The Rise and Fall of Immediate and Delayed Memory for Verbal and Visuospatial Information from Late Childhood to Late Adulthood

Jaap M. J. Murre[1], Steve M. J. Janssen[2], Romke Rouw[1], & Martijn Meeter[3]
[1]University of Amsterdam, [2]Duke University, [3]VU University Amsterdam

Abstract

Over 100,000 verbal and visuospatial immediate and delayed memory tests were presented via the Internet to over 28,000 participants in the age range of 11 to 80. Structural equation modeling pointed to the verbal versus visuospatial dimension as an important factor in individual differences, but not the immediate versus delayed dimension. We found a linear decrease of 1% to 3% per year in overall memory performance past the age of 25. For visuospatial tests, this decrease started at age 18 and was twice as fast as the decrease of verbal memory. There were strong effects of education, with the highest educated group sometimes scoring one full standard deviation above the lowest educated group. Gender effects were small but as expected: women outperformed men on the verbal memory tasks; men outperformed women on the visuospatial tasks. We also found evidence of increasing proneness to false memory with age. Memory for recent news events did not show a decrease with age.

Keywords: memory, aging, visual memory, verbal memory, short-term memory, working memory, long-term memory

The Rise and Fall of Immediate and Delayed Memory for Verbal and Visuospatial Information from Late Childhood to Late Adulthood

1. Introduction

When first introduced, theories and models of short-term memory (Atkinson & Shiffrin, 1968; Miller, 1956/1994; Peterson & Peterson, 1959) and working memory (Baddeley & Hitch, 1974) typically considered these immediate memory stores to have roughly the same capacity from person to person. Miller (1956/1994) noticed that short-term memory seemed to be able to contain about seven 'chunks', and Baddeley and Hitch (1974) proposed a phonological loop that can contain about two seconds worth of phonological information. It became clear, however, that there is in fact great variation between individuals and that these differences are predictive of other types of skills and capacities. For example, individual differences in working memory capacity may be highly predictive of reading comprehension (Daneman & Carpenter, 1980) and correlate with measures of IQ (Deary, Penke, & Johnson, 2010).

One may argue that short-term memory or working memory relies on temporary activation of long-term memory traces (Murre, Wolters, & Raffone, 2006) and that it is, thus, the 'gateway' to long-term memory. Based on this idea, one might conjecture that individuals with an excellent short-term or working memory will also have a good long-term memory. After all, the longer one keeps items in short-term memory, the higher the chances of transference to long-term memory (Atkinson & Shiffrin, 1968).

Evidence from brain imaging, however, suggests that the relationship between immediate (short-term or working memory) and delayed (long-term memory) may be more complicated and involves aspects, such as efficiency of encoding, that are not related to capacity issues. Individual differences in visual working memory, for example, could be predicted with EEG/ERPs (Vogel, McCollough, & Machizawa, 2005). This study found that individuals with a high capacity were much more efficient at representing only the relevant items than the low capacity individuals. In fact, low capacity individuals sometimes stored more information in memory than high capacity individuals. There are more factors that may disrupt correlations between immediate and delayed memory. For example, individual differences in neuromodulators, such as acetylcholine and dopamine, could affect long-term memory storage more than immediate memory. So, the question remains whether the connection between immediate and delayed memory is strong enough to find a high correlation between these two systems.

The immediate-versus-delayed dimension may be contrasted with the other major dimension of working memory, namely that of verbal versus visuospatial modality, which form the main 'slave systems' in the Baddeley and Hitch Working Memory model (Baddeley, 2003; Baddeley & Hitch, 1974). Several studies have shown that individuals may differ greatly on this dimension (e.g., Alloway, Gathercole, & Pickering, 2006; Ardila & Rosselli, 1994; Bopp & Verhaeghen, 2007; Gathercole, Pickering, Ambridge, & Wearing, 2004; Herlitz, Airaksinen, & Nordstrom, 1999; Park, et al., 2002). In this study, we contrasted the immediate-versus-delayed dimension with verbal-versus-visuospatial dimension, examining whether people have a good memory for either verbal or visual (or spatial) materials and whether they have either a good immediate or a good delayed memory for these materials.

Because it is well known that age, education and gender interact with memory performance, the effects of these demographic variables were investigated as well. We had a particular interest in the effects of age on memory. To examine whether verbal and visuospatial memory and immediate and delayed memory show a similar decrease with age, we designed an online battery with a wide array of tests. Some studies have found a higher rate of decrease with age for the visuospatial tests than for the verbal tests (Bopp & Verhaeghen, 2007; Jenkins, Myerson, Joerding, & Hale, 2000; Turcotte, Gagnon, & Poirier, 2005), but other studies did not replicate this finding (Park, et al., 2002). We wanted to unravel more details of the pattern of decreasing memory performance with age. Recent studies (Lovden, 2003; Schacter, Israel, & Racine, 1999) suggested, for example, that older subjects are more prone to false memories. We, therefore, also included tests in the online battery to see whether we could find further evidence for this effect and how it relates to other aspects of memory performance.

Using Internet-based testing allowed us to administer about 100,000 memory tests to a total of about 28,000 Dutch people. These large numbers enabled us to use Structural Equation Modeling to disentangle the effects of age, education, and gender and the other dimensions of interest. Internet-based testing also has clear limitations, though many of these can be circumvented with the correct precautions (Gosling, Vazire, Srivastava, & John, 2004; Reips, 2000, 2002; Skitka & Sargis, 2006). The most notable limitation for our purposes was that the length of a single test session should be limited to about 20 minutes. We have found that participants are often well motivated during twenty-odd minutes and make for excellent subjects for research in experimental psychology, but they start to drop out of online experiments after this interval. The twenty-minute window was used to present the participants with a quasi-randomly assigned subsets of the larger test battery, with the option -but not the requirement- to take several subsets on one or different occasions. The study was therefore planned with missing data in mind, assuming that the majority of participants would take only a subset of all tests.

To assert the validity of our approach, the experiment was first conducted in a normal psychological laboratory on a small sample of participants with satisfactory results (not reported here). Other measures were taken as well to ensure validity and reliability, based on our earlier experiences with Internet-based research (e.g., Janssen, Chessa, & Murre, 2005, 2006, 2007; Janssen & Murre, 2008; Janssen, Murre, & Meeter, 2008; Kristo, Janssen, & Murre, 2009; Meeter, Murre, & Janssen, 2005; Meeter, Ochtman, Janssen, & Murre, 2010). The test battery was promoted nationally (in the Netherlands) as *The National Memory Test* with coverage in several national newspapers and other media, making additional advertising and rewards or other incentives superfluous.

In this paper, we will thus examine the relative importance of the two dimensions *visuospatial versus verbal memory* and *immediate versus delayed memory* in individual differences of memory performance. Our study furthermore investigates the dependence of these dimensions across age, gender, and level of education. Recent research by Johnson, Logie, and Brockmole (2010) also suggests that working memory is not unitary and that its structure changes across the life span, though this study did not contrast the verbal/visual nor an immediate/delayed dimensions directly.

2. Method

2.1. Participants

All participants took part on their own volition via the Internet. They could come into contact with our website in at least four ways: (1) through links on other websites, (2) through search engines, (3) through promotion in traditional media, such as articles in newspapers and magazines, which included our web address, or (4) through word of mouth. At the end of the test, participants could invite relatives, friends, and colleagues by sending them standardized e-mails.

In total, 28,116 Dutch volunteers, who were between the ages 11 and 80 years (*M* age = 37.34, *SD* = 16.15, of whom 34.7% males and 65.3% females), participated. The age distribution is given in Figure 1 and the age distribution per test is given in Table 1.b. Participants who were younger than 11 or older 80 years or who did not reside in the Netherlands were allowed to take the tests, but their results were not included in the analyses.

---INSERT FIGURE 1 ABOUT HERE---

As part of the registration procedure, in addition to age, gender and country of residence, we asked for the level of education using an eight-point scale that is customary in the Netherlands. Participants could have finished primary school (*N* = 2776) or one of four levels of secondary schooling (from low to high): preparatory vocational school (designated in Dutch as LBO; *N* = 1759), mid-level general education (MAVO; *N* = 3705), high-level general education (HAVO; *N* = 3049), or secondary education preparing for university (VWO; *N* = 2469). At the tertiary level, the highest level attained could be vocational school (roughly equivalent to community college; MBO; *N* = 4228), higher professional education (HBO; *N* = 5934), or university (WO; *N* = 3504). (692 participants did not provide their level of education.)

2.2. Materials

The choice of tests was rather clear-cut. We needed tests that measure immediate and delayed memory and verbal and visuospatial memory. We selected a number of well-known tests and added some that we felt were suitable for our goal. We were also interested in the effects of age on gist (and hence false) memory and added some tests for that as well (see Table 1a-b for an overview of the tests). Each test is described in some detail below. To facilitate referring to the tests, they are given abbreviations that indicate whether they measured (primarily) immediate memory (labeled as STM for brevity) and delayed memory (labeled as LTM).

---INSERT TABLE 1a-b ABOUT HERE---

2.2.1. Verbal Immediate Memory

2.2.1.1. Ten-Words Test

In the Ten-Words Test (Spaan, Raaijmakers, & Jonker, 2005), participants were shown ten words one by one (10WT-STM). The order of the words was randomized between participants. All words were Dutch, concrete, medium-frequency nouns. Each word was presented for 2.5 seconds. After the presentation, participants were instructed to count backwards from a three-digit number (e.g., 317) in steps of three (e.g., 314, 311, 308, etc.). After twenty seconds of counting backwards, participants saw a text field in which they could enter the words they had seen during the presentation. Participants could recall the words in any order (i.e., free recall) using a form with ten text fields. The cycle of presentation and testing was repeated three times. The test score was the sum of correct responses over all three phases. The scores could therefore range from 0 to 30. Participants with the maximum score were removed because of possible cheating.

2.2.1.2. Deese-Roediger-McDermott Immediate Recall

The test (DRM-STM) was based on the Deese-Roediger-McDermott procedure for creating false memories (Deese, 1959; Roediger & McDermott, 1995). Four lists with the strongest associates of a certain word that was not included (the *critical lure*) were presented (2.5 s per word). For example, the twelve strongest associates of “chair” were on the list, but not the word “chair” itself. We used the Dutch translations of the English words of four of the lists used by Roediger and McDermott. Participants were first told to learn the lists of words. Immediately after each list presentation, the participants had to enter the words that they could remember in a list of text fields in any order (i.e., free recall). The score on this test, the total number of words correctly recalled, could range from 0 to 48 (4 lists of 12 words each). This test was always followed by its delayed variant (see DRM-LTM below), after an intermediate non-verbal task (Memory Game).

2.2.1.3. Story Telling Immediate Recognition

Participants were asked in this test (Story-STM) to learn ten consecutively presented sentences. Each sentence was visible for five seconds. After the last sentence, a twenty-second delay followed in which the participants were requested to engage in backward counting from 211 in steps of three. After counting backwards, ten sentences were presented. Half of the sentences had been slightly changed, whereas the other half was unchanged. The participants had to indicate whether each test sentence was exactly the same as the sentence in the learning phase. Scores could possibly range from 0 to 10 points, but scores below chance level (5 points) were not used in the data analyses. This test was later followed by a delayed variant (see Story-LTM below), after an intermediate non-verbal task (Texture-STM).

2.2.1.4. Digit Span Task

A series of digits was presented in this test (Digit-STM) one-by-one in a large format (200 x 200 pixels) on the upper half of the screen. Each digit was visible for 2.5 seconds, whereas the inter-item interval was 1.5 seconds. After the sequence had ended, participants had to click on pictures of the digits presented as a row of small pictures (80 x 80 pixels) on the lower half of the screen. The responses had to be in the order of presentation. The task started with a sequence length of two. If a participant had reproduced the sequence correctly twice, the sequence length increased with one. If a participant had failed to reproduce the sequence correctly twice, the test ended and the sequence-length was taken as the digit span score. The maximum sequence length was nine digits.

2.2.2. Visuospatial Immediate Memory

2.2.2.1. Pattern Span Task

The pattern span task (Pattern-STM) was an adaptation of the task used in Della Sala, Gray, Baddeley, Allamano, & Wilson (1999). In the Pattern-STM, participants saw an array of, initially, two by two blocks. When they pressed the start button, half of the blocks in the array changed color simultaneously. This configuration of colored blocks was presented for two seconds, after which all blocks returned to their original white color. The participants had to click the blocks that had changed color. When at least two out of three answers were correct the array increased by size. If not, the task ended. There were three trials for every array size. The array sizes used were 2x2, 2x3, 3x4, 4x5, 5x6 and 6x7 blocks. Rather than merely reporting the span, we used a more sensitive measure which was the sum total of all blocks seen. Thus, after completing all 2x2 arrays correctly, the score would be 3x2x2 = 12, after the 2x3 arrays it would be 12 + 3x2x3 = 30, etc. A score of 126 indicates a pattern span (in the sense of Della Sala et al., 1999) of about 8.5 to 9.

2.2.2.2. Corsi Block Tapping Task

This spatial working memory test (Corsi-STM) was an adaptation of the widely used Corsi Block Tapping Task (Corsi, 1972; Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000; Milner, 1971). The participants saw nine blocks randomly placed on the screen in a fixed layout. After the participants had pressed the start button, a sequence started and the blocks lit up one by one. Each block lit up for 1.5 seconds, whereas the inter-item interval was 0.5 seconds. The first sequence consisted of five blocks. When participants had reproduced a sequence correctly, by clicking the blocks in the correct order, the sequence-length increased with one block on the subsequent trial. When they failed to reproduce the sequence, the sequence-length decreased with one block. The minimum sequence-length was two blocks, the maximum nine. The score was the maximum reproduced sequence-length.

2.2.2.3. Texture Span Task

In this test (Texture-STM), participants had to remember 11 pictures with pseudo-random lines and figures on them (textures). The textures were selected such they were hard to encode verbally. After participants had pressed a start button, a sequence of pictures (200 x 200 pixels) was presented on the upper half of the screen. Each picture was visible for 2.5 seconds. The inter-item interval was 1.5 seconds. After presentation of the sequence, a row of smaller pictures of the stimuli appeared on the lower half of the screen. Participants had to click on the smaller pictures in the same order in which the textures had been presented. The test started with a sequence length of two. This was repeated three times. If the participant remembered two sequences correctly, the sequence length increased. If not, the test ended and the maximum correctly answered sequence-length was taken as the visual texture span score. The maximum sequence-length presented was nine pictures.

2.2.3. Verbal Delayed Memory

2.2.3.1. Deese-Roediger-McDermott Delayed Recognition

This test (DRM-LTM) was follow-up of the immediate recall variant of this test, the DRM-STM. It was always presented last in the fixed test series: DRM-STM, Memory Game (described below), and DRM-LTM. From each of the four lists of the learning phase (i.e., of the DRM-STM), two 'old' words were selected. These were presented with the critical lure (i.e., the non-presented word with which all list words were associated), two new words that were weakly associated with the critical lure but not given on the original studied list, and two new words that were not associated with the critical lure. Words were presented one by one in random order. For each word, participants indicated on a four-point scale how certain they were whether the word had been presented during the learning phase (i.e., during the preceding DRM-STM), ranging from 'I am sure the word is new', 'I think the word is new', and 'I think the word is old', to 'I am sure the word is old'.

Participants were awarded three points for each correct answer about which they were certain. Two points were given to the participants when the answer was correct, but they were not confident. Participants received one point when the answer was incorrect, but they were not confident. No points were awarded when the participants were certain about their incorrect answer. Thus, the total test score on four lists of seven items could possibly range from 0 to 84. Scores below chance level (42 points) were not used in the data analyses.

2.2.3.2. Story Telling Delayed Recognition

This task was presented after the Story-STM and Texture-STM, both described above. In the Story Telling Delayed Recognition test (Story-LTM), ten sentences were presented. Half of these sentences were correct paraphrases of sentences presented in the learning phase of the Story-STM, half were incorrect paraphrases. Participants had to indicate whether the gist of sentences was similar to the gist of a sentence presented in the learning phase. They were instructed to focus on global similarity (gist) only. Scores could range from 0 to 10 points, but scores below chance level (5 points) were not used in the data analyses.

2.2.4. Visual Delayed Memory

2.2.4.1. Memory Game

The task (MemGame-LTM) has all the characteristics of the well-known game *Memory*. It was based on the task developed by Duff and Hampson (2001). Participants were presented with a grid of five by four squares or ‘cards’. At first, all the squares were white, but when the participant clicked on a square, it changed its color (color of the ‘card’ was revealed). There were ten pairs in different colors. The goal of the task was to find all matching squares. When participants had clicked on two squares of different colors, both squares turned white again after a one-second interval. Even though there were only twenty ‘cards’, the fact that all ‘cards’ remained in the game until the end made the test difficult. To make the task easier, the participants were kept abreast of colors they had already matched successfully (but not their positions). After the participants had solved the task the first time, they started over with the same underlying color configuration. In total, they performed the task three times. With a perfect memory for the configuration, the second and third attempt could be done in 10 trials.

2.2.5. Other Delayed Memory Test

2.2.5.1. Daily News Memory Test

This test involves memory for public events that occurred within the past two years. In the Daily News Memory Test (DNMT), participants were given ten open-ended and twenty multiple-choice questions about events that had briefly been front page news in the past two years (see Meeter, et al., 2005; Meeter, et al., 2010, p. for more detailed descriptions). If participants did not know the correct answer to the open-ended questions, they were encouraged to guess. If they really did not know the correct answer, they were asked to enter a question mark. The multiple-choice questions were four-alternative forced-choice questions. To proceed to the next question, participants had to select one of the four options. The range of the total score ran from 0 to 30. Results of participants who scored below chance-level on the twenty multiple-choice questions (5) were omitted from the analyses.

2.3. Procedure

The participants visited the website and, after registering or logging in, made one or more tests in random order. As described above, some tests had an immediate and delayed equivalent (e.g., DRM-STM and DRM-LTM), which were presented before and after an intermediate test (e.g., MemGame-LTM). Presentation of all tests was visual and computer-controlled.

Results of incomplete tests were recorded, but they were omitted from the analyses. Results of tests on which performance was below chance were also not included in the analyses. Participants had to complete at least two memory tests to which they had been assigned to be included in the analyses. Participants completed on average 3.80 memory tests.

3. Results

We first analyzed the results of the tests separately. Table 2 gives the number of participants after we omitted the results of incomplete tests and the results of tests with scores that were below chance. Table 2 also gives the mean scores per test. Three-way ANOVAs were conducted for each test. The F-values of each test for the effects of gender (2 levels), age group (14 levels) and level of education (8 levels) are given in Table 3, with higher-order interactions given in Table 4. These levels were the same for all ANOVAs. We generally found weak effects of gender and strong effects of age and education. In the following sections, the results for gender, education, and age will be reviewed in more detail. More extensive analyses with structural equation modeling are presented later.

---INSERT TABLES 2, 3 AND 4 ABOUT HERE---

3.1. Effects of gender

Female participants performed better on the DRM-STM, DRM-LTM, 10WT-STM and MemGame-LTM (*p*s < .001), whereas male participants performed better on the Digit-STM, Pattern-STM, Corsi-STM and DNMT (*p*s < .05). No significant differences between men and women were found on the scores of the Story-STM, Story-LTM and Texture-STM (*p*s ≥ .350). Whereas the analyses of variance were conducted with the raw scores on the tests, z-transformed scores were calculated to compare performance the verbal and visuospatial tests (taking STM and LTM together). When averaging the z-scores, the often reported finding that women tend to perform better on verbal tasks was obtained, although the difference was small in our study. Women performed better on the visuospatial MemGame-LTM, $F(1, 9185) = 96.71$, $p < .001$, $\eta_p^2 = .010$. When this test is excluded from the analyses, men performed better on the visuospatial tasks (see Figure 2).

---INSERT FIGURE 2 ABOUT HERE---

3.2. Effects of education

Education had a clear effect on the test scores: participants with higher levels of education (i.e., HAVO, HBO, VWO and WO) outperformed those with lower levels of education (i.e., lagere school, LBO, MAVO and MBO) on all tests (*p*s < .001). The effect is shown in Figure 3. To make the figure easier to read, z-transformed scores were calculated and the results of participants with MAVO and MBO, with HAVO and HBO, and with VWO and WO were combined. We combined these levels because they represent similar achievement levels of education, where MAVO is completed before a student (typically) moves on to MBO, HAVO students continue to HBO and VWO students to WO. In other words, these pairs represent three tiers of the Dutch educational system, with VWO/WO being the highest one (Scientific/University level). Education affects performance on verbal and visuospatial tests in a similar manner and the age-related decrease gives rise to parallel curves ordered by level of education.

---INSERT FIGURE 3 ABOUT HERE---

3.3. Effects of age

All tests showed a significant age effect (*p*s < .001), except Story-LTM ($p = .12$). There was a clear peak in the performance as a function of age. For visuospatial memory, the peak was centered on the 16-18 age-bin. There was no clear peak for the verbal tests; memory scores reached maximum performance between ages 14-26 (see Figure 4). After these age periods, there was a linear decrease in performance as measured in averaged Z-scores. Using linear regression, we calculated slopes, intercepts, r2, and average values in the age-range 42-70, which was most stable for most tests (see Table 5). Using larger age ranges (e.g., 42-80 years) gave similar results, but are not reported here.

The average decrease rate with age was more than twice as high for the visuospatial tests as for the verbal tests (26.6% per 10 years, $R^2 = 98.7\%$, vs. 12% per 10 years, $R^2 = 97.0\%$, respectively). If we plot averaged Z-scores based on the immediate versus delayed distinction, we see no clear difference, though past the age of 40 immediate memory seems to show a smaller drop compared with delayed memory (see Figure 5). Linear regression lines from 42 to 80 years, give slopes of -0.204 and -0.170 (with intercepts of 0.825 and 0.753; r^2 is 0.950 and 0.935) for, immediate and delayed memory, respectively. On average, immediate memory is 0.136 standard deviations below delayed memory in this age range.

--- INSERT TABLE 5 ABOUT HERE ---

Figure 6 gives the grand average over all tests, with the results of the DNMT plotted separately. A straight line fits this combined data well for the averaged Z-scores after the age of 25, explaining 98.5% of the variance. The Z-scores decrease with 18.5% per ten years. Performance on the DNMT shows a different pattern with a steady increase in performance that reaches a steady-state after age 24.

There are, in general, large differences between the age-related decreases in performances on individual tests. The 10WT-STM, for example, shows a decrease of 26.6% (i.e., 0.26 standard deviations) per 10 years ($R^2 = 96.5\%$), whereas the Digit-STM decreases only 9.2% per 10 years ($R^2 = 67.4\%$). The Digit Span Task is thus a less sensitive test for age differences than the Ten-Words Test, which may explain the value of the latter test in the early assessment of dementia (Spaan, et al., 2005).

---INSERT FIGURES 4, 5 AND 6 ABOUT HERE---

3.4. Interactions between gender, education, and age

There were also thirteen significant interaction effects. Women outperformed men on the MemGame-LMT to a greater degree for older participants (F(13, 9185) = 2.71, $p < .01$, η_p^2 = .004), whereas men outperformed women to a greater degree for older participants on the Pattern-STM and the DNMT (F(13, 10392) = 1.85, $p < .05$, η_p^2 = .002; F(13, 6317) = 2.12, $p < .05$, η_p^2 = .004). As participants became older, effects of education became smaller on six tests (MemGame-LTM, F(95, 9185) = 2.16, $p < .001$, η_p^2 = .022; DRM-STM, F(96, 11967) = 1.83, $p < .001$, η_p^2 = .014; 10WT-STM, F(96, 10498) = 1.76, $p < .001$, η_p^2 = .016; DNMT, F(94, 6317) = 1.52, $p < .01$, η_p^2 = .022; DRM-LTM, F(95, 7912) = 1.49, $p < .01$, η_p^2 = .018; Pattern-STM, F(96, 10392) = 1.34, $p < .05$, η_p^2 = .012). The effect of education was smaller on the scores of female participants than on the scores of male participants in the MemGame-LTM (F(8, 9185) = 2.30, $p < .05$, η_p^2 = .002). There were also two three-way interactions between gender, age group and level of education on scores of the Corsi-STM and the MemGame-LTM, F(95, 10119) = 1.29, $p < .05$, η_p^2 = .012 and F(91, 9185) = 1.84, $p < .001$, η_p^2 = .018.

3.5. False memory

One task measured to what extent participants tend to falsely recognize an unseen semantic prototype, namely the Deese-Roediger-McDermott Delayed Recognition Test (DRM-LTM). In this task, participants indicated how certain they were that a word had or had not been presented in a preceding test. This earlier DRM-STM test can best be seen as a straightforward immediate free recall task, whereas the subsequent DRM-LTM test is an old-new recognition task in which participants have to answer on a four-point scale (i.e., surely new, probably new, probably old, and surely old).

For the previous analyses of variance, we had calculated one score for the DRM-LTM, but Table 6 gives the answers on the four-point scale separately for the studied words and the three types of new words: unrelated words, weakly related words and critical lures. Participants almost always reported correctly (surely old or probably old) that they had seen the studied words during the learning phase (M = .934). They also indicated often correctly (surely new or probably new) that the unrelated words (M = .985) and weakly related words (M = .828) were new, but they less often reported that they had not seen the critical lures before (M = .328). For the majority of the critical lures, participants indicated incorrectly that they had studied them during the learning phase (M = .672). As can be observed in Figure 7 and Table 5, performance on both tasks decreases with age, but the decrease is stronger on the DRM-STM test.

An interesting pattern of age-related change is observed when we plot the scores on DRM-LTM separately for the four types of words, as shown in Figure 8. In this figure, only the proportions of correct responses about which the participants were certain (‘surely old’ or ‘surely new’, whichever is correct) were taken. The decrease of correct responses to critical lures (i.e., a ‘surely new’ judgment) seems to be faster than for the other types of words. A linear regression on the decrease (excluding the two youngest age groups for reasons of linearity) confirmed this. The slopes of the four types words were 0.0005 (unrelated words), -0.0071 (studied words), -0.0026 (weakly related words), and -0.0121 (critical lures). Thus, the tendency to incorrectly recognize the critical lure seems to grow with age, whereas the tendency to incorrectly recognize the unrelated or weakly related words does not.

---INSERT TABLE 6 AND FIGURES 7 AND 8 ABOUT HERE---

3.6. Analyses with structural equation modeling

One of the main questions in this study is whether individuals tend to perform better or worse on specific types of memory tasks. Is it the case that, if someone tends to score high on a task that measures pattern memory, this person will also perform well on other types of visuospatial memory? Or is the distinction rather along the time dimension, such that certain individuals perform well on most immediate memory tasks, whereas others on delayed memory tasks?

Complex models of memory and age can be developed, especially if many different types of tests have been administered. Here, we limited the analysis to the major dimensions immediate-versus-delayed memory and verbal-versus-visuospatial memory. We also examined how age and education factor into the resulting models.

We investigated the structure of the test battery using confirmatory factor analysis. This was done by fitting models with the structural equation modeling package for the statistical programming language R (Fox, 2006). Before conducting analyses with structural equation modeling, we analyzed our data with a factor analysis. This yielded a strong, single factor. We suspected that simply the covariation with age caused this (spurious) factor and re-analyzed the data with smaller age bins of ten years. We still found a single dominant factor within each age bin, with varying structures for smaller factors per age bin. We suspected that even within these ten-year age bins, age would still remain a dominant covariate given our rather large sample.

Structure equation modeling (SEM) is a technique that allows us to control for linearly varying factors like age. It assumes linear relationships and cannot handle nonlinearity (e.g., inverted U-shaped curves). Because the results in Figure 4 suggested that age has a nonlinear effect on some tests, only data of participants 25 years and older were used for the analyses to ensure that effects of age would be monotonic. The correlations between age, education and the tests are given in Table 7. The scores on the MemGame-LTM were given in the opposite direction of the scores on the other tests (i.e., a higher score meant a poorer performance), which caused the positive correlation with age and the negative correlation with education. The direction of the scores did not affect the results of the SEM analyses.

The SEM analyses were carried out over all data. Fits were based on the covariance matrix (Table 7) using a pair-wise missing values criterion, which means that each covariance is computed with different participants. To remain conservative, in estimating the goodness-of-fit of the models we used the lowest N with which a covariance was computed: 1631. Most covariances were based on more participants, with the highest N being 8712. Model choice was based on the Bayesian Information Criterion (BIC). To ascertain that the choice of criterion did not overly affect our conclusions, model choice was repeated with Bentler's Comparative Fit Index (CFI) as criterion; this led to the same choices in all cases. We also report chi-square tests of model fit. All fitted models were rejected by this test, but this is not unusual when modeling very large data sets.

We analyzed the set of tests listed in Table 1. There are two dimensions on which the tests differ: modality (verbal vs. visuospatial) and retention interval (immediate memory [STM] vs. delayed memory [LTM]). We first tested a model in which all tests were grouped together (the 'one memory factor' model in Table 8). Next, we tested a model in which tests were grouped on the basis of retention interval ('STM & LTM' in Table 8). We also tested a model in which the tests were grouped on the basis of modality ('verbal & visuospatial memory'). Distinguishing between verbal and visuospatial tests led to a better fit of the data than distinguishing between immediate and delayed memory tests (see Table 8). In fact, once tests were divided into verbal and visual, further division into immediate and delayed memory tests (the 'STM & LTM, spatial & verbal' model in Table 8) led to a worse fit of the data. We can conclude from fits of the models that in our experiment individual differences in memory were dominated by a distinction between verbal and visuospatial memory.

---INSERT TABLES 7 AND 8 ABOUT HERE---

We also looked at the contribution of age to examine whether verbal and visuospatial memory decreased in the same way with age. Our previous analyses suggested that visuospatial memory decreases much faster than verbal memory. This result was confirmed by the SEM analyses. A model with shared parameters for the effect of age on verbal and visuospatial tests was rejected in favor of a model with independent parameters (see Table 8, second group of models), suggesting that the effect of age was different for verbal and visuospatial memory. In fact, the effect of age on visuospatial memory was about three times as large as that on verbal memory tests (lambda 0.026 vs. 0.067, with both memory factors having a variance set to 1). A model in which age affected all tests individually, and not at the level of any superseding factors like verbal versus visuospatial, was also tested. This model performed worse than the one in which age affects visuospatial and verbal tests as a group.

We subsequently looked at the same question for the effect of education by fitting models in which parameters for the effect of age and education were either shared or separated by the visuospatial and verbal memory factors. For education, no difference was found between its effect on visuospatial and verbal memory (see Table 8, third group of models).

We noted above that one test, the DNMT (recent news events), had a very different pattern of correlations with age and education than the other tests. This was also revealed by further modeling. When we added specific parameters for the effect of age and education on the DNMT, fits markedly improved – even more when only a separate education parameter was included ('separate age, shared education, extra for DNMT' model in Table 8). This extra parameter incorporates the stronger effect of education on the DNMT than on any other test (see Figure 9 for a schematic of this final model).

---INSERT FIGURE 9 ABOUT HERE---

4. Discussion

The main result that emerged from our administration of a test battery to more than 28,000 participants via the Internet was a steady decrease of memory performance after the age of 25. The effects of age were different for different types of material. For visuospatial tests, this decrease started at 18 years or even earlier and was twice as fast as the decrease of verbal memory. When expressed in Z-scores, the decrease fitted a straight line with a substantial decrease rate, varying from 2.7% per year for visuospatial memory to 0.9% per year for verbal memory. The highest rate of decrease (3.3% per year, R^2 =98.5%) was for the scores on the Pattern Span Task (Pattern-STM), which showed an early peak at the 16-18 year bin (i.e., having age 17 at its center). An exception to the general pattern of decrease was the Daily News Memory Test (DNMT) which measures memory for news events that occurred up to two years prior to the moment of testing. For the DNMT, we did not find an age-related decrease, replicating our earlier findings (Meeter, et al., 2005; Meeter, et al., 2010). This lack of decrease makes it unlikely that the decrease with age found in the other tests can be attributed primarily to extraneous factors, such as computer skills.

Our study yielded large effects of level of education, where participants who had received vocational education performed about 0.8 standard deviations below those who had received the highest level of secondary schooling, preparing for university. We also found gender effects, though these were small, with women outperforming men on verbal tests and the reverse pattern on visuospatial tests. These patterns have been reported several times in the literature (Herlitz, et al., 1999; Herlitz, Nilsson, & Backman, 1997; Robert & Savoie, 2006).

Analyses with structural equation modeling confirmed the findings above, in particular the notion that individuals tend to differ in visuospatial versus verbal memory, such that if they perform well on a few verbal tests, they tend to perform well on the other verbal tests but not on the visuospatial tests. A similar difference was not found for immediate versus delayed memory. Individuals, who perform well on a few immediate memory tests, do not necessarily perform well on other immediate memory tests.

4.1. The rise and fall of memory

Our results suggest an optimally efficient memory around 16 to 24 years with a steady decrease afterwards. Other authors have reported that the distinction between visuospatial and verbal memory is in place in children from about 4 years (Alloway, et al., 2006), with an increase in performance between the ages 4 and 15 (Alloway, et al., 2006; Gathercole, et al., 2004; Nichelli, Bulgheroni, & Riva, 2001; van Leeuwen, van den Berg, Hoekstra, & Boomsma, 2009). Studies that include the age range around the peak observed here have also found such a pattern (Li, et al., 2004; Swanson, 1999), although the exact location of the peak varies with the tasks. Studies focusing on adults (usually taken as 18 years and older) report a pattern similar to ours with a linear decrease after 20 to 25 years (Logie & Maylor, 2009; Nilsson, 2003; Park, et al., 2002; Ronnlund, Nyberg, Backman, & Nilsson, 2005; Salthouse, 2009).

A large-scale Internet-based study by Logie and Maylor (2009) included two tests similar to ours, the digit span and pattern span task (also see Johnson, et al., 2010, for additional analyses). Like in our study, they also reported that the digit span declines relatively little with age and the pattern span very much (highest decline in both studies). A somewhat different pattern was found for another delayed memory task in a related large-scale Internet-based study by Maylor and Logie (2010). They found that prospective memory (remembering to click a 'smiley' picture later) peaks around 16 to 18 years. They also included a remote memory task in which subjects had to indicate whether a change had earlier occurred in one image of a scene (out of three shown earlier, delayed recall). There did not seem to be a clear peak in the data, which continued to increase slowly until late thirties with a gradual decline afterwards. This pattern resembles more our verbal memory data, which may be because the subjects verbally recoded their earlier encounter with the scenes.

Rapid decrease of memory with age does not occur on all memory tasks, as the results of the DNMT indicate. This test has strong components of crystallized intelligence, such as knowledge-of-the-world and verbal knowledge (as evidenced by the strong effect of education on this test), making the test more similar to a semantic memory test. For semantic memory tests, other studies have found that decrease may occur much later, after about 55 years of age, with a shallower slope (Li, et al., 2004; Nyberg, Backman, Erngrund, Olofsson, & Nilsson, 1996).

4.2. Visuospatial and verbal memory

Our results show a higher rate of decrease with age for the visuospatial tests than for the verbal tests included in this study. These results have also been found by other studies (Bopp & Verhaeghen, 2007; Jenkins, et al., 2000; Logie & Maylor, 2009; Turcotte, et al., 2005), but one study, using a different selection of tests, did not replicate this finding (Park, et al., 2002). The finding that visuospatial and verbal memory are the dominant factors in our structural equation analysis is corroborated by genetic studies. Two studies with adult twins reared apart (McGue & Bouchard, 1989; Pedersen, Plomin, Nesselroade, & McClearn, 1992) have found substantial heritability estimates for memory (43% and 38%, respectively) and for verbal (57% and 58%) and spatial (71% and 46%) abilities. A recent twin study with 12-year-old children and young adults (aged 18) directly investigated the heritability of visuospatial and verbal memory (van Leeuwen, et al., 2009). For the 12-year-old children, heritability estimates of 25% for visuospatial immediate memory and 48% for verbal memory were reported. Estimates for the 18-year-olds were 43% and 42%. Visuospatial and verbal memory, thus, show independent—though not completely uncorrelated—patterns of heritability in the order of 40% or higher.

4.3. False memory and age

The DRM paradigm (Deese, 1959; Roediger & McDermott, 1995) can induce false memories for critical lures, which is often used to study memory distortion in the laboratory. Like several other studies (Koutstaal & Schacter, 1997; Koutstaal, Schacter, Galluccio, & Stofer, 1999; Lovden, 2003; Schacter, et al., 1999; Schacter, Koutstaal, & Norman, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998), we found that older participants are more likely to falsely recognize the critical lure than younger participants, though the effect was small. It is not entirely clear what promotes this effect of age (Henkel, Johnson, & De Leonardis, 1998; Reyna & Lloyd, 1997; Schacter, et al., 1999). The effect of age could not be explained by an overall impairment in word recognition. Studied words, non-studied weakly related words, and non-studied unrelated words were not recognized differently at later ages. Detailed experimentation and analyses show that the false memory effect itself can be decomposed into several sub-processes, notably semantic activation of the critical lure and failure of strategic monitoring processes (Roediger, Watson, McDermott, & Gallo, 2001; Unsworth & Brewer, 2010), whereas another study emphasized episodic memory (Lovden, 2003) as an important factor.

Though the age effect with respect to critical lures seems clear, we did not find corroborating evidence that individuals differed consistently in proneness to recall the gist (critical lure) rather than the details (Salthouse & Siedlecki, 2007). Our test battery included another test that emphasized ‘holistic’ memory, the Story-LTM study, which asked for the gist of the story rather than the exact phrasing. Further analyses (not reported above) with structural equation models yielded, however, no evidence that proneness to ‘gist memory’ or ‘holistic memory’ played a role as an individual difference factor.

4.4. Conclusions

The main finding in this study was that individual differences in memory are dominated by the verbal-versus-visuospatial modality dimension. Individuals are thus best characterized in terms of having a good verbal memory or a good visuospatial memory. We did not find evidence for such a distinction between immediate and delayed memory. After the age of 25, memory gets 1% to 3% worse each year (in percentage of standard deviations) with immediate and delayed memory following nearly identical patterns of decrease. There were large differences between visuospatial versus verbal memory. For visuospatial memory, the decrease sets in earlier, around the age of 18, and is almost twice as steep as the decrease of verbal memory. We also found an increasing proneness to report false memories with age.

Author Note

Jaap M. J. Murre, Department of Psychology, University of Amsterdam; Steve M. J. Janssen, Department of Psychology and Neuroscience, Duke University; Romke Rouw, Department of Psychology, University of Amsterdam; Martijn Meeter, Department of Psychology, VU University Amsterdam.

Steve M. J. Janssen is now at the School of Psychology, Flinders University.

We would like to thank Wouter Pasman for his work on creating the software for the test battery. The first, second and fourth authors all benefited from grants from the Netherlands Organization for Scientific Research (NWO; an PIONIER grant to JM, and grants numbered 446-06-031 to SJ, and 451-05-006 to MM).

Correspondence concerning this manuscript should be addressed to: Jaap Murre, University of Amsterdam, Department of Psychology, Weesperplein 4, 1018 XA, Amsterdam, the Netherlands. E-mail address: jaap@murre.com.

References

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuospatial short-term and working memory in children: Are they separable? *Child Development, 77*(6), 1698-1716.

Ardila, A., & Rosselli, M. (1994). Development of language, memory, and visuospatial abilities in 5-year-old to 12-year-old children using a neuropsychological battery. *Developmental Neuropsychology, 10*(2), 97-120.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory (Vol. 2).* New York:: Academic Press.

Baddeley, A. D. (2003). Working memory: looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839. doi: 10.1038/nrn1201

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). New York: Academic press.

Bopp, K. L., & Verhaeghen, P. (2007). Age-related differences in control processes in verbal and visuospatial working memory: Storage, transformation, supervision, and coordination. *Journals of Gerontology Series B-Psychological Sciences and Social Sciences, 62*(5), P239-P246.

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.* Ph.D. Thesis, McGill University, Montreal.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. doi: 10.1016/S0022-5371(80)90312-6

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58*(1), 17-22.

Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo–spatial memory. *Neuropsychologia, 37*(10), 1189-1199.

Duff, S. J., & Hampson, E. (2001). A sex difference on a novel spatial working memory task in humans. *Brain and Cognition, 47*(3), 470-493. doi: 10.1006/brcg.2001.1326

Fox, J. (2006). Structural Equation Modeling with the sem Package in R. *Structural Equation Modeling, 13*, 465-486.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40*(2), 177-190. doi: 10.1037/0012-1649.40.2.177

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist, 59*(2), 93-104. doi: 10.1037/0003-066x.59.2.93

Henkel, L. A., Johnson, M. K., & De Leonardis, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General, 127*(3), 251-268.

Herlitz, A., Airaksinen, E., & Nordstrom, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology, 13*(4), 590-597.

Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*(6), 801-811.

Janssen, S. M. J., Chessa, A. G., & Murre, J. M. J. (2005). The reminiscence bump in autobiographical memory: Effects of age, gender, education, and culture. *Memory, 13*(6), 658-668. doi: 10.1080/09658210444000322

Janssen, S. M. J., Chessa, A. G., & Murre, J. M. J. (2006). Memory for time: How people date events. *Memory & Cognition, 34*(1), 138-147. doi: 10.3758/BF03193393

Janssen, S. M. J., Chessa, A. G., & Murre, J. M. J. (2007). Temporal distribution of favourite books, movies, and records: Differential encoding and re-sampling. *Memory, 15*(7), 755-767. doi: 10.1080/09658210701539646

Janssen, S. M. J., & Murre, J. M. J. (2008). Reminiscence bump in autobiographical memory: Unexplained by novelty, emotionality, valence, or importance of personal events. *Quarterly Journal of Experimental Psychology, 61*(12), 1847-1860. doi: 10.1080/17470210701774242

Janssen, S. M. J., Murre, J. M. J., & Meeter, M. (2008). Reminiscence bump in memory for public events. *European Journal of Cognitive Psychology, 20*(4), 738-764. doi: 10.1080/09541440701554409

Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology and Aging, 15*(1), 157-175. doi: 10.1037//0882-7974.15.1.157

Johnson, W., Logie, R. H., & Brockmole, J. R. (2010). Working memory tasks differ in factor structure across age cohorts: Implications for dedifferentiation. *Intelligence, 38*(5), 513-528. doi: 10.1016/j.intell.2010.06.005

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The corsi block-tapping task: standardization and normative data. *Applied Neuropsychology, 7*(4), 252-258.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*(4), 555-583.
Koutstaal, W., Schacter, D. L., Galluccio, L., & Stofer, K. A. (1999). Reducing gist-based false recognition in older adults: Encoding and retrieval manipulations. *Psychology and Aging, 14*(2), 220-237.
Kristo, G., Janssen, S. M. J., & Murre, J. M. J. (2009). Retention of autobiographical memories: An Internet-based diary study. *Memory, 17*(8), 816-829. doi: 10.1080/09658210903143841
Li, S.-C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P., B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science, 15*(3), 155-163.
Logie, R. H., & Maylor, E. A. (2009). An Internet study of prospective memory across adulthood. *Psychology and Aging, 24*(3), 767-774. doi: 10.1037/a0015479
Lovden, M. (2003). The episodic memory and inhibition accounts of age-related increases in false memories: A consistency check. *Journal of Memory and Language, 49*(2), 268-283. doi: 10.1016/s0749-596x(03)00069-x
Maylor, E. A., & Logie, R. H. (2010). A large-scale comparison of prospective and retrospective memory development from childhood to middle age. *The Quarterly Journal of Experimental Psychology, 63*(3), 442-451. doi: 10.1080/17470210903469872
McGue, M., & Bouchard, T. J. (1989). Genetic and environmental determinants of information-processing and special mental abilities - A twins reared apart analysis. *Behavior Genetics, 19*(6), 768-768.
Meeter, M., Murre, J. M. J., & Janssen, S. M. J. (2005). Remembering the news: Modeling retention data from a study with 14,000 participants. *Memory & Cognition, 33*(5), 793-810. doi: 10.3758/BF03193075
Meeter, M., Ochtman, D. J. C., Janssen, S. M. J., & Murre, J. M. J. (2010). Of sports and politics: Predicting category-specific retention of news events from demographic variables. *European Journal of Cognitive Psychology, 22*(1), 117-129. doi: 10.1080/09541440802708037
Miller, G. A. (1956/1994). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 101*(2), 343-352. doi: 10.1037/0033-295X.101.2.343
Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*, 272-277.
Murre, J. M. J., Wolters, G., & Raffone, A. (2006). Binding in working memory and long-term memory: towards an integrative model. In Z. Zimmer, A. Mecklinger & U. U. Lindenberger (Eds.), *Handbook of binding and memory: Perspectives from cognitive neuroscience.* (pp. 221-250). Oxford: Oxford University Press.
Nichelli, F., Bulgheroni, S., & Riva, D. (2001). Developmental patterns of verbal and visuospatial spans. *Neurological Sciences, 22*(5), 377-384.
Nilsson, L. G. (2003). Memory function in normal aging. *Acta Neurologica Scandinavica, 107*, 7-13.
Nyberg, L., Backman, L., Erngrund, K., Olofsson, U., & Nilsson, L. G. (1996). Age differences in episodic memory, semantic memory, and priming: relationships to demographic, intellectual, and biological factors. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 51*(4), 234-240.
Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*(2), 299-320. doi: 10.1037//0882-7974.17.2.299
Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic-analysis of cognitive-abilities during the second half of the life-span. *Psychological Science, 3*(6), 346-353.
Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology: General, 58*, 193-198.
Reips, U. D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbaum (Ed.), *Psychological experiments on the Internet* (pp. 89-114). San Diego, CA: Academic Press.
Reips, U. D. (2002). Standards for Internet-based experimenting. *Experimental Psychology, 49*(4), 243-256. doi: 10.1027//1618-3169.49.4.243
Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*(2), 95-123.
Robert, M., & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working-memory resources? *European Journal of Cognitive Psychology, 18*(3), 378-397. doi: 10.1080/09541440500234104
Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning Memory and Cognition, 21*(4), 803-814.
Roediger, H. L., Watson, J. M., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review, 8*(3), 385-405.

Ronnlund, M., Nyberg, L., Backman, L., & Nilsson, L. G. (2005). Stability, growth, and decline in adult life span development of declarative memory: Cross-sectional and longitudinal data from a population-based study. *Psychology and Aging, 20*(1), 3-18. doi: 10.1037/0882-7974.20.1.3

Salthouse, T. A. (2009). When does age-related cognitive decline begin? [Article]. *Neurobiology of Aging, 30*(4), 507-514. doi: 10.1016/j.neurobiolaging.2008.09.023

Salthouse, T. A., & Siedlecki, K. L. (2007). An individual difference analysis of false recognition. *American Journal of Psychology, 120*(3), 429-458.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*(1), 1-24.

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences, 1*(6), 229-236.

Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. [Review]. *Annual Review of Psychology, 57*, 529-555. doi: 10.1146/annurev.psych.57.102904.190048

Spaan, P. E. J., Raaijmakers, J. G. W., & Jonker, C. (2005). Early assessment of dementia: The contribution of different memory components. *Neuropsychology, 19*, 629-640.

Swanson, H. L. (1999). What develops in working memory? A life span perspective. *Developmental Psychology, 35*(4), 986-1000.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13*(2), 230-241.

Turcotte, J., Gagnon, S., & Poirier, M. (2005). The effect of old age on the learning of supraspan sequences. *Psychology and Aging, 20*(2), 251-260. doi: 10.1037/0882-7974.20.2.251

Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. [Article]. *Journal of Memory and Language, 62*(1), 19-34. doi: 10.1016/j.jml.2009.08.002

van Leeuwen, M., van den Berg, S. M., Hoekstra, R. A., & Boomsma, D. I. (2009). The genetic and environmental structure of verbal and visuospatial memory in young adults and children. *Neuropsychology, 23*(6), 792-802. doi: 10.1037/a0016526

Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature, 438*(7067), 500-503.

Table 1.a. Memory tests in the battery.

Immediate Memory Tests	
	Verbal
10WT-STM	Ten-Words Test
DRM-STM	Deese-Roediger-McDermott Immediate Recall
Story-STM	Story Telling Immediate Recognition
Digit-STM	Digit Span Task
	Visuospatial
Pattern-STM	Pattern Span Task
Corsi-STM	Corsi Block Tapping Task
Texture-STM	Texture Span Task
Delayed Memory Tests	
	Verbal
DRM-LTM	Deese-Roediger-McDermott Delayed Recognition
Story-LTM	Story Telling Delayed Recognition
	Visuospatial
MemGame-LTM	Memory Game
	Other
DNMT	Daily News Memory Test

Table 1.b. Number of tests taken per five-year age group (after removing tests below chance).

	Immediate Memory Tests							**Delayed Memory Tests**			
	Verbal				*Visuospatial*			*Verbal*		*Visuospatial*	*Other*
Age Bin	10WT-STM	DRM-STM	Story-STM	Digit-STM	Pattern-STM	Corsi-STM	Texture-STM	DRM-LTM	Story-LTM	MemGame-LTM	DNMT
11-15	349	420	488	321	480	399	304	266	292	336	84
16-20	1451	1703	1666	1282	1661	1447	1159	1175	1141	1413	501
21-25	1394	1563	1404	1167	1448	1317	1006	1098	1058	1282	568
26-30	1064	1220	1098	888	1072	1013	792	845	850	962	494
31-35	814	943	846	720	780	788	628	663	678	742	451
36-40	874	961	941	803	840	818	705	697	769	779	487
41-45	1010	1083	1075	918	940	933	746	754	891	866	675
46-50	1111	1206	1257	1040	1012	1051	860	807	1035	922	869
51-55	1084	1238	1283	1039	1026	1065	806	773	1067	888	921
56-60	759	875	908	733	677	759	525	536	759	606	668
61-65	391	482	456	387	330	361	239	280	391	322	401
66-70	215	257	243	218	189	204	106	140	207	164	223
71-75	144	163	169	124	108	121	53	69	132	80	137
76-80	70	87	88	61	61	78	22	39	65	53	67
Total	**10730**	**12201**	**11922**	**9701**	**10624**	**10354**	**7951**	**8142**	**9335**	**9415**	**6546**

Table 2. For each test, the mean performance (*Mean*), standard deviation (*SD*), standard error (*SE*) and number of times it was completed (*N*) are listed.

	Mean	*SD*	*SE*	*N*
	Verbal STM			
10WT-STM	20.14	5.37	0.0518	10730
DRM-STM	29.80	7.47	0.0676	12201
Story-STM	7.07	1.29	0.0118	11922
Digit-STM	6.04	1.39	0.0141	9701
	Visuospatial STM			
Pattern-STM	195.91	61.04	0.5922	10624
Corsi-STM	5.35	0.73	0.0072	10354
Texture-STM	3.56	1.41	0.0158	7951
	Verbal LTM			
DRM-LTM	69.63	7.12	0.0789	8142
Story-LTM	7.90	1.29	0.0134	9335
	Visuospatial LTM			
MemGame-LTM	98.09	39.96	0.4118	9415
	Other LTM			
DNMT	14.33	4.33	0.0536	6546

Table 3. F-values for main effects in the analysis of variance of each test.

	Gender	*Education*	*Age Group*
		Verbal STM	
10WT-STM	78.238***	23.920***	33.793***
DRM-STM	30.636***	39.959***	17.791***
Story-STM	0.654	3.852***	5.281***
Digit-STM	4.909*	16.193***	3.732***
		Visuospatial STM	
Pattern-STM	7.061**	10.982***	93.324***
Corsi-STM	31.546***	6.602***	27.671***
Texture-STM	0.005	7.005***	14.229***
		Verbal LTM	
DRM-LTM	19.333***	17.462***	4.695***
Story-LTM	0.872	5.180***	1.482
		Visuospatial LTM	
MemGame-LTM	96.713***	22.419***	57.371***
		Other LTM	
DNMT	40.075***	26.601***	6.779***

*$p < .05$, **$p < .01$, ***$p < .001$

Table 4. F-values for interactions in the analysis of variance of each test.

	*Gender*Age*	*Age*Edu*	*Gender*Edu*	*Gender*Edu*Age*
		Verbal STM		
10WT-STM	1.568	1.761***	1.412	1.188
DRM-STM	1.342	1.832***	1.201	1.038
Story-STM	0.756	1.076	0.808	1.098
Digit-STM	1.600	1.092	0.541	0.978*
		Visuospatial STM		
Pattern-STM	1.846*	1.337*	1.227	1.008
Corsi-STM	1.258	1.096	1.257	1.289*
Texture-STM	1.130	1.113	0.143	0.859
		Verbal LTM		
DRM-LTM	0.977	1.488**	0.691	1.103
Story-LTM	1.113	1.074	1.124	1.221
		Visuospatial LTM		
MemGame-LTM	2.710**	2.158***	2.302*	1.837***
		Other LTM		
DNMT	2.121*	1.516**	1.376	1.173

*$p < .05$; **$p < .01$; ***$p < .001$

Table 5. For each test, using linear regression in the age range 42 to 70 years, the slope, intercept, r^2 and average value (i.e., for age range 42-70) was calculated.

	Slope	*Intercept*	r^2	*Average*
	Verbal STM			
10WT-STM	-0.0294	1.278	.883	-0.346
DRM-STM	-0.0177	0.775	.828	-0.201
Story-STM	-0.0072	0.309	.613	-0.085
Digit-STM	-0.0070	0.312	.502	-0.076
	Visuospatial STM			
Pattern-STM	-0.0309	1.132	.959	-0.570
Corsi-STM	-0.0204	0.829	.851	-0.300
Texture-STM	-0.0230	0.924	.972	-0.343
	Verbal LTM			
DRM-LTM	-0.0043	0.152	.120	-0.082
Story-LTM	-0.0093	0.465	.772	-0.051
	Visuospatial LTM			
MemGame-LTM	-0.0381	1.606	.958	-0.498
	Other LTM			
DNMT	0.0042	-0.129	.313	0.101

Table 6. The proportion of words from the DRM-LTM test classified as surely old, probably old, probably new, and surely new.

	Surely Old	Probably Old	Probably New	Surely New
Studied Words	.881	.053	.021	.045
Non-studied Words				
Unrelated	.010	.005	.019	.966
Weakly Related	.081	.090	.119	.710
Critical Lure	.539	.133	.089	.240

Table 7. Correlations (above diagonal) and variances/covariances (on/below diagonal) between age, education and the individual tests for participants who were older than 25 years.

	Age	Education	DRM-STM	Digit-STM	Story-STM	Pat-STM	Corsi-STM	Texture-STM	10WT-STM	DRM-LTM	Story-LTM	DNMT	MemGame-LTM
Age	252.1	.162**	-.176**	-.071**	-.095**	-.463**	-.248**	-.249**	-.270**	-.078**	-.046**	.139**	.371**
Education	5.88	5.26	.191**	.152**	.027**	.052**	.083**	.093**	.147**	.148**	.127**	.286**	-.094**
DRM-STM	-20.56	3.23	55.5	.251**	.149**	.244**	.185**	.289**	.540**	.434**	.155**	.155**	-.346**
Digit-STM	-1.54	0.47	2.49	1.92	.060**	.200**	.177**	.220**	.227**	.144**	.088**	.063**	-.180**
Story-STM	-1.93	0.08	1.40	0.11	1.66	.086**	.061**	.105**	.129**	.127**	.099**	.040*	-.136**
Pattern-STM	-436.7	7.2	111.3	16.5	6.81	3703.2	.309**	.274**	.268**	.156**	.106**	.033	-.360**
Corsi-STM	-2.83	0.14	1.01	0.17	0.06	13.48	0.53	.188**	.186**	.125**	.077**	.079**	-.246**
Texture-STM	-5.34	0.30	2.96	0.43	0.19	24.11	0.20	2.06	.266**	.179**	.091**	-.018	-.315**
10WT-STM	-22.5	1.80	20.6	1.65	0.88	85.3	0.71	1.96	28.6	.314**	.152**	.077**	-.349**
DRM-LTM	-8.45	2.38	22.6	1.36	1.16	67.9	0.64	1.83	11.5	50.8	.122**	.085**	-.236**
Story-LTM	-0.93	0.37	1.47	0.16	0.16	8.37	0.07	0.17	1.04	1.11	1.67	.132**	-.136**
DNMT	9.31	2.67	4.61	0.36	0.22	8.3	0.24	-0.12	1.71	2.42	0.72	18.8	-.053*
MemGame-LTM	227.3	-8.6	-102.1	-9.5	-7.03	-853.9	-7.10	-17.2	-71.7	-66.1	-7.00	-9.1	1674

$^{*}p < .05$; $^{**}p < .01$.

Table 8. Summary of structural equation models fit on the data. For each model, the number of parameters (# param) is given and the Goodness of Fit index (GoF), which is comparable to R^2. The Bayesian Information Criterion (BIC) of the best fit and Bentler's Comparative Fit Index (CFI) are reported as well, as the results of a chi-square null hypothesis test against the null model.

Data	Model description	# param	GoF	BIC	CFI	Test against null model
10 STM / LTM tests	One memory factor	20	0.965	-2.3	0.893	$\chi^2(35)=256.6$, $p<.001$
	STM & LTM	21	0.979	-0.9	0.896	$\chi^2(34)=250.6$, $p<.001$
	Verbal & visuospatial memory	21	0.985	**-134.9**	**0.960**	$\chi^2(34)=116.6$, $p<.001$
	STM & LTM, visuospatial and verbal	22	0.948	255.1	0.776	$\chi^2(33)=499.2$, $p<.001$
	STM & LTM, visuospatial & verbal, and semantic	23	0.984	238.76	0.787	$\chi^2(32)=475.4$, $p<.001$
Age	Shared influence age on verbal & spatial	23	0.966	-4.6	0.892	$\chi^2(43)=313.4$, $p<.001$
	Separate for verbal and visuospatial	**24**	**0.979**	**-123.1**	**0.942**	$\boldsymbol{\chi^2(42)=187.5}$**, **$\boldsymbol{p<.001}$
	Separate for each test	32	0.958	175.9	0.844	$\chi^2(34)=427.4$, $p<.001$
Age & Education	**Separate age-, shared education influence**	26	0.973	**-113.7**	**0.921**	$\chi^2(52)=270.9$, $p<.001$
	Separate age and education influence	27	0.973	-106.6	0.921	$\chi^2(51)=270.7$, $p<.001$
	Separate age-, shared education, extra for DNMT	27	0.979	**-172.1**	**0.944**	$\chi^2(51)=205.1$, $p<.001$

Figure Captions

Figure 1. Distribution of number of male and female participants as a function of age.

Figure 2. Averaged Z-transformed scores for male and female participants for verbal and visuospatial tests.

Figure 3. Averaged Z-transformed scores as a function of age and increasing levels of education. LBO, MAVO, HAVO and VWO are increasing levels of secondary education, MBO, HBO and WO of tertiary education. In this graphs immediate and delayed tests have been combined. (a) Averages taken over verbal tests. (b) Averages taken over visuospatial tests. Participants who completed primary school only are not included here (e.g., combined with LBO category), because primary school only is not predictive of intelligence for elderly participants. Participants over 70 have not been included, because of the small numbers per education category, which makes the graphs past 70 years quite erratic and hard to read.

Figure 4. Averaged Z-transformed scores as a function of age for verbal and visuospatial tests.

Figure 5. Averaged Z-transformed scores as a function of age for immediate and delayed tests.

Figure 6. Grand average of Z-transformed scores of all tests and averaged Z-transformed score on the Daily News Memory Test (DNMT) as a function of age (two-year bins).

Figure 7. Averaged Z-transformed scores for the DRM Immediate Recall and Delayed Recognition.

Figure 8. Proportion correct for the four types of stimuli in the DRM Delayed Recognition test as a function of age group: Not-studied unrelated words (squares with dotted line), studied words (triangles with dashed line), not-studied weakly related words (squares with dashed line) and not-studied critical lures (diamonds with solid line). (Participants age > 75 have not been included because of very small sample sizes.)

Figure 9. Final SEM model as described at the bottom of Table 7. Tests and other measurements are given by squares, circles denote latent variables. Latent variables were standardized, but measurements were not.

Figure 1

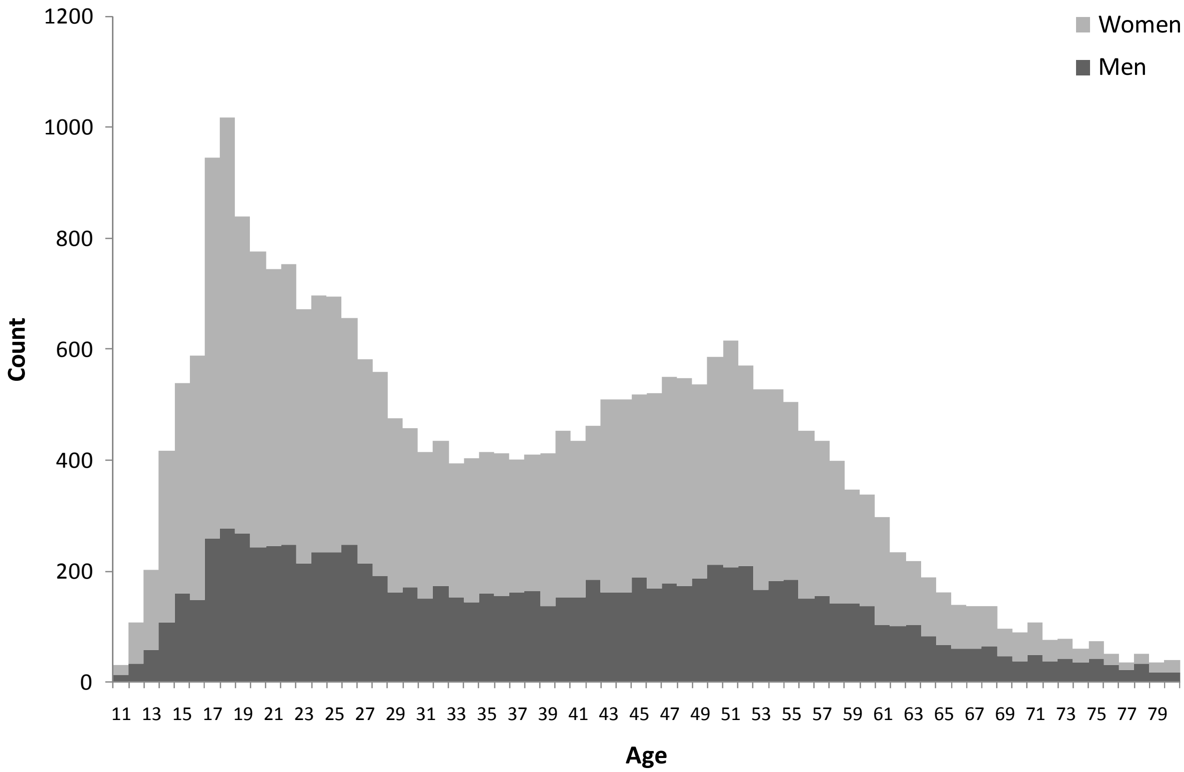
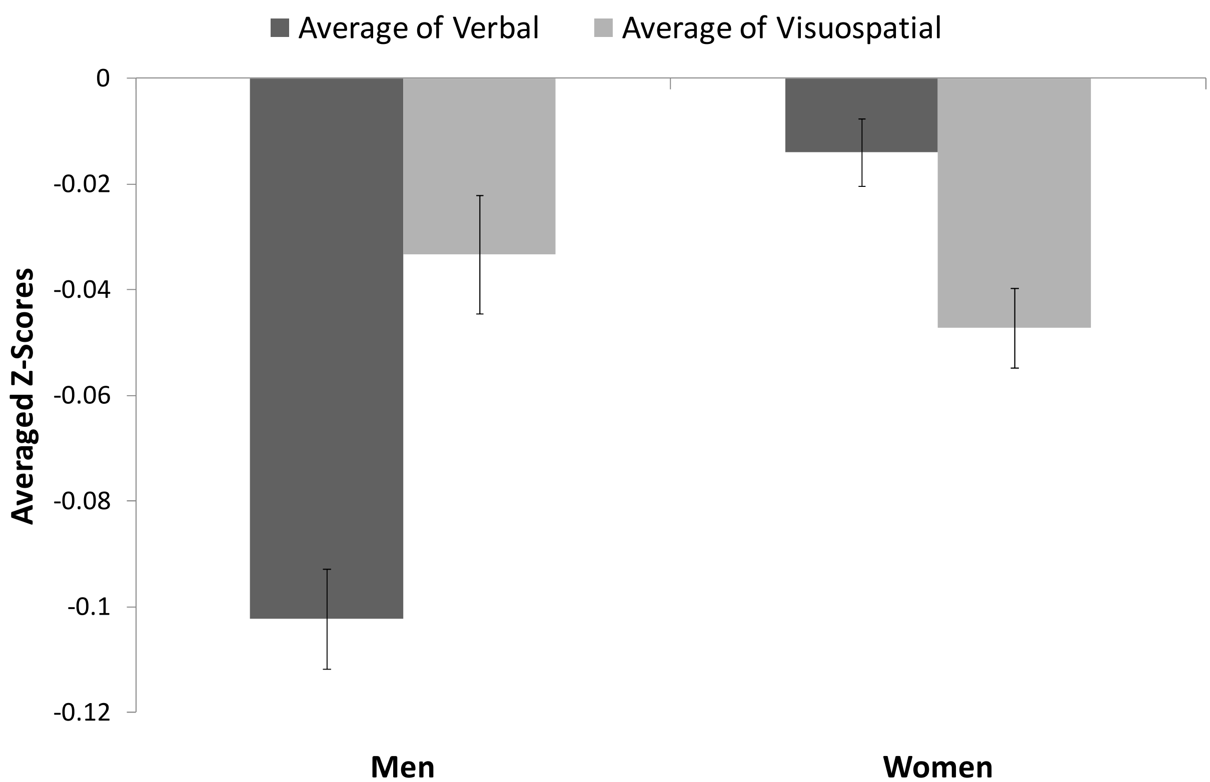

Women
Men
Count
1200
1000
800
600
400
200
0
11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49 51 53 55 57 59 61 63 65 67 69 71 73 75 77 79
Age

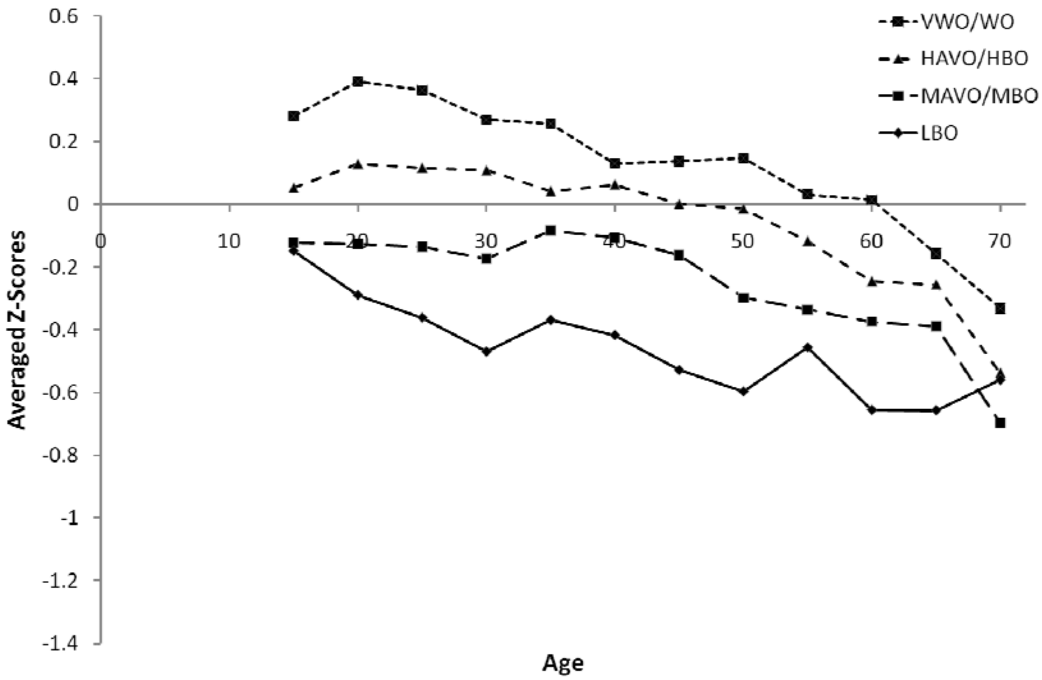
Figure 2



Average of Verbal
Average of Visuospatial
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
Averaged Z-Scores
Men
Women

Figure 3

(a)


0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70
Averaged Z-Scores
Age
VWO/WO
HAVO/HBO
MAVO/MBO
LBO

(b)

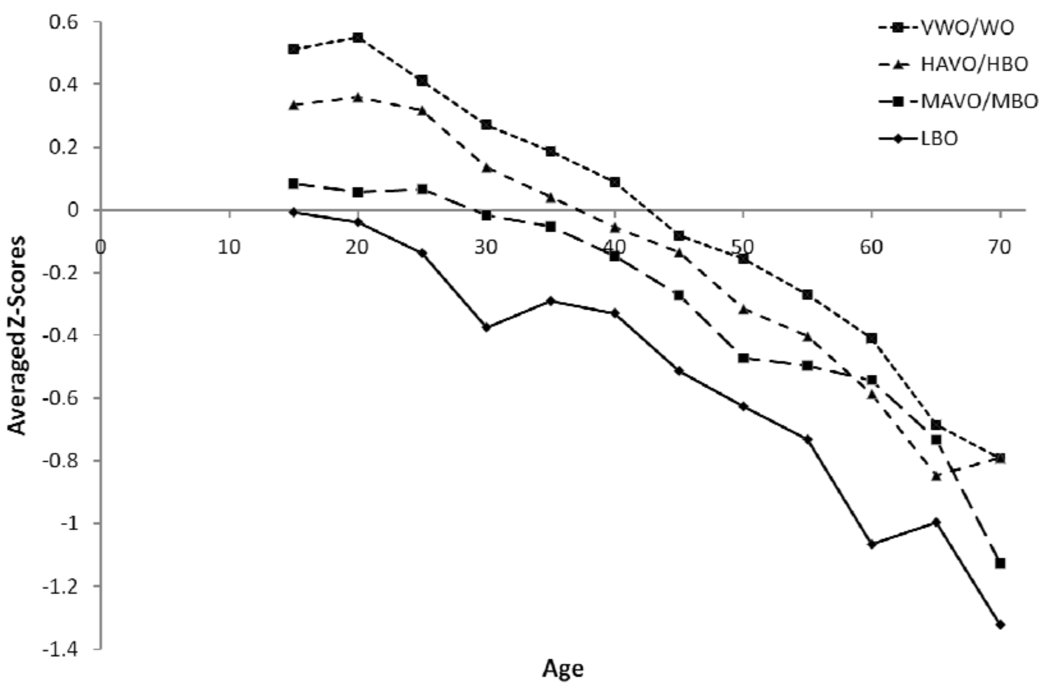
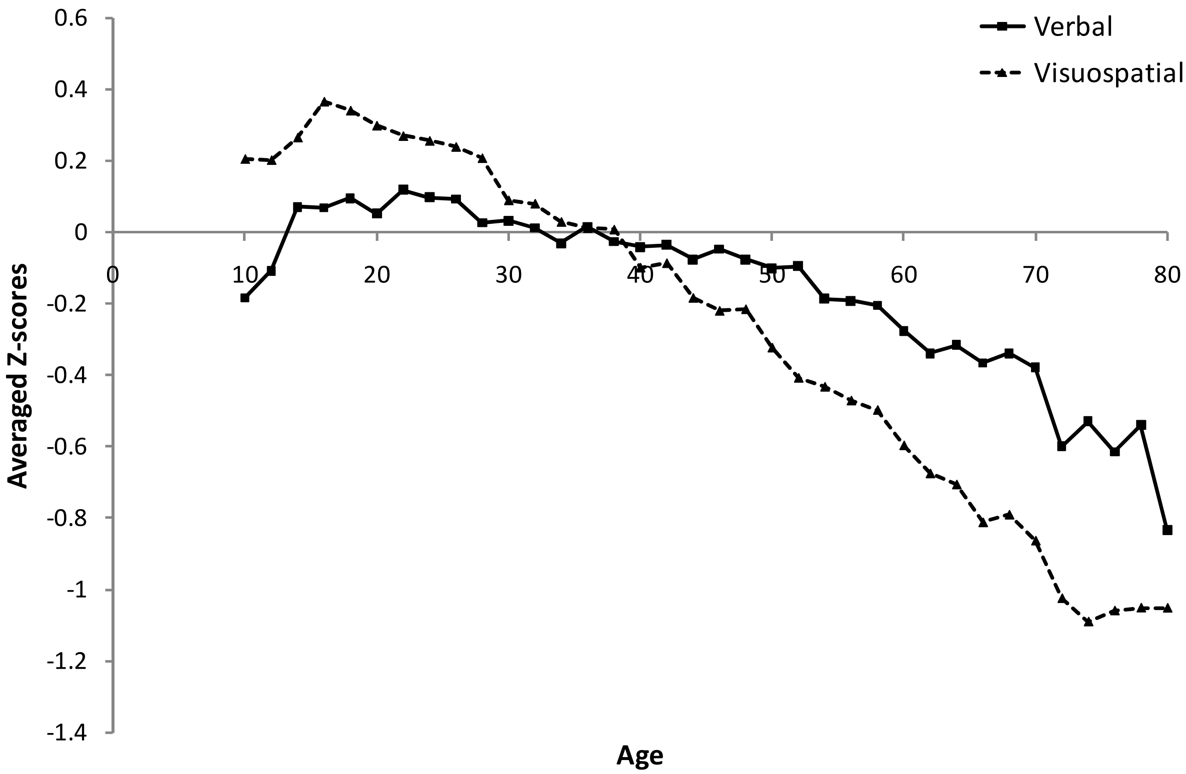

0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70
Averaged Z-Scores
Age
VWO/WO
HAVO/HBO
MAVO/MBO
LBO

Figure 4



Verbal
Visuospatial
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70
80
Averaged Z-scores
Age

Figure 5

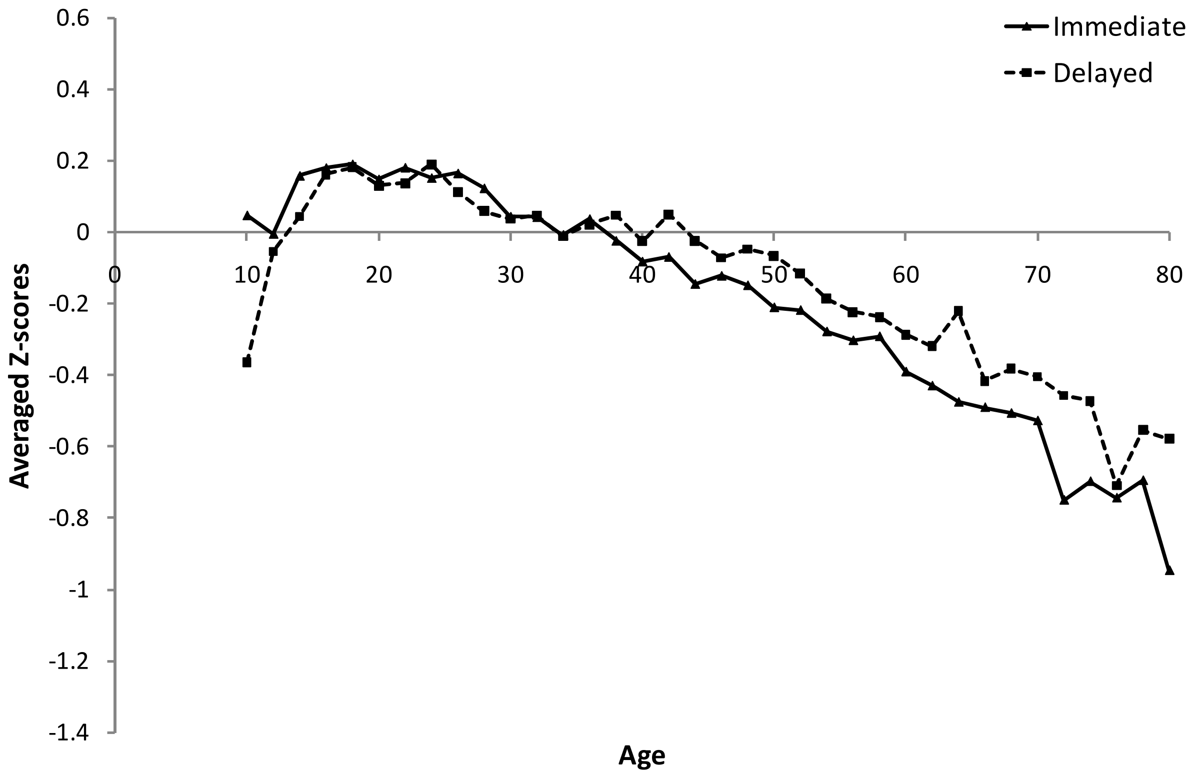

Immediate
Delayed
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70
80
Averaged Z-scores
Age

Figure 6

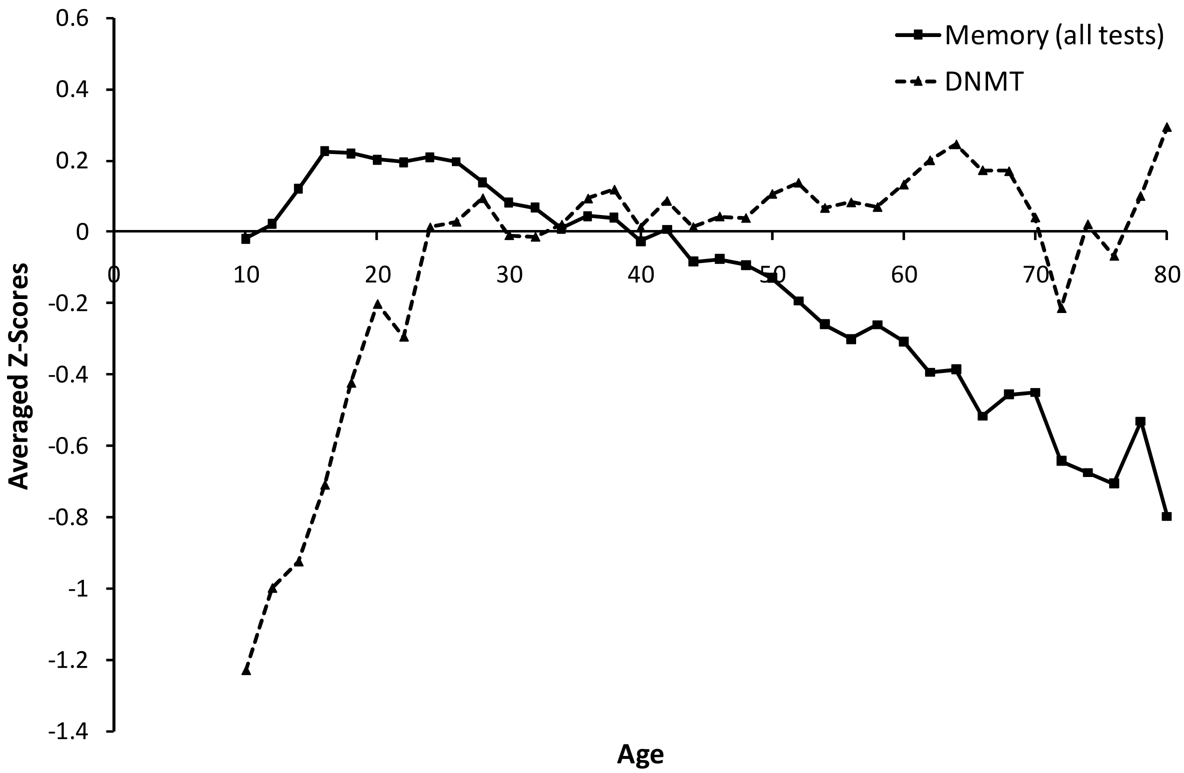
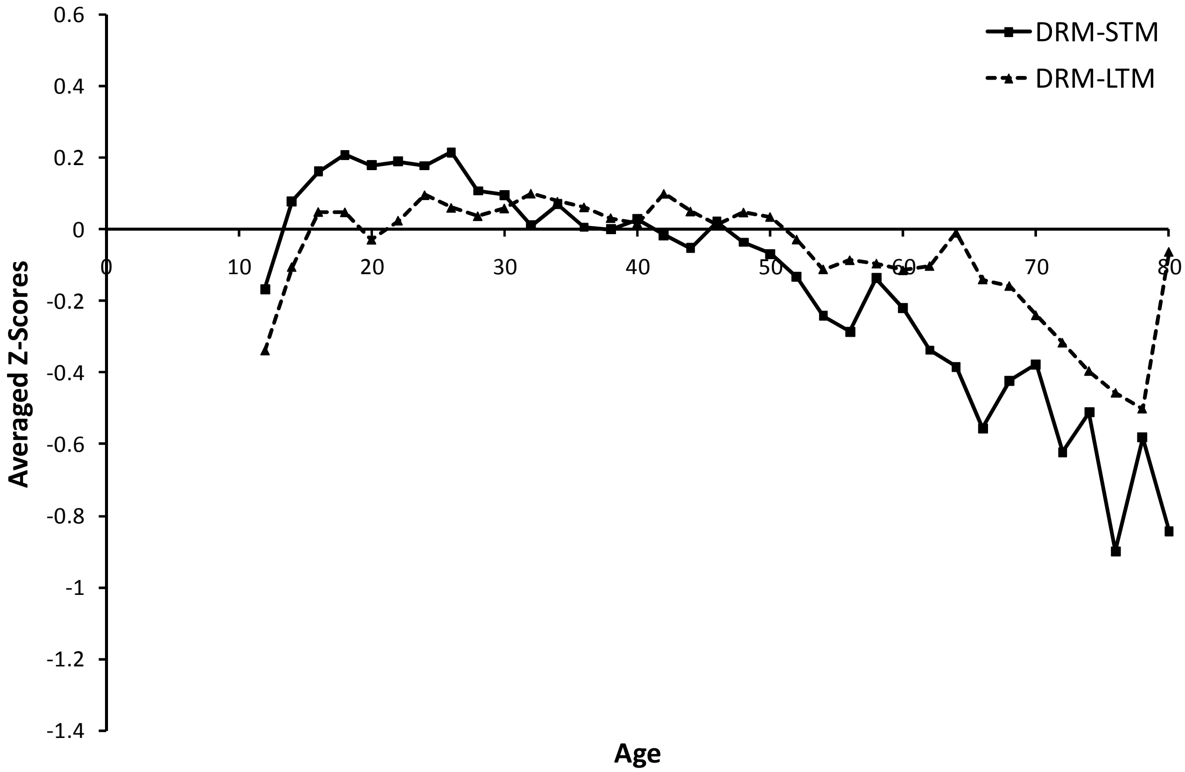
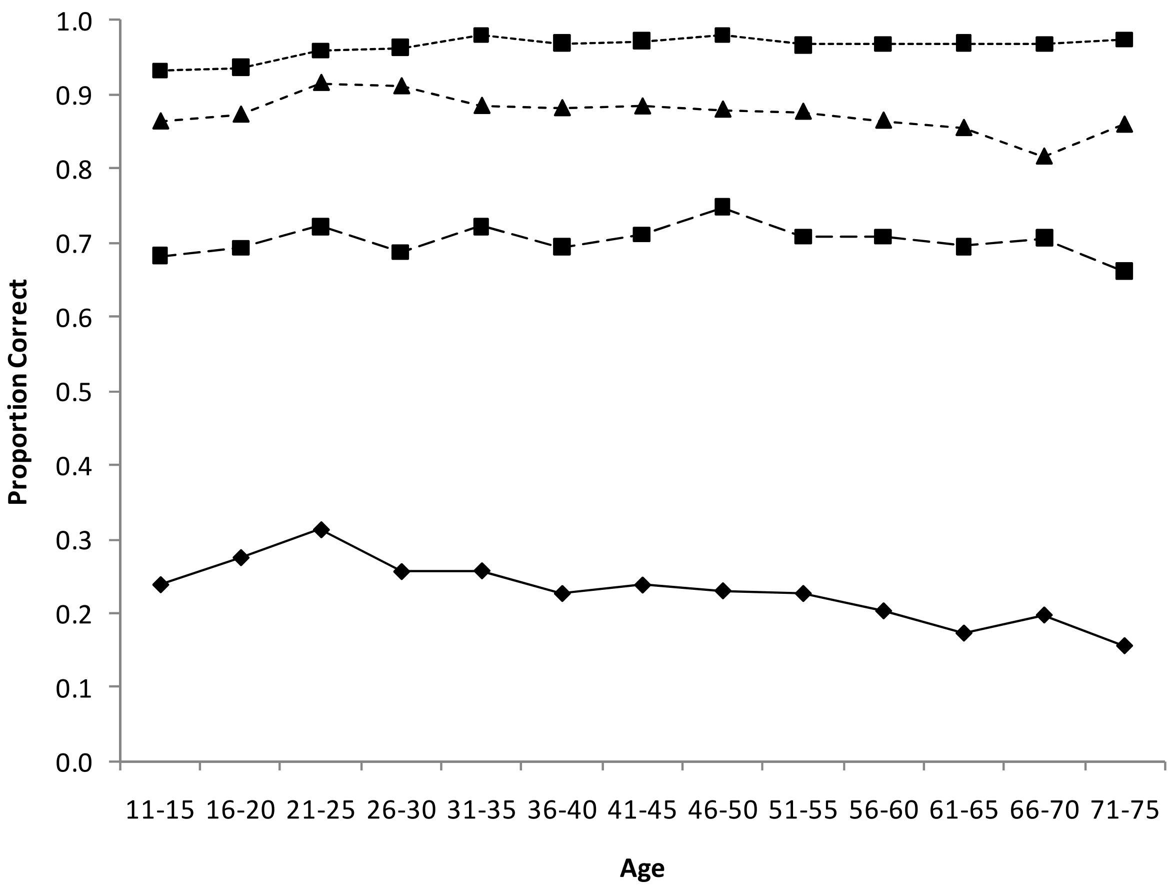

Memory (all tests)
DNMT
Averaged Z-Scores
Age
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70
80

Figure 7



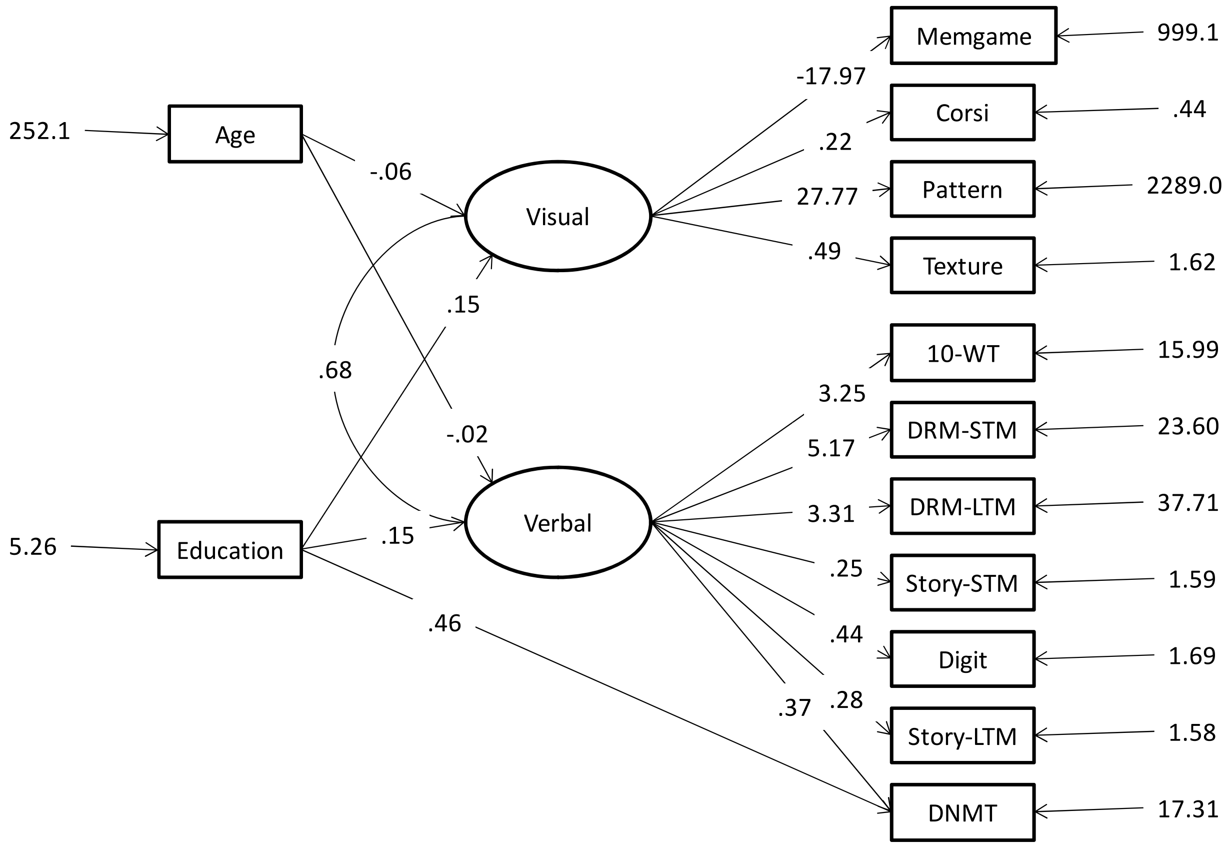
DRM-STM
DRM-LTM
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
Averaged Z-Scores
0
10
20
30
40
50
60
70
80
Age

Figure 8



1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Proportion Correct
11-15
16-20
21-25
26-30
31-35
36-40
41-45
46-50
51-55
56-60
61-65
66-70
71-75
Age

Figure 9


Age
252.1
-.06
.15
Visual
Education
5.26
.15
.46
.68
-.02
Verbal
Memgame
999.1
-17.97
Corsi
.44
.22
Pattern
2289.0
27.77
Texture
1.62
.49
10-WT
15.99
3.25
DRM-STM
23.60
5.17
DRM-LTM
37.71
3.31
Story-STM
1.59
.25
Digit
1.69
.44
Story-LTM
1.58
.28
DNMT
17.31
.37