Running head: RISE AND FALL OF IMMEDIATE AND DELAYED MEMORY

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

Correspondence concerning this manuscript should be addressed to:

Prof. dr. Jaap Murre

University of Amsterdam

Department of Psychology

Roetersstraat 15

1018 WB Amsterdam

The Netherlands

E-mail address: jaap@murre.com

Telephone number: +31-20-5256722

Fax number: +31-20-6391656

Abstract

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 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.

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

1. Introduction

With an Internet-based study in which more than 28,000 volunteers participated, we investigated how memory first rises and then decreases with age. Our research focused on individual differences in immediate versus delayed memory and in verbal versus visuospatial memory. We examined 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. We also looked at which of these two dimensions characterized the individual differences in the population better. The large and varied sample, which spanned the ages from 11 to 80, allowed us to carry out more detailed analyses than some of the previous studies have done that focused on these research questions.

Since it is well known that age, education and gender interact with memory performance, the effects of these 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. While some studies have found a higher rate of decrease with age for the visuospatial tests than for the verbal tests (e.g., Bopp & Verhaeghen, 2007; Jenkins, Myerson, Joerding, & Hale, 2000; Turcotte, Gagnon, & Poirier, 2005), other studies did not replicate this finding (Park, Lautenschlager, Hedden, Davidson, Smith, & Smith, 2002). Furthermore, we wanted to unravel more of the age pattern of decrease. Recent studies (e.g., Lovden, 2003; Schacter, Israel, & Racine, 1999) suggested, for example, that people become more prone to false memories with increasing age. We, therefore, also included a test in the online battery to measure false memories to see whether we could find further evidence for this.

Internet-based testing has clear limitations, though many of these can be circumvented with the correct precautions (e.g., 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 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.

In order 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., Friedman & Janssen, 2010a, 2010b; Janssen, Chessa, & Murre, 2005, 2006, 2007; Janssen & Murre, 2008; Janssen, Murre, & Meeter, 2008; Janssen & Rubin, in press; Janssen, Rubin, & St. Jacques, 2011; Kawasaki, Janssen, & Inoue, 2011; 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.

This study is not the first one to report the results of a large internet sample. Logie and Maylor (2009) looked at prospective memory and examined the results of over 70,000 participants. Johnson, Logie, and Brockmole (2010) looked at

Our study does not look at prospective memory, but at

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. 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).

2.2. Materials

For the study, we selected a number of memory tests that covered immediate and delayed verbal and visuospatial memory (see Table 1). Each test is described below. To facilitate referring to the tests, they are given abbreviations that indicate whether they measured mainly immediate memory (labeled as STM for brevity) and delayed memory (labeled as LTM).

---INSERT TABLE 1 ABOUT HERE---

2.2.1. Verbal Immediate Memory

2.2.1.1. Ten Words Test

In the Ten Words Test (10WT-STM; Spaan, Raaijmakers, & Jonker, 2005), participants were shown ten words one by one. Order 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. 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).

2.2.1.3. Story Telling Immediate Recognition

Participants were in this test (Story-STM) asked 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 this, 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, while 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) in 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 twice correctly, 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 visual span score. The maximum sequence length was nine digits.

2.2.2. Visuospatial Immediate Memory

2.2.2.1. Pattern Span

The pattern span task (Pattern-STM) was an adaptation of that of Della Sala, Gray, Baddeley, Allamano, and 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 test 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 as 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 over 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 was an adaptation of the widely used Corsi Block Tapping Task (Corsi-STM; 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, while the inter-item interval was 0.5 seconds. The first sequence consisted of five blocks. When participants reproduced sequence correctly, by clicking the blocks in the correct order, the sequence-length increased with one block on the next 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. 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. 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 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 list studied, 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 DTM-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, total error scores on four lists of seven items in the test 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. After a participant had clicked on two squares of different colors, after a one-second interval both squares turned white again. Even though there were only twenty 'cards', the fact that all 'cards' remained in the game until the end made the test quite difficult. To make it a little easier, the participants were kept abreast of the 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. We expected this task to test aspects of delayed memory for newly learned patterns.

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, 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 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), presented before and after a third test (e.g., MemGame-LTM). The participants had to finish at least two memory tests, to which they had been randomly assigned to be included in the analyses. Participants completed on average 3.80 memory tests. (Unfinished tests were recorded as well, but they were not included in the analyses.) Presentation of all tests was visual and computer-controlled.

3. Results

We first analyzed the results of the tests separately. Table 2 gives the number of completed tests and their mean score. Three-way ANOVAs were calculated for each test. The F-values of each test for the effects of gender, age group and level of education are given in Table 3, with higher-order interactions given in Table 4. We generally found weak effects of gender and stronger effects of age and education. In the following paragraphs, 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, while male participants performed better on the Digit-STM, Pattern-STM, Corsi-STM and DNMT. No significant differences between men and women were found on the scores of the Story-STM, Story-LTM, and Texture-STM. When overall scores were calculated by averaging over z-scores on verbal and visuospatial tests (taking STM and LTM together, see below), the often reported finding that women tend to perform better on verbal tasks was obtained, although these differences are small in our experiment. Women performed better on the visuospatial MemGame-LTM. Without this test, 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: participants with higher education outperformed those with lower education on all tests. The effect is shown in Figure 3. 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, 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. 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%, versus 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 small drop compared with delayed memory (see Figure 5).

Figure 6 gives the grand average overall 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 Daily News Memory Test shows a quite 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 decrease in performances between 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%). Digit Span is thus a far less sensitive test for age differences than the 10WT-STM, 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 ten significant interaction effects. Women outperformed men on the MemGame-LMT to a greater degree for older participants (F(13, 9185) = 2.71, p < .01), while men outperformed women to a greater degree for older participants on the Pattern-STM and the DNMT (F(13, 10392) = 1.85, p < .05; F(13, 6317) = 2.12, p < .05). As participants become older, effects of education became smaller on six tests (MemGame-LTM, F(95, 9185) = 2.16, p < .001; DRM-STM, F(96, 11967) = 1.83, p < .001; 10WT-STM, F(96, 10498) = 1.76, p < .001; DNMT, F(94, 6317) = 1.52, p < .01; DRM-LTM, F(95, 7912) = 1.49, p < .01; Pattern-STM, F(96, 10392) = 1.34, p < .05). 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). 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 and F(91, 9185) = 1.84, p < .001.

3.5. False memory

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

Table 5 gives the proportions correct for the studied words (old) and for the three types of new words: unrelated words, weakly related words, and words from which the lists were derived (critical lure). Participants almost always indicated correctly that they had seen the studied words during the learning phase (M = .934) and that the unrelated not-studied words were new (M = .975), but they often incorrectly believed that they had seen the critical lures before (M = .672).

As can be observed in Figure 7, performance on both tasks decreases with age, but the decrease is faster 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 were taken. Decrease of the critical lures seems to be faster than for the other types of words. A linear regression on the decrease (excluding the two lowest 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, while the tendency to incorrectly recognize the unrelated or weakly related words did not.

---INSERT TABLE 5 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?

Quite complex models of memory and age can be developed, especially if many different types of tests have been administered (more than here). 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 model package (S.E.M.) for the statistical programming language R (Dalgaard, 2008). This technique allows one to model and test complex patterns of relationships as a whole rather than with separate analyses. Structural equation models assume linear relationships and cannot handle nonlinearity (e.g., inverted U-shaped curves). Since the results in Figure 4 suggested that age has a nonlinear effect on some tests, only data of participants 25 and older were used for the analyses to ensure that effects of age would be monotonic. Fits were based on the covariance matrix (Table 6) using a pair-wise missing values criterion, which means that each covariance is computed with a different set of participants. The correlations between age, education and the tests are also given in Table 6. 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 S.E.M. analyses.

The S.E.M. analyses were carried out over all data. Fits were based on the covariance matrix (Table 6) 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 many more participants, with the highest *N* being 8712. Model choice was based on the Bayesian Information Criterion (BIC). 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 (visuospatial versus verbal) and retention interval (yielding the immediate memory [STM] versus delayed memory [LTM] distinction). We first tested a model in which all tests were grouped together (the 'one memory factor' model in Table 7). Next, we tested a model in which tests were grouped on the basis of retention interval ('STM & LTM' in Table 7). 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 7). 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 7) led to a worse fit of the data. For reasons that will become clear later, we also tested a model in which our test of recent news events (DNMT) was treated as its own 'semantic memory' category ('STM & LTM, spatial & verbal, and semantic' in Table 7). This model did not perform well either. 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 6 AND 7 ABOUT HERE---

We also looked at the contribution of age. Do verbal and visuospatial memory decrease in the same way with age or not? Our previous analyses suggest 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 7, 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 much 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. Again, we found that age affects the scores on visuospatial memory tests more than the scores on verbal memory tests: a model with separate age parameters for the two factors was superior to a model with shared parameters (see Table 7, third group of models). For education, no difference was found between the effect on visuospatial and on verbal memory.

We noted above that one test, the DNMT (news events), has 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 ('separate age, shared education, extra for DNMT' model in Table 7). Both extra parameters received positive values, neutralizing the negative relation between age and performance on the verbal memory factor and strengthening the positive relation between education and all memory tests. In fact, the DNMT was uniquely affected by age and education, to such an extent that the best-fitting model was one in which the DNMT was placed in its own category. In this model ('separate age-, shared education influence, semantic with education' in Table 7), there were three groups of tests: verbal, visuospatial, and semantic (i.e., the DNMT). Age affected visuospatial tests strongly, verbal tests to a lesser degree, and the DNMT not at all. Education strongly affected the DNMT, and all other tests to a much smaller degree.

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 test (Pattern-STM), which showed an early peak at the 16-18 year bin (i.e., having age 17 at its center). Interestingly, at least one other study has also found that Pattern Span may decrease faster than other memory measures, with a peak around 18 years or earlier (Logie &

Maylor, 2009). 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 (e.g., Herlitz, Airaksinen, & Nordstrom, 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 one immediate memory test, 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, Gathercole, & Pickering, 2006), with an increase in performance between the ages 4 and 15 (Alloway et al., 2006; Gathercole, Pickering, Ambridge, & Wearing, 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).

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 rather shallow 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 (e.g., Bopp & Verhaeghen, 2007; Jenkins et al., 2000; 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 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 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 (e.g., 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 tendency to recognize words. Studied words, non-studied weakly related words, and non-studied unrelated words were not recognized more 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), while another study emphasized episodic memory (Lovden, 2003) as an important factor.

Though the age effect seems undisputed, there was no compelling 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 most important finding in this study was that individual differences are dominated by the type of items, verbal versus visuospatial. 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 for immediate versus delayed memory. After the age of 25, immediate and delayed memory gets 1-3% worse each year (in percentage of standard deviations) with immediate and delayed memory following a nearly identical pattern 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 create 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 Department of Psychology, Hokkaido 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, Roetersstraat 15, 1018 WB, 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,* 1698-1716.

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,* P239-P246.

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

Dalgaard, P. (2008). *Introductory statistics with R, 2nd edition*. New York: Springer.

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

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

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

Friedman, W. J., & Janssen, S. M. J. (2010a). Aging and the speed of time. *Acta Psychologica, 134*, 130-141.

Friedman, W. J., & Janssen, S. M. J. (2010b). Do people remember the temporal proximity of unrelated events? *Memory & Cognition, 38,* 1122-1136.

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*, 177-190.

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*, 93-104.

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*, 251-268.

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

Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*, 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*, 658-668.

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

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*, 755-767.

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*, 1847-1860.

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

Janssen, S. M. J., & Rubin, D. C. (in press). *Age effects in cultural life scripts.* Applied Cognitive Psychology.

Janssen, S. M. J., Rubin, D. C., & St. Jacques, P. L. (2011). The temporal distribution in autobiographical memory: Changes in reliving and vividness over the life span do not explain the reminiscence bump. *Memory & Cognition, 39,* 1-11.

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*, 157-175.

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

Kawasaki, Y., Janssen, S. M .J., & Inoue, T. (2011). Temporal distribution of autobiographical memory: Uncovering the reminiscence bump in Japanese young and middle-aged adults. *Japanese Psychological Research, 53,* 86-96.

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*, 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*, 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*, 220-237.

Kristo, G., Janssen, S. M. J., & Murre, J. M. J. (2009). Retention of autobiographical memories: An Internet-based diary study. *Memory, 17*, 816-829.

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*, 155-163.

Logie, R. H., & Maylor, E. A. (2009). An Internet study of prospective memory across adulthood. *Psychology and Aging, 24*, 767-774.

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,* 268-283.

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,* 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*, 793-810.

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,* 117-129.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27*, 272-277.

Nichelli, F., Bulgheroni, S., & Riva, D. (2001). Developmental patterns of verbal and visuospatial spans. *Neurological Sciences, 22*, 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*, 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,* 299-320.

Pedersen, N. L., Plomin, R., Nesselroade, J. R., & McClearn, G. E. (1992). A quantitative genetic-analysis of cognitive-abilities during the 2nd-half of the life-span. *Psychological Science, 3,* 346-353.

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,* 243-256.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9,* 95-123.

Robert, M., & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working-memory resources? *European Journal of Cognitive Psychology, 18,* 378-397.

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*, 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,* 385-405.

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

Salthouse, T. A. (2009). When does age-related cognitive decrease begin? *Neurobiology of Aging*, *30*, 507-514.

Salthouse, T. A., & Siedlecki, K. L. (2007). An individual difference analysis of false recognition. *American Journal of Psychology, 120*, 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*, 1-24.

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

Skitka, L. J., & Sargis, E. G. (2006). The Internet as psychological laboratory. *Annual Review of Psychology*, *57*, 529-555.

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*, 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*, 230-241.

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

Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62*, 19-34.

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*, 792-802.

Table 1. 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 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-value for each test and each main effect.

	Gender	*Age Group*	*Education*
		Verbal STM	
10WT-STM	33.939***	17.724***	9.218***
DRM-STM	6.265*	2.981***	18.869***
Story-STM	0.265	2.981***	1.234
Digit-STM	3.007*	3.436***	5.543***
		Visuospatial STM	
Pattern-STM	19.270***	34.871***	2.659*
Corsi-STM	13.939***	12.936***	2.962**
Texture-STM	0.607	8.519***	4.846***
		Verbal LTM	
DRM-LTM	11.947**	2.802***	10.728***
Story-LTM	0.159	1.561	6.057***
		Visuospatial LTM	
MemGame-LTM	38.126***	31.021***	13.720***
		Other LTM	
DNMT	79.263***	17.513***	34.314***

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

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

	*Gender*Age*	*Gender*Edu*	*Age*Edu*	*Gender*Age*Edu*
		Verbal STM		
10WT-STM	1.401	2.425*	1.458**	1.398*
DRM-STM	1.916*	0.436	1.363*	1.158
Story-STM	0.817	0.765	1.184	1.171
Digit-STM	1.588	0.63	1.249	1.320*
		Visuospatial STM		
Pattern-STM	1.226	1.151	0.94	1.005
Corsi-STM	0.901	0.712	1.212	1.101
Texture-STM	1.343	1.586	0.994	1.011
		Verbal LTM		
DRM-LTM	1.198	0.25	1.266	0.877
Story-LTM	1.171	1.265	1.241	1.19
		Visuospatial LTM		
MemGame-LTM	3.025***	4.181***	1.933***	2.292***
		Other LTM		
DNMT	3.478***	2.772**	5.110***	2.920***

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

Table 5. The proportion of words in DRM test classified as sure old, probably old, probably new, and sure new.

	Sure Old	Probably Old	Probably New	Sure New
Studied Words	0.881	0.053	0.021	0.045
Non-studied Words				
Unrelated	0.010	0.005	0.019	0.966
Related	0.081	0.090	0.119	0.710
Critical Lure	0.539	0.133	0.089	0.240

Table 6. Correlations (above diagonal) and variances/covariances (below diagonal) between age, education and the individual tests for participants who were older than 25 years old.

	Age	Edu	DRM-STM	Digit-STM	Story-STM	Pattern-STM	Corsi-STM	Texture-STM	10WT-LTM	DRM-LTM	Story-LTM	Mem Game-LTM	DNMT
Age	145.2	-.117**	-.203**	-.110**	-.101**	-.417**	-.271**	-.249**	-.296**	-.098**	-.072**	.400**	.032*
Education	-2.9	4.37	.224**	.179**	.058**	.205**	.150**	.147**	.201**	.166**	.147**	-.166**	.267**
DRM-STM	-17.2	3.34	50.82	.267**	.158**	.261**	.220**	.238**	.531**	.428**	.153**	-.341**	.178**
Digit-STM	-1.7	0.49	2.37	1.75	.083**	.190**	.205**	.234**	.233**	.177**	.082**	-.164**	.070**
Story-STM	-1.5	0.15	1.41	0.14	1.70	.086**	.075**	.106**	.144**	.132**	.077**	-.115**	.038
Pattern-STM	-272.3	23.41	105.20	13.62	6.29	3057	.325**	.236**	.281**	.173**	.106**	-.349**	.076**
Corsi-STM	-2.3	0.22	1.14	0.19	0.07	12.92	0.53	.203**	.180**	.107**	.084**	-.267**	.071**
Texture-STM	-4.1	0.43	2.26	0.43	0.21	20.09	0.23	2.15	.188**	.090*	.094**	-.261**	.010
10WT-LTM	-19.0	2.25	18.94	1.59	0.95	82.29	0.68	1.39	28.59	.304**	.187**	-.315**	.133**
DRM-LTM	-7.8	2.33	20.10	1.56	1.15	66.68	0.54	0.88	10.92	46.25	.137**	-.218**	.109**
Story-LTM	-1.1	0.39	1.37	0.14	0.13	7.86	0.08	0.17	1.23	1.18	1.66	-.156**	.156**
MemGame-LTM	191.8	-14.10	-97.40	-8.65	-6.18	-793.90	-7.90	-15.15	-65.66	-59.81	-8.28	1674	.070**
DNMT	1.6	2.34	5.15	0.38	0.21	17.51	0.21	0.06	2.91	3.07	0.84	-11.98	17.64

**p* < .05; ***p* < .01.

Table 7. 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 are reported as well as the results of a chi-square null hypothesis test against the null model.

Data	Model description	# param	GoF	BIC	Test against null model
10 STM / LTM tests	One memory factor	20	0.965	-2.3	$X^2(35) = 256.6$, $p < .001$
	STM & LTM	21	0.979	-0.9	$X^2(34) = 250.6$, $p < .001$
	Verbal & spatial memory	21	0.985	-134.9	$X^2(34) = 116.6$, $p < .001$
	STM & LTM, spatial and verbal	22	0.948	255.1	$X^2(33) = 499.2$, $p < .001$
	STM & LTM, spatial & verbal, and semantic	23	0.984	238.8	$X^2(32) = 475.4$, $p < .001$
Age	Shared influence age on verbal & spatial	23	0.966	-4.6	$X^2(43) = 313.4$, $p < .001$
	Separate for verbal and spatial	24	0.979	-123.1	$X^2(42) = 187.5$, $p < .001$
	Separate for each test	32	0.958	175.9	$X^2(34) = 427.4$, $p < .001$

	Shared age-, separate education influence	26	0.968	-51.5	$X^2(52) = 318.3$, $p < .001$
	Separate age-, shared education influence	26	0.973	-113.7	$X^2(52) = 270.9$, $p < .001$
Age & Education	Separate age and education influence	27	0.973	-106.6	$X^2(51) = 270.7$, $p < .001$
	Separate age-, shared education, extra for DNMT	28	0.979	-171.5	$X^2(50) = 198.3$, $p < .001$
	Separate age-, shared education influence, semantic with education	27	0.979	-171.8	$X^2(51) = 205.4$, $p < .001$

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.

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

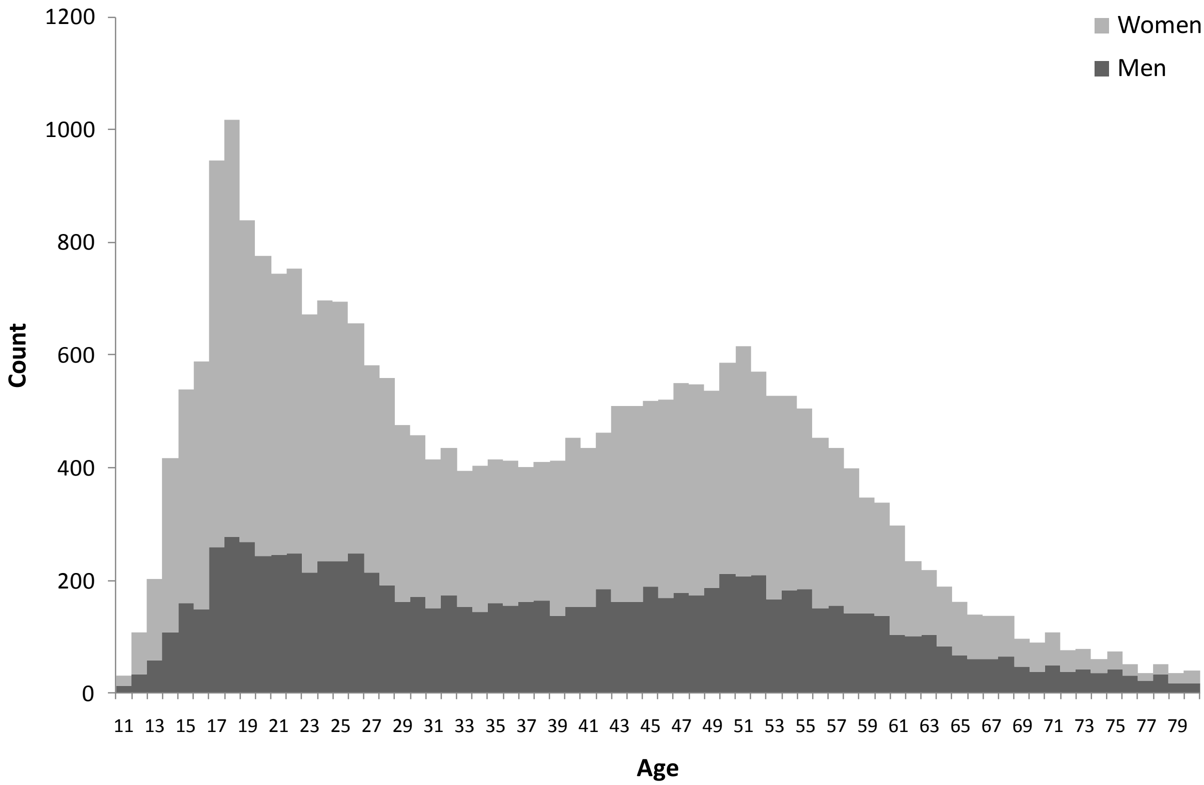
Figure 5. Averaged Z-transformed scores as a function of age for immediate and delayed tests.

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 dashed line), studied related words (triangles with dashed line), not-studied related words (squares with solid line) and critical lures (triangles with solid line).

Figure 1



1200
1000
800
600
400
200
0
Count
Women
Men
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

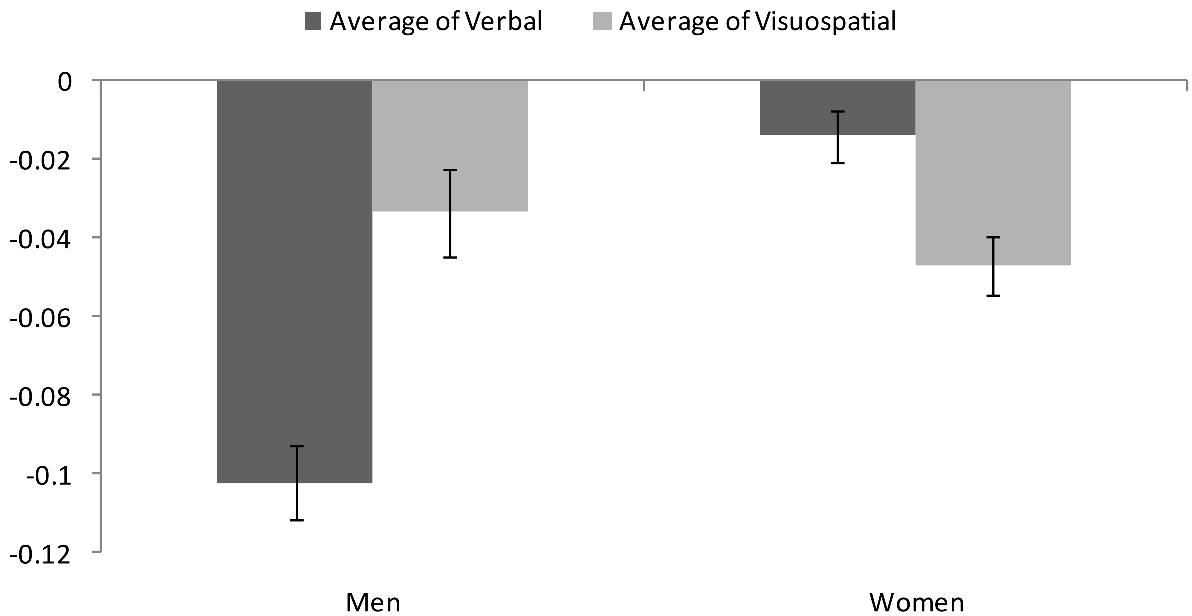
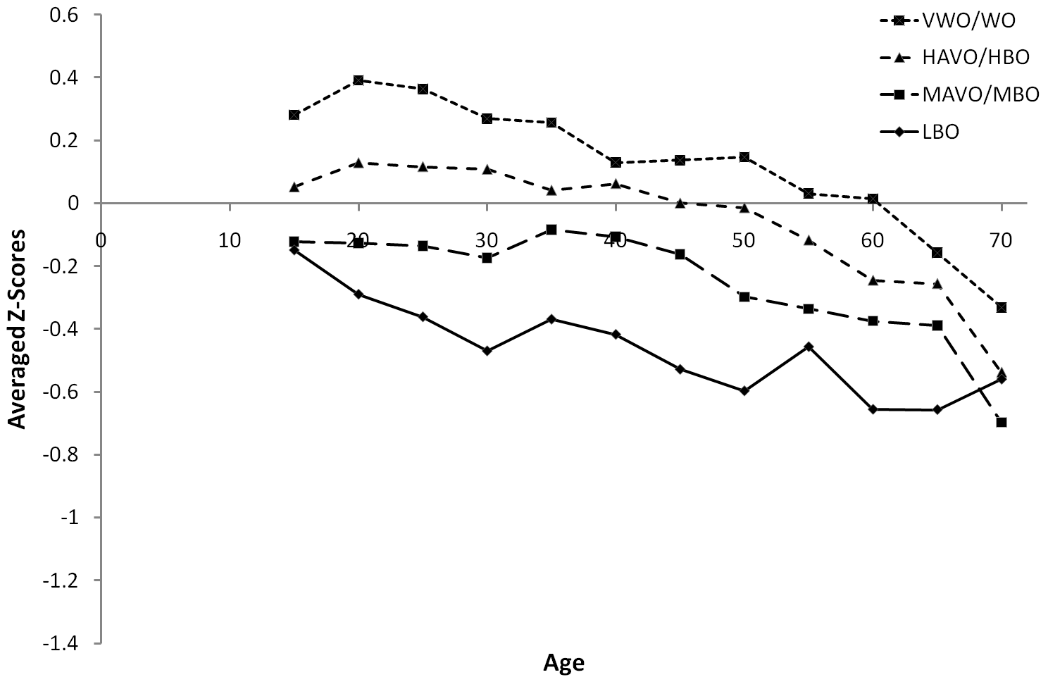

Average of Verbal
Average of Visuospatial
0
-0.02
-0.04
-0.06
-0.08
-0.1
-0.12
Men
Women

Figure 3

(a)


VWO/WO
HAVO/HBO
MAVO/MBO
LBO
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

(b)

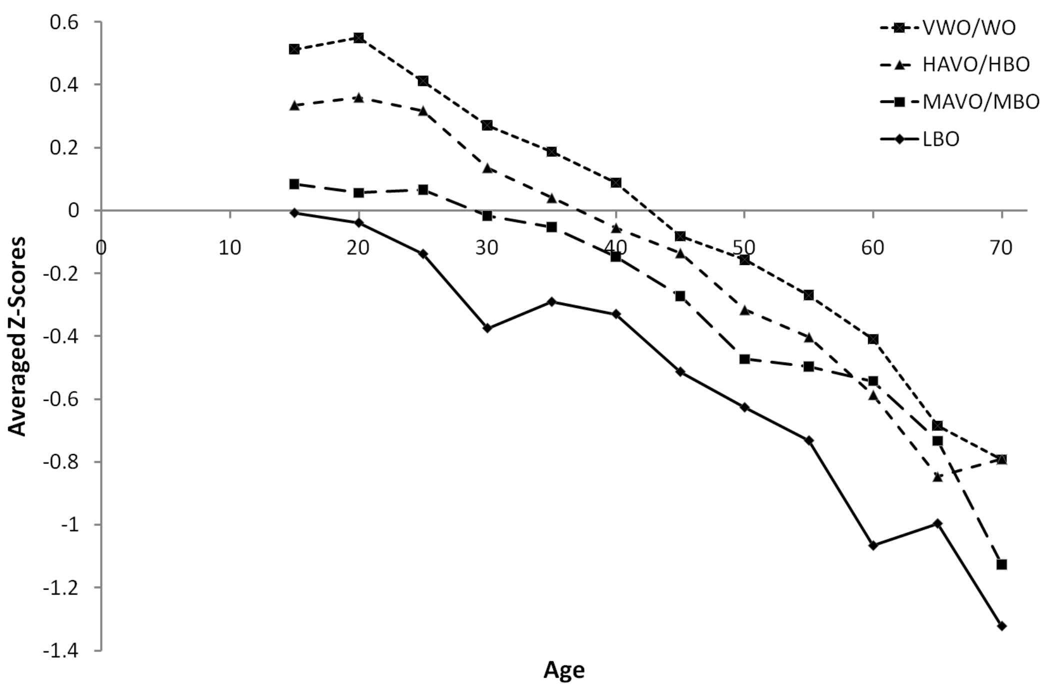
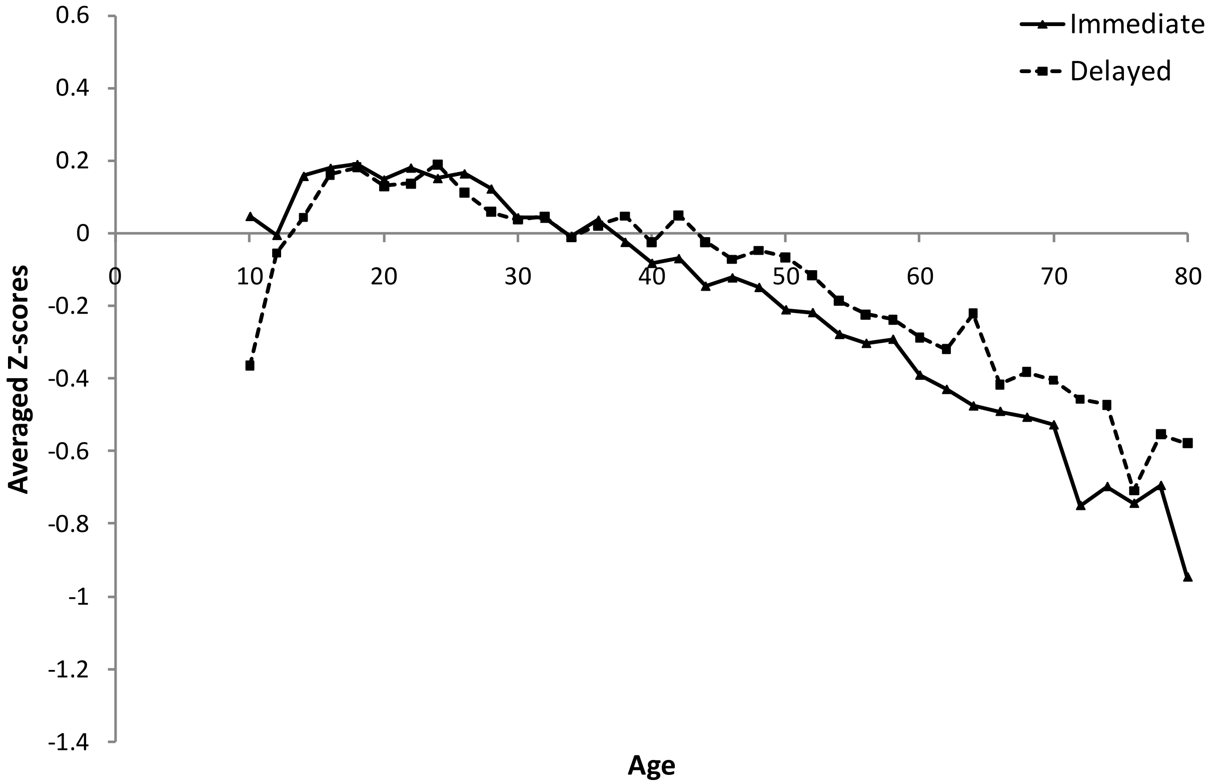

VWO/WO
HAVO/HBO
MAVO/MBO
LBO
Averaged Z-Scores
Age
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70

Figure 4

Verbal
Visuospatial
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 5



Immediate
Delayed
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 6

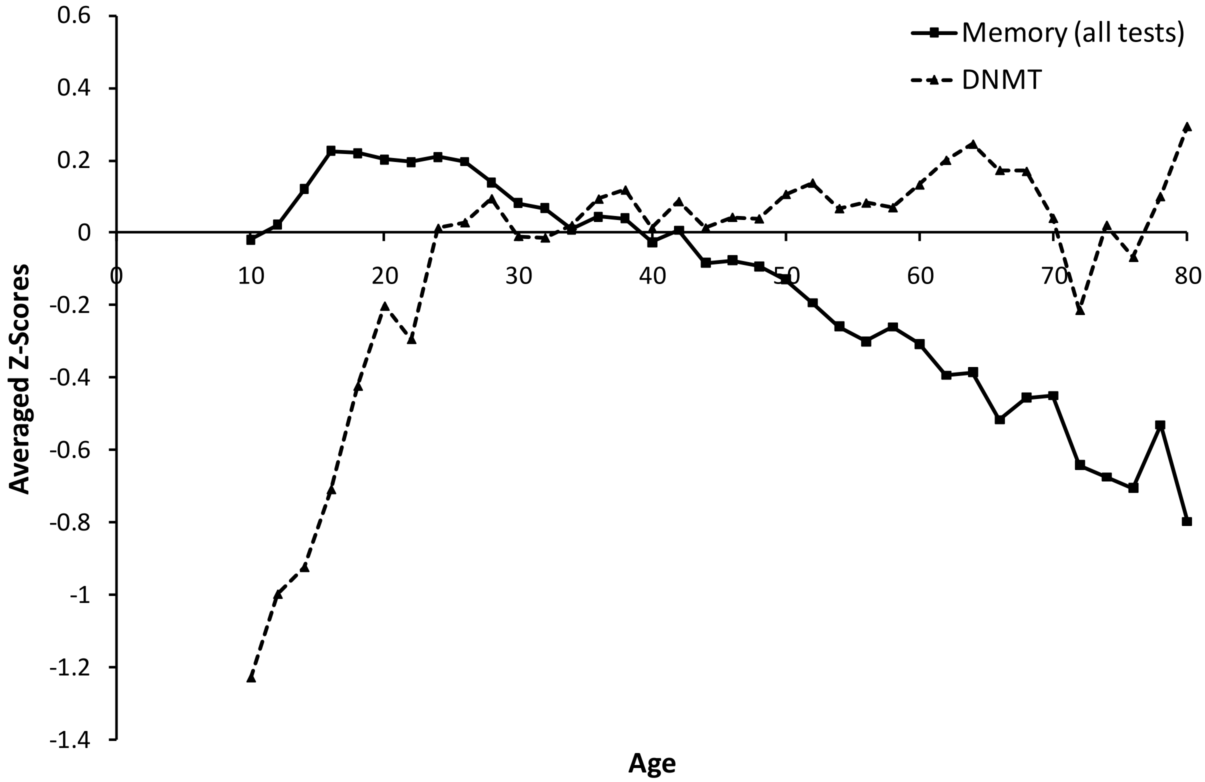
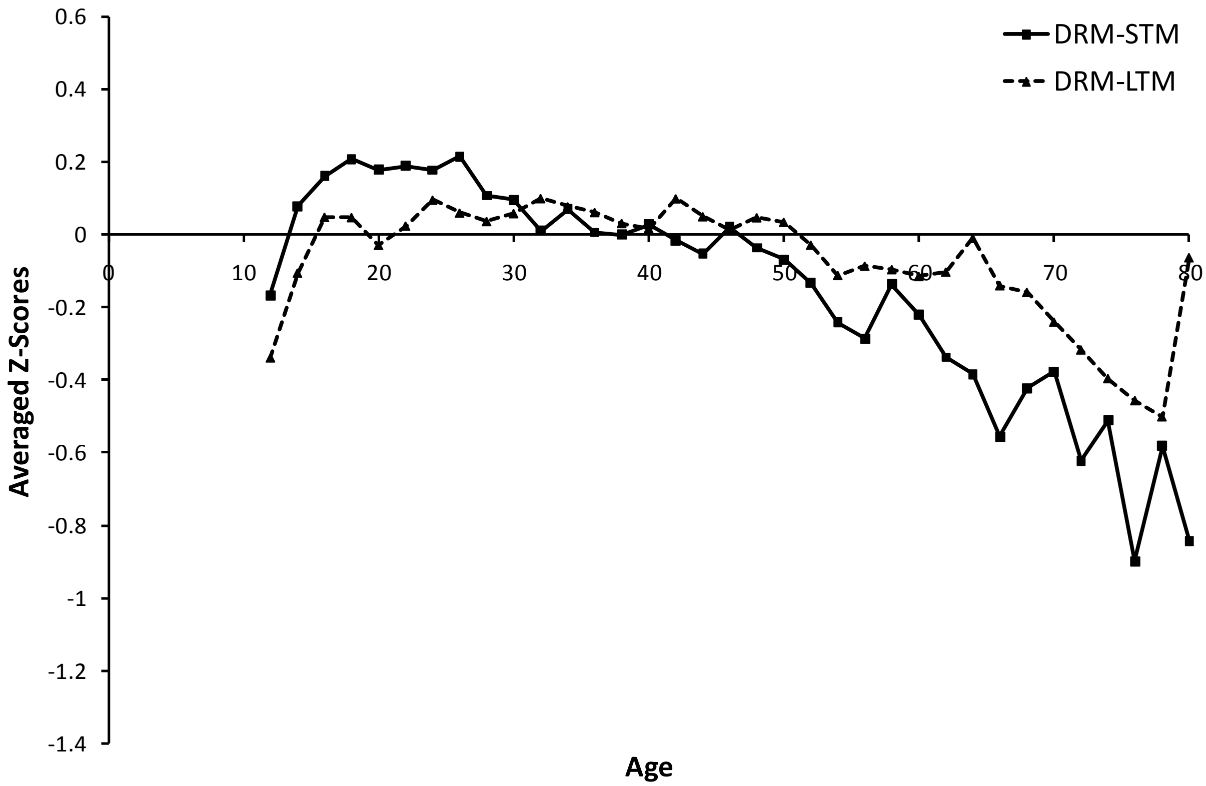
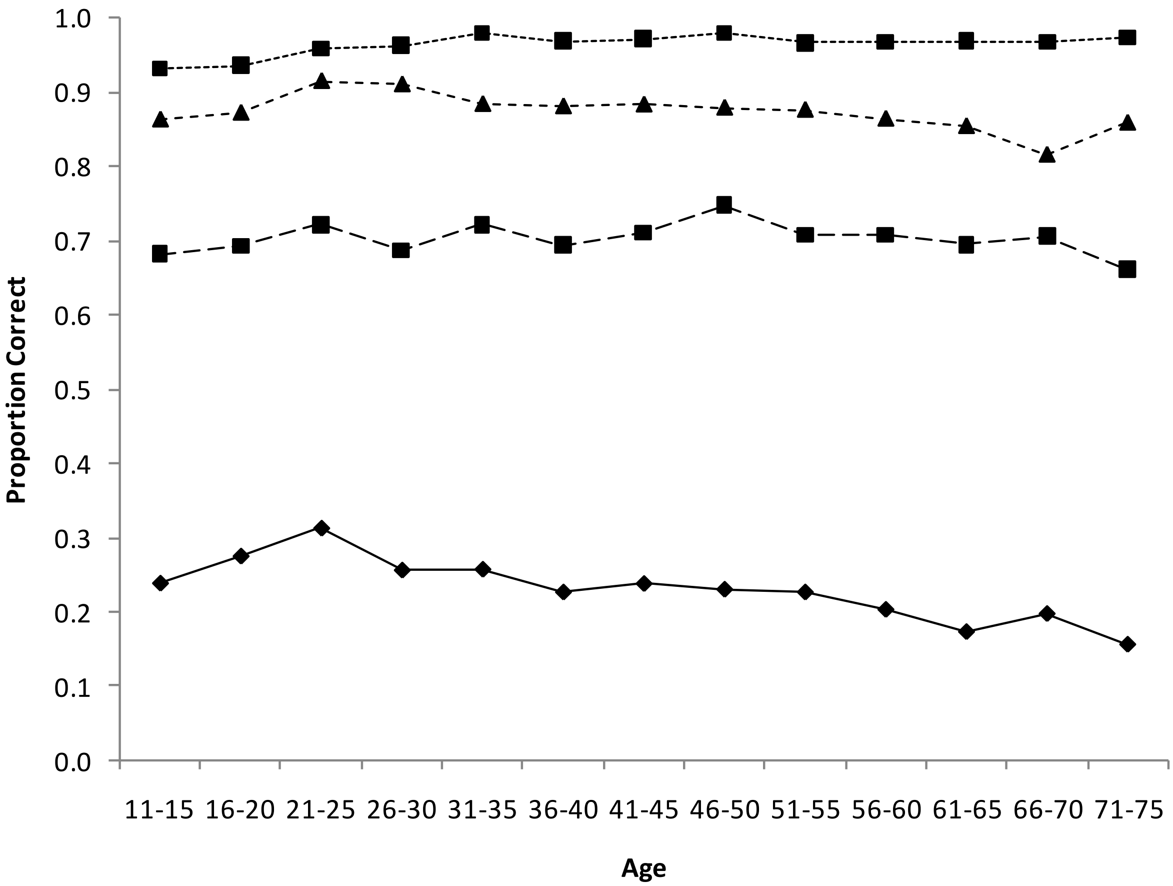

Memory (all tests)
DNMT
Averaged Z-Scores
Age
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
-1.4
0
10
20
30
40
50
60
70
80

Figure 7



DRM-STM
DRM-LTM
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 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