. Accordingly, the difficulties among older participants in benefiting from drawing in associative memory might therefore not be explained by an isolated deficit in one of the three processes. Our findings suggest that the simultaneous involvement of the processes during drawing (cf. Fernandes et al., 2018) was not helpful for older adults to encode associations. It seems that older adults were able to cope with the simultaneously involved processes of memorizing individual words (in Experiment 2, older adults benefited from drawing in item memory), but were overstrained when additionally creating the association.

One possible explanation for the results observed in older participants is that the processes involved in drawing, together with the creation of an association, led to divided attention. It has previously been shown that divided attention has a negative impact on associative memory in older adults (Naveh-Benjamin et al., 2003). On the one hand, it is conceivable that some of the processes involved in the drawing task might have worked as a distraction when older adults' cognitive processes were already occupied by creating the associations. As a result, older participants would have been able to create a relation between the word pairs, but the dual-task nature of the encoding task might have lowered its enhancing effect on subsequent associative memory. On the other hand, our observation that older adults benefited from drawing in item memory supports the opposite idea. It is thus more conceivable that the additional process of creating the association worked as a distraction that affected the enhancing processes involved in drawing. This assumption is supported by findings showing better item memory for pictures than words, but no such superior effect for pictures in associative memory (Guez & Lev, 2016; Ratcliff & McKoon, 2015). These findings indicate that the associative deficit in older age is modality independent (Guez & Lev, 2016).

Another potential explanation for why older adults did not benefit from drawing in associative memory is that they were unable to create a mental image of the association. We addressed this post hoc assumption in an explorative approach and rated the association in drawings from a randomly selected subsample. The ratings indeed indicated that older adults, compared to younger adults, seemed to have more difficulty finding a mental image that combined the two words.[2] This observation might suggest

---

[2] Two persons (not involved in the original study) rated the drawings from randomly selected older (n = 20) and younger (n = 20) participants regarding their level of association. Only participants from Experiment 2 were considered. Ratings were given on a scale between 1 and 5, with higher values pointing to stronger associations (1: two separated images for the word pair, or only one image for one of the words; 2: merely spatial relation between two images, e.g., beside each other, or related with arrows; 3: spatial relation

that the decline in episodic memory is situated at the creation of associations rather than at the retrieval of associations (cf. the associative deficit hypothesis, which states that the problem is based on difficulties in encoding and retrieval processes; Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). If our explorative results are confirmed in further experimental studies, they would have specific implications for experimental research using relational imagery as an encoding strategy. Specifically, self-report measures might not always be reliable for assessing whether older participants were indeed able to create relational images. In such studies, the use of drawing would provide an ideal tool for assessing the created relational images.

Although we did not observe age differences in the VVIQ, which reflects the ability to create vivid mental images (Cui et al., 2007; Marks, 1973), we found lower scores in older compared to younger participants on the VVQ2 scales, which indicates a preference for visual and imagery processes (Kirby et al., 1988; Richardson, 1977). This pattern is in line with previous findings, showing that despite having distinguishable brain activation patterns during mental imagery and sentence creation, older adults had difficulties using mental images successfully as a memory strategy (Bartsch et al., 2019). Together, these results suggest that the ability to create vivid mental images seems to be preserved in healthy aging, but that older adults nevertheless do not deploy imagery as an encoding strategy.

We might also mention the decrease in the age-related associative deficit in the delay after one week, although it should be interpreted with caution because it was

---

between the two images, e.g., nested or on top of one another, 4: substantial relation, but still both images distinguishable; 5: substantial relation to one 'new' image). Drawings were rated lower for the older (M = 2.6, SD = 0.71) than the younger subsample (M = 3.0, SD = 0.54) $t_{Welch}(35.67) = 2.41$, p = .032.

unspecific to the encoding tool. This observation is in conflict with previous research suggesting that the weakest consolidation effects exist in older adults' associative memory. In particular, studies comparing the consolidation of associative memory to item memory found stronger consolidation in item memory than associative memory in younger adults (e.g., Sopp et al., 2018; Studte, Bridger, & Mecklinger, 2015). Another study investigating the consolidation of associative memory reported less efficient consolidation in older adults than in younger adults (e.g., Mary et al., 2013). It might thus be that drawing as an encoding tool attenuated the expected large decay in older adults' associative memory. If so, older adults might have benefited from system consolidation after drawing in associative memory.

Taken together, we have shown that drawing relational images as an encoding tool is efficient not only for item memory but also for associative memory in younger adults, and that the observed benefits remained over one week. These results indicate that the processes simultaneously involved in drawing can also be used to enhance binding and remembering associations and to further promote system consolidation. However, in contrast to younger adults, older adults benefited from drawing in item memory, but not in associative memory. Tasking participants with manifesting relational images in drawings was not enough to support older participants in successfully creating associations and thus using appropriate encoding strategies. The pattern of results suggests that the age-related associative deficit hinges partly on the decreased use of appropriate strategies and partly on difficulties in the mere creation of an association.

To conclude, our drawing task did not provide the appropriate environmental support to overcome the problem of older adults spontaneously using efficient strategies for associative memory (cf. Craik, 1983; Shing et al., 2010). Thus, further research might investigate whether more specific instructions on how to draw the relations can also improve the benefit of drawing as an encoding tool in older adults. Moreover, understanding the underlying processes that led to the differences between age groups and memory types might provide a better understanding of the associative deficit in healthy aging and how to overcome it.

Author Contributions

R. Ovalle-Fresa developed the study concept and design, with critical input from C. Martarelli. R. Ovalle-Fresa programmed the tasks, performed the data analysis, and drafted the manuscript. R. Ovalle-Fresa supervised the data collection. C. Martarelli provided critical revisions to the manuscript and supervised the project. Both authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Ajanigha Arumaithurai, Vivianne Corrodi, Jana Gassan, Raphael Gfeller, Lilian Rau, and Tina Siegenthaler for their help with recruiting and testing. Moreover, we thank Urs Chalupny and Sebastian Totzauer for rating a subset of the drawings.

Open Practices Statement

The data and analyses for all experiments are available at https://osf.io/3y8eb/. The experiments were preregistered.

References

Bäckman, L., & Nilsson, L.-G. (1984). Aging effects in free recall: An exception to the rule. Human Learning: Journal of Practical Research & Applications, 3(1), 53–69.

Bartsch, L. M., Loaiza, V. M., Jäncke, L., Oberauer, K., & Lewis-Peacock, J. A. (2019). Dissociating refreshing and elaboration and their impacts on memory. *NeuroImage*, *199*(March), 585–597. https://doi.org/10.1016/j.neuroimage.2019.06.028

Bender, A. R., & Raz, N. (2012). Age-Related differences in recognition memory for items and associations: Contribution of individual differences in working memory and metamemory. *Psychology and Aging*, *27*(3), 691–700. https://doi.org/10.1037/a0026714

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *302*, 341–359. https://doi.org/10.1098/rstb.1983.0059

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684.

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *12*, 599–607. https://doi.org/10.1016/S0022-5371(73)80039-8

Cui, X., Jeter, C. B., Yang, D., Montague, P. R., & Eagleman, D. M. (2007). Vividness of mental imagery: Individual variability can be measured objectively. *Vision Research*, *47*(4), 474–478. https://doi.org/10.1016/j.visres.2006.11.013

Dudai, Y. (2004). The neurobiology of consolidations, or, how stable is the engram? *Annual Review of Psychology*, *55*, 51–86. https://doi.org/10.1146/annurev.psych.55.090902.142050

Dunlosky, J., Hertzog, C., & Powell-Moman, A. (2005). The contribution of mediator-based deficiencies to age differences in associative learning. *Developmental Psychology*, *41*(2), 389–400. https://doi.org/10.1037/0012-1649.41.2.389

Engelkamp, J., & Zimmer, H. D. (1997). Sensory factors in memory for subject-performed tasks. *Acta Psychologica*, *96*(1–2), 43–60. https://doi.org/10.1016/s0001-6918(97)00005-x

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fernandes, M. A., Wammes, J. D., & Meade, M. E. (2018). The surprisingly powerful influence of drawing on memory. *Current Directions in Psychological Science*, *27*(5), 1–7. https://doi.org/10.1177/0963721418755385

Feyereisen, P. (2009). Enactment effects and integration processes in younger and older adults' memory for actions. *Memory*, *17*(4), 374–385. https://doi.org/10.1080/09658210902731851

Fox, M. C., Baldock, Z., Freeman, S. P., & Berry, J. M. (2016). The role of encoding strategy in younger and older adult associative recognition: A think-aloud analysis. *Psychology and Aging*, *31*(5), 471–480. https://doi.org/10.1037/pag0000111

Goldhahn, D., Eckart, T., & Quasthoff, U. (2012). Building large monolingual dictionaries at the leipzig corpora collection: From 100 to 200 languages. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 759–765.

Grady, C. L., & Craik, F. I. M. (2000). Changes in memory processing with age. *Current Opinion in Neuro Biology*, *10*, 224–231. https://doi.org/10.1093/ejil/chq032

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, *102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Guez, J., & Lev, D. (2016). A picture is worth a thousand words? Not when it comes to associative memory of older adults. *Psychology and Aging*, *31*(1), 37–41. https://doi.org/10.1037/pag0000069

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2022). lab.js: A free, open, online study builder. *Behavior Research Methods*, *54*, 556–573. https://doi.org/10.3758/s13428-019-01283-5

Hertzog, C., Fulton, E. K., Mandviwala, L., & Dunlosky, J. (2013). Older adults show deficits in retrieving and decoding associative mediators generated at study. *Developmental Psychology*, *49*(6), 1127–1131. https://doi.org/10.1037/a0029414

Kamp, S.-M., & Zimmer, H. D. (2015). Contributions of attention and elaboration to associative encoding in young and older adults. *Neuropsychologia*, *75*, 252–264. https://doi.org/10.1016/j.neuropsychologia.2015.06.026

Kirby, J. R., Moore, P. J., & Schofield, N. J. (1988). Verbal and visual learning styles. *Contemporary Educational Psychology*, *13*(2), 169–184. https://doi.org/10.1016/0361-476X(88)90017-3

Lakens, D., & Caldwell, A. R. (2021). Simulation-based power analysis for factorial analysis of variance designs. *Advances in Methods and Practices in Psychological Science*, *4*(1), 1–14. https://doi.org/10.1177/2515245920951503

Limesurvey GmbH. (2018). *LimeSurvey: An open source survey tool*.

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(3), 671–685. https://doi.org/10.1037/a0018785

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Mary, A., Schreiner, S., & Peigneux, P. (2013). Accelerated long-term forgetting in aging and intra-sleep awakenings. *Frontiers in Psychology*, *4*(October), 1–11. https://doi.org/10.3389/fpsyg.2013.00750

Meade, M. E., Ahmad, M., & Fernandes, M. A. (2019). Drawing pictures at encoding enhances memory in healthy older adults and in individuals with probable dementia. *Aging, Neuropsychology, and Cognition*, 1–22. https://doi.org/10.1080/13825585.2019.1700899

Meade, M. E., Wammes, J. D., & Fernandes, M. A. (2018). Drawing as an encoding tool: memorial benefits in younger and older adults. *Experimental Aging Research*, *44*(5), 369–396. https://doi.org/10.1080/0361073X.2018.1521432

Meier, B., & Rothen, N. (2013). Grapheme-color synaesthesia is associated with a distinct cognitive style. *Frontiers in Psychology*, *4*, 632. https://doi.org/10.3389/fpsyg.2013.00632

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J. L., & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695–699. https://doi.org/10.1111/j.1532-5415.2005.53221.x

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cogition*, *26*(5), 1170–1187. https://doi.org/10.1037//0278-7393.26.5.1170

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, *22*(1), 202–208. https://doi.org/10.1037/0882-7974.22.1.202

Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory and Cognition*, *31*(7), 1021–1035. https://doi.org/10.3758/BF03196123

O'Hanlon, L., Wilcox, K. A., & Kemper, S. (2001). Age differences in implicit and explicit associative memory: Exploring elaborative processing effects. *Experimental Aging Research*, *27*(4), 341–359. https://doi.org/10.1080/03610730109342353

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, *76*(3), 241–263. https://doi.org/10.1037/h0027272

R Core Team, R. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ratcliff, R., & McKoon, G. (2015). Aging effects in item and associative recognition memory for pictures and words. *Psychology and Aging*, *30*(3), 669–674. https://doi.org/10.1037/pag0000030

Richardson, A. (1977). Verbalizer-visualizer: A cognitive style dimension. *Journal of Mental Imagery*, *1*(1), 109–126.

Salthouse, T. A. (1995). Selective influences of age and speed on associative memory. *The American Journal of Psychology*, *108*(3), 381–396. https://doi.org/10.2307/1422896

Shing, Y. L., Werkle-Bergner, M., Brehmer, Y., Müller, V., Li, S. C., & Lindenberger, U. (2010). Episodic memory across the lifespan: The contributions of associative and strategic components. *Neuroscience and Biobehavioral Reviews*, *34*(7), 1080–1091. https://doi.org/10.1016/j.neubiorev.2009.11.002

Silva, A. R., Pinho, M. S., Souchay, C., & Moulin, C. J. A. (2015). Evaluating the subject-performed task effect in healthy older adults: Relationship with neuropsychological tests. *Socioaffective Neuroscience & Psychology*, *5*(1), 24068. https://doi.org/10.3402/snp.v5.24068

Sopp, M. R., Michael, T., & Mecklinger, A. (2018). Effects of early morning nap sleep on associative memory for neutral and emotional stimuli. *Brain Research*, *1698*, 29–42. https://doi.org/10.1016/j.brainres.2018.06.020

Studte, S., Bridger, E., & Mecklinger, A. (2015). Nap sleep preserves associative but not item memory performance. *Neurobiology of Learning and Memory*, *120*, 84–93. https://doi.org/10.1016/j.nlm.2015.02.012

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). Academic Press.

Unesco Institute for Statistics. (2012). *International Standard Classification of Education (ISCED) 2011*.

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Wammes, J. D., Jonker, T. R., & Fernandes, M. A. (2019). Drawing improves memory: The importance of multimodal encoding context. *Cognition*, *191*, 103955. https://doi.org/10.1016/j.cognition.2019.04.024

Wilson, R. S., Boyle, P. A., Segawa, E., Yu, L., Begeny, C. T., Anagnos, S. E., & Bennet, D. A. (2013). The influence of cognitive decline on well-being in old age. *Psychology and Aging*, *28*(2), 304–313. https://doi.org/doi:10.1037/a0031196

Wittich, W., Phillips, N., Nasreddine, Z. S., & Chertkow, H. (2010). Sensitivity and specificity of the Montreal Cognitive Assessment modified for individuals who are visually impaired. *Journal of Visual Impairment and Blindness*, *104*(6), 360–368. https://doi.org/10.1177/0145482x1010400606