Intelligence 38 (2010) 513–528

Contents lists available at ScienceDirect

Intelligence

Working memory tasks differ in factor structure across age cohorts: Implications for dedifferentiation

Wendy Johnson [a,b,c,*], Robert H. Logie [a,b,*], James R. Brockmole [d]

[a] Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, United Kingdom
[b] Department of Psychology, University of Edinburgh, United Kingdom
[c] Department of Psychology, University of Minnesota – Twin Cities, United States
[d] Department of Psychology, University of Notre Dame, United States

ARTICLE INFO

Article history:
Received 3 November 2009
Received in revised form 8 June 2010
Accepted 10 June 2010

Keywords:
Working memory
Measurement invariance
Internet testing
Cognitive aging

ABSTRACT

Researchers interested in working memory have debated whether it should be considered a single latent cognitive ability or a set of essentially independent latent abilities distinguished by domain-specific memory and/or processing resources. Simultaneously, researchers interested in cognitive aging have established that there are substantial differences in rates of change in various aspects of cognitive function with age. In general, so-called fluid measures of cognitive function including working memory decline at faster rates in later adulthood than so-called crystallized measures. Using an Internet working and short-term memory test battery completed by over 95,000 people aged 18–90, we used multiple-group confirmatory factor analysis to assess the extent to which working memory could be considered a single latent ability as well as how its common and unique variance components varied with age. Results indicated a single latent factor, but the tests did not reflect this factor consistently across age groups. Both individual test residual variances and factor intercepts showed different patterns of variation with age. We discuss the implications for understanding age differences in working memory function.

Individuals' performances on tests of different cognitive abilities are positively inter-correlated. The general factor that characterizes these correlations was observed around the time the first cognitive ability tests were developed, and has come to be known as Spearman's *g*, after one of the first psychometricians to study it (Spearman, 1904, 1927). Spearman focused his analytical attention on the tests themselves and the extent to which specific tests reflected this general factor (Spearman, 1927), but he also developed hypotheses about the psychological natures of the traits represented by the tests. He proposed that any individual's performance on any cognitive task arose from a general ability (*g*) that could also be applied to any other cognitive task, and a cognitive ability specific to that particular task. Almost immediately, others took issue with this proposition (beginning with Thurstone, 1931), maintaining that cognitive abilities are distinct and independent from each other, with no general factor. Despite the strong evidence for the pervasiveness of the general factor across test batteries and samples that has developed since then regarding *g* (Jensen, 1998), the debate about the existence of some general ability has continued. This is most notable in studies of individual differences in working memory, which has been shown to correlate highly with a wide range of fluid ability assessments and in particular with 'g' (e.g. Conway, Kane, & Engle, 2003; Cowan, 2005; Kane & Engle, 2002; Unsworth, Redick, Heitz, Broadway, & Engle, 2009).

In the context of working memory, the debate about the existence of a general form has been fueled by the large literature that shows that a wide range of tasks can be distinguished and

* Corresponding authors. Centre for Cognitive Ageing and Cognitive Epidemiology and Department of Psychology, 7 George Square, Edinburgh EH8 9JZ, UK. Johnson is to be contacted at Tel.: +1 952 473 1673; fax: +1 952 626 2079. Logie, Tel.: +44 0131 651 1394.

E-mail addresses: wendy.johnson@ed.ac.uk (W. Johnson), rlogie@staffmail.ed.ac.uk (R.H. Logie).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.06.005

made to appear independent through experimental manipulations (see reviews in Baddeley, 2007; Conway, Jarrold, Kane, Miyake, & Towse, 2007; Logie & van der Meulen, 2009), as well as by studies of individual differences (e.g. Bayliss, Jarrold, Gunn, & Baddeley, 2003; Friedman et al., 2008; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). Case studies of the specificity of deficits experienced by victims of neurological damage (reviewed in e.g. Alloway & Gathercole, 2006; Baddeley, Kopelman, & Wilson, 2002; Logie & Della Sala, 2005; Vallar & Shallice, 1990) provide additional direct evidence of task distinctiveness, and the consistently observed fact that different kinds of cognitive abilities show very different patterns of change with age (e.g., Maylor & Logie, 2010; Park et al., 2002; Salthouse, 1995) provides further indirect evidence. Though initially focused on the question of the separability or generality of cognitive ability broadly construed, more narrowly defined aspects of cognition such as executive function, working memory, and attention have been drawn into the separability versus generality debate as well. For example, Baddeley and Logie (1999) posited a multiple-component model of working memory. Under their model, working memory consists of a central executive controller, one separable subsidiary system for temporary storage of phonologically-based material, and a second separable subsidiary system for temporary storage of visuospatially-based material. In actual task performance at the level of day-to-day activities or even tests of cognitive ability as usually construed, the central executive coordinates the subsidiary memory systems, controls long-term memory storage and retrieval processes and placement of attention, and carries out the manipulation of the material stored in the subsidiary memory systems. In contrast to Baddeley and Logie (1999) proposal, other researchers (e.g. Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007; Cowan, 2005) have suggested a model based on centrality of processing that emphasizes the interference between verbal and spatial processing that takes place when attentional resources are constrained.

At the same time, the different patterns of cognitive change with age have received considerable research attention. In both cross-sectional and longitudinal data, tests that rely on factual knowledge tend to show gradual average increases until even as late as age 70, and then relatively slow declines with age. In contrast, tests of nonverbal reasoning tend to show faster declines with age that begin around age 30, and tests of perceptual speed show even sharper declines beginning as early as age 20. (Hunt, 1949; Jones & Conrad, 1933; Logie & Maylor, 2009; Lovden, Ghisletta, & Lindenberger, 2004; Salthouse, 2009; Schaie, 1994). In all cases, however, longitudinal data show less extreme differences with age than do cross-sectional data. Whether cross-sectional or longitudinal, the age differences are thought by many to reflect cognitive aging processes, with the differences in rates of change with age reflecting differences between the aging processes underlying tests that measure primarily current efficiency of various kinds of information processing and memory and other tests that measure cumulative products of processing carried out in the past (Salthouse, 2006). To some, these differences suggest the presence of relatively modular cognitive capacities that change with age at different rates regardless of a common factor such as general working memory capacity (e.g., Logie, Della Sala, MacPherson, & Cooper, 2007; Logie & Maylor, 2009; Park et al., 2002; Perfect & Maylor, 2000). Others, however, see changes in *g* as the main drivers of cognitive aging (e.g., Gow et al., 2008; McGue & Christensen, 2002; Rabbitt, Lunn, Wong, & Cobain, 2008). If changes in *g* are to be understood as the main drivers, *g* and its components such as working memory must be measured consistently with age. By 'measured consistently with age,' we mean that performances on tests used to generate *g* or one of its components must reflect that general factor in the same way and to the same extent at different ages. To date whether this is the case has not been investigated systematically throughout the adult lifespan, but it was the primary purpose of the present study.

Consistency of measurement of *g* and its components with age has relevance for the age differentiation hypothesis as well. This is the idea that, during childhood, general ability gradually develops into more specific abilities, while in later life the more specific abilities face global biological constraints that tend to cause their distinctions to blur, that is, later-life dedifferentiation (Balinsky, 1941; Baltes & Lindenberger, 1997; Garrett, 1938; Li, Lindenberger, & Sikstrom, 2001). Support for this proposition has been mixed at both ends of the lifespan (e.g., Anstey, Hofer, & Luscez, 2003; Bickley, Keith, & Wolfle, 1995; DeFrias, Lovden, Lindenberger, & Nilsson, 2007; Juan-Espinosa et al., 2002; Tucker-Drob, 2009; Tucker-Drob & Salthouse, 2008 for studies evaluating later-life dedifferentiation). Differences in measurement properties can provide support for either differentiation or dedifferentiation. Support for the later-life dedifferentiation hypothesis would be provided if there were particular kinds of inconsistencies of measurement of *g* and its components such as working memory in different adult age groups. The inconsistencies should suggest that individual tests become more closely related to each other with age (controlling for simple changes in variance of the general factor), and/or that individual tests show progressively less test-specific variance with age, especially if the latter is true to different degrees for different tests.

1. Establishing consistent measurement across groups of people

Our main goal in this study was to contribute to the debate regarding the existence of a general working memory construct by assessing how psychometric tests measure the construct across age. To do this, we used confirmatory factor analytic tests of the factor structure of a group of working memory tasks, and of the measurement invariance (Meredith, 1993) of this structure across age groups reflecting the adult lifespan. When a construct is *measurement invariant* across some group division, it is measured in the same way from group to group, and both within-group individual differences and between-group differences in means can be considered to reflect differences in level of the underlying latent construct (Lubke, Dolan, Kelderman, & Mellenburgh, 2003; Meredith, 1993). When a construct is not measurement invariant across groups, between-group mean differences do not reflect differences in level of the underlying construct alone; in addition to construct differences among the groups, there are differences in the relative importance of the various marker variables used to define the latent construct (Hofer, Horn, & Eber, 1997), and/or some or all of the tests may be somehow measuring different ways of approaching the tasks. When measurement invariance fails, the methodological work involved in addressing its extent and source can be quite informative in identifying to what degree, in what form, and on what tests measurement varies across the groups, leading to both further theoretical developments to improve understanding of cognitive ability and practical applications to improve tests.

The methodological work in addressing measurement invariance takes place in several steps that impose increasingly restrictive equality constraints on the models used to describe the data in the groups under consideration. Each step is important for establishing full measurement invariance, but the implications of each step for the measurement properties across groups differ. There is a logical order to the steps, and they are carried out in the order given.

The first step is to define a basic model of the factor pattern to describe the data in all groups. This is generally termed establishing *configural invariance*. If this cannot be accomplished, not only is there no point in progressing to the second step, but the basic factor structure of the data is not the same in the groups being compared: there may be more factors in one group than another, or certain items or scales may load on completely different factors in the groups. The second step is to constrain the factor loadings equal across the groups, termed *metric invariance*. When this can be done without loss of overall model fit, the relations among the variables (though not necessarily the factor variances themselves) are the same across the groups and the relative contributions of the various tests to the factor(s) are the same in all groups. When it cannot be done, there are important differences in the extent to which some or all of the tests represent the factors. Next, residual variances are constrained equal. When this is possible, the reliabilities of the tests are functions only of the factor variances in all groups. When it is not possible, there are differences either in error variance or in systematic test-specific variance across groups. Alternatively, residual variances may be left free and factor intercepts constrained equal to test *strong invariance.* When this is possible, latent factor mean differences are interpretable, though mean differences in the tests themselves may not be attributable to mean differences in the latent factor(s). Finally, factor intercepts and residual variances are constrained equal, and *strict invariance* is tested. When this can be done without loss of overall model fit, group mean differences in the tests can be attributed to mean differences in the latent factor(s), and the tests can be considered to reflect the latent construct in the same way in all groups. When there are differences in intercepts but not residual variances among groups, there is some difference between groups that affects mean test levels but not the latent factor(s) (Brand, 1987), such as different problem-solving strategies, levels of background knowledge or problem-solving sophistication, or familiarity with procedures (Johnson & Bouchard, 2007b; Wicherts, 2007; Wicherts, Dolan, & Hesson, 2005). The tests involved in establishing measurement invariance tend to have low power (Molenaar, Dolan, & Wicherts, 2009). That is, there may be important group differences in measurement properties that cannot be detected with these statistical tests. Thus, working with large samples is important.

2. The present study

The purpose of the present study was to investigate systematically the extent to which one component of *g*, working memory, is a unitary construct measured consistently across the adult lifespan, in order to evaluate the extent to which it may differentiate with age. Age is a particularly important variable on which to evaluate the extent to which working memory represents a single unitary construct because individuals pass through different ages in the course of their lives. This is not true of many grouping variables: for example, in general individuals are either male or female; they are not male at some times and female at others. In addition, any failures of measurement invariance of working memory with age that we observed would provide new evidence for the degree to which the dedifferentiation hypothesis holds in later life. To accomplish these two goals, we made use of multigroup confirmatory factor analysis testing measurement invariance in a battery of tasks put together to address different aspects of working and short-term memory function and administered over the Internet in collaboration with the British Broadcasting Corporation (BBC). Ideally, our sample would be longitudinal, so that we could evaluate the extent to which the common working memory factor can be measured consistently in the same people as they change with age. This was not the case for this sample, but the extent to which working memory can be measured consistently in people born at different times has not been evaluated either, so our cross-sectional sample served as a good introduction to the general subject of its consistency of measurement across age.

Our Internet collaboration with the BBC made it possible to accumulate an unusually large sample in excess of 95,000 individuals ranging in age from 18 to 90. Collecting psychological data over the Internet is becoming increasingly common (see Birnbaum, 2004; Reips, 2002; Skitka & Sargis, 2006). It makes common the accumulation of large samples, and these samples tend to represent broader demographics than most laboratory-based studies and even than many studies that rely on community volunteers, with substantial savings of research time, energy, and resources. In addition, there may be advantages particular to research on aging. Participants do not have to travel and may be less anxious when tested in their own familiar environments (Maylor & Reimers, 2007), and older adults are increasingly being encouraged to use the Internet (Cutler, Hendricks, & Guyer, 2003; Selwyn, Gorard, Furlong, & Madden, 2003). Moreover, accumulating evidence suggests that Internet studies can reliably replicate more conventional data collection methods even in studies of aging (Della Sala, Darling, & Logie, 2010; Logie & Maylor, 2009; Maylor & Logie, 2010; Maylor et al., 2007; Robins, Trzesniewski, Gosling, & Potter, 2002).

3. Method

3.1. Sample

Participants were volunteers who accessed the Science pages of the BBC's official website between May 25, 2006 and March 19, 2007. The study was clearly advertised on the site, and it was also advertised on BBC radio programs, was the topic of an article in the August, 2006, *Radio Times* program guide magazine (Logie, 2006), and was featured in a major BBC television documentary on human memory aired on August 9, 2006 (Cadman, 2006). The first participant completed the tests within an hour of their posting on the site and approximately one third of the participants had completed the tests prior to the August television broadcast, suggesting that many participants regularly and spontaneously accessed the BBC site. Defining a data record as the data set for a single attempt to undertake the test battery including partially completed attempts, 160,405 data records were collected during the 10-month data collection period. Participants were requested to provide information on country of residence, highest level of completed education, sex, and age.

Participants reported 156 different countries of residence, but the vast majority reported residing in the United Kingdom or United States. Initial data analysis revealed no differences between those reporting residence in English language-dominated countries and those not. Though participants were not specifically asked to indicate fluency in English, we considered it reasonable to assume that they had a high level of English fluency because they would have had to find and choose to access the website through English language web pages maintained by the BBC.

For this study, we made use of a subset of the 160,405 data records collected. We excluded all those who did not provide age and/or education, as well as those who reported ages under 18 and over 90. This left 111,497 data records. Among the data records, it was not possible to determine whether individuals completed the tests multiple times. It was, however, possible to ascertain how many times a particular computer had been used to attempt the tests. To minimise the possibility of multiple attempts by the same individual, we made use of only the first data record from each computer for all those 65 and under. There were relatively few participants above age 65, and inspection of the multiple data records from the same computers indicated that few if any were from the same individuals (age, education, and sex did not match and scores varied widely). We therefore included these data records in order to maintain sample sizes as large as possible for the older groups. This left a total of 95,201 data records. Among them were participants who reported unrealistic levels of education such as a postgraduate degree at age 18. We deleted the education variable for these data records. Some participants did not report sex, but we did not exclude their data records on this basis.

3.2. Measures

We made use of the five tasks in the data set that had scale scores that could reasonably be considered continuous and measured in scales with roughly equal numbers of intervals. There were several other dichotomous measures in the set (see Logie & Maylor, 2009; Maylor & Logie, 2010). Our only reason for not including them in our analyses was that dichotomous variables have very limited measurement properties that would not contribute meaningfully to estimation of a common factor in combination with the other quantitative variables. The five tasks tapped a wide range of aspects of working, visual, and verbal short-term memory. Three of the tests (Digit Span, Working Memory Span and Spatial Orientation) have been used in multiple previously published assessments of fluid intelligence, a fourth (visual pattern span) has been published with normative data. One of the tests (Feature Binding) was more novel, but variations of this test recently have been used to assess cognitive decline with age. Specific descriptions of our variables follow.

3.2.1. Education

Participants reported highest level of attained education on a scale from 1 to 7, where 1 indicated no education; 2 primary education; 3 secondary education; 4 technical or vocational college; 5 other college; 6 college graduate with a first degree; and 7 postgraduate education.

3.2.2. Feature Binding (adapted from Brockmole, Parra, Della Sala, & Logie, 2008; Logie, Brockmole, & Vandenbroucke, 2009)

Participants were presented with screens containing 1 to 4 objects for 1 second each. The test started with two trials, each consisting of one object, and followed with two trials consisting of two objects, two trials consisting of three objects, and finally two trials consisting of four objects, for a total of 20 objects across all trials. After the presentation of each screen, there was a 1-second gap followed by a test screen showing a selection of colors, shapes, and locations. The task was to report the color, shape, and location of each object that had been on the previously presented screen by clicking on the color and shape of each object and then clicking on its presented location. The test stopped when participants failed accurately to report anything from the two trials at a particular array size.[1] Performance was scored as the number of trials for which color, shape, and location of all objects were reported correctly and thus ranged from 0 to 8. We presumed that specific abilities tested by this task involved visual working memory, visual attention, and binding of features to form representations of integrated objects (e.g. Brockmole et al., 2008; Logie et al., 2009; Mitchell, Johnson, D'Esposito, Raye, & Mather, 2000; Treisman, 2006).

3.2.3. Visual Pattern Span (adapted from Logie & Pearson, 1997)

Participants were presented with screens containing matrices consisting of patterns of white and blue squares for 2 s each. Immediately after presentation of each screen, a matrix of blank squares was presented, and participants were to click on the squares in the matrix that had been blue on the previous screen. There were two trials each of 3 × 3 (5 blue squares), 3 × 4 (5 blue squares), 4 × 4 (8 blue squares), 4 × 5 (9 blue squares), and 5 × 5 (9 blue squares), for a total of 10 trials. The test stopped when participants failed to recall all of the blue squares at a single matrix size. Performance was scored as the number of trials for which all blue squares were recalled and thus ranged from 0 to 10. A standard laboratory version of this test has been published with normative data and has been shown to be a robust measure of immediate visual memory (Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999).

3.2.4. Digit Span

Participants were presented with sequences of screens containing randomly generated single digits for 1 s each. At the end of each sequence, a screen containing a blank box appeared, and participants were to type in the sequence of

[1] This was purely for practical reasons, to prevent participant frustration. As each step of the tasks increased in difficulty, it should have had no impact on participants' scores, as participants who could not recall all the features of two objects in either of two trials presumably would also not have been able to recall all the features of three trials. Participants were scored for all the trials for which they gave correct answers, regardless of how far within the full test they progressed. This was true for Visual Pattern Span as well.

digits they had seen. Two sequences at each length from 3 to 9 were presented. The test stopped when participants failed to recall both sequences at a single sequence length. Performance was scored as the total number of digits recalled in the correct orders and thus ranged from 0 to 84. Variations of this test, originally devised by Jacobs (1887), have been included in standard test batteries of fluid intelligence almost since their inception, and appear in contemporary standard assessments such as the Wechsler Adult Intelligence Scale. It is assumed to assess the specific ability to retain a verbal sequence in immediate memory, and is often associated with the operation of the phonological loop component of working memory (see review in Baddeley, 2007).

3.2.5. Working Memory Span (adapted from Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Duff & Logie, 2001)

Participants were presented with sequences of screens containing simple sentences such as 'Flies are insects' or 'Mobile phones are made of cheese,' and buttons indicating 'true' or 'false.' As fast as possible, participants were to click on the appropriate button for the presented sentence and were to remember the last word of each sentence. A new sentence was then presented until the sequence was complete. Participants were then presented with a 4×5 array of words in the left two-thirds of the screen. The final words of each sentence in the sequence were randomly distributed throughout the array, and the remaining words were unrelated foils. Participants were to click on the final words from the presented sentences in the correct order and drag them to boxes arranged vertically on the right third of the screen. Two sequences at each length from 2 to 6 were presented. The test stopped if the participant failed to recall all of the last words of the sentences at a single sequence length. Performance was scored as the total number of final sentence words recalled in the correct order and thus ranged from 0 to 40. The Baddeley et al. (1985) version of this test was shown to correlate with a range of cognitive abilities. Different versions of this kind of test, thought to involve both cognitive processing and immediate memory, have been shown to correlate strongly with fluid intelligence (e.g. Baddeley et al., 1985; Kane & Engle, 2002).

3.2.6. Spatial Orientation (adapted from Logie & Baddeley, 1983)

Participants were presented with a series of screens showing male figures facing away or towards the observer and either upright or inverted. Below each figure were buttons indicating 'left' and 'right.'. Each figure had a blue ball in one hand and a white ball in the other. Participants were to click on the button indicating the hand holding the blue ball. Performance was scored as the number correct in 30 s and ranged from 0 to 41. This task is thought to reflect speed of response as well as spatial ability. Originally devised by Benson and Gedye (1963), versions of this task have been used in a wide range of psychometric test batteries, and it has been shown to correlate with fluid intelligence as well as with other measures of speed of processing (e.g. Carter & Woldstad, 1985; Turnage & Kennedy, 1992).

3.3. Statistical analyses

The Working Memory Span variable was negatively skewed (−1.033); all others were reasonably normally distributed. We reduced the degree of skew of the Working Memory Span variable by squaring it and then standardized all variables in the full sample, in order to place them on the same scale. We then created 14 age groups beginning with ages 18–20 and extending in 5-year increments (21–25, 26–30, etc.) to 81 and over. In the full sample, the covariance matrix had a single eigenvalue of 2.09 greater than 1; the next largest eigenvalue was .85, and this pattern was very consistent throughout the age groups.[2] This indicated the presence of a single dominant general factor, which was our primary interest in this study, regardless of whether other factors might or might not be present. We therefore fit a single-factor model separately to the data for each age group using confirmatory factor analysis as implemented in Mplus (Muthen & Muthen, 1998–2006) and proceeded from there to apply the constraints described above that were needed to test measurement invariance (Meredith, 1993).

To assess relative model fit, we report chi-squared and -2*loglikelihood, Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), Akaike Information Criterion (AIC; Akaike, 1983), and Sample-Size Adjusted Bayesian Information Criterion (BIC; Schwartz, 1978). Because the progressively restrictive tests of measurement invariance create nested models, the differences in chi-square and −2*loglikelihood between pairs of models are distributed as chi-square with the differences in degrees of freedom or numbers of estimated parameters as degrees of freedom. In large samples such as ours, however, these tests are generally too restrictive. That is, they show statistically significant differences in model fit due simply to sample size. AIC and BIC are information-theoretic fit statistics that are not subject to this problem. Moreover, they explicitly recognize model parsimony, particularly BIC, and they tend to be more sensitive than CFI, TLI, and RMSEA. In our application, this means that they tend to indicate preference for models allowing the constraints that indicate measurement invariance, so they provided the strongest tests of failure of measurement invariance and we relied on them most heavily. For larger samples, BIC generally provides better estimates than AIC (Markon & Krueger, 2004). Lower (smaller) fit statistics indicate preferred models for both AIC and BIC, with differences of 10 or more considered substantive for BIC.

[2] Eigenvalues, or characteristic roots, of a matrix summarize the extent to which the variance in the matrix can be consolidated or reduced to underlying dimensions, often the primary purpose of factor analysis, though not in this case. A 5×5 variance-covariance matrix such as we used here generates five eigenvalues that might commonly range in size from 3 to .2, though the presence of no single eigenvalue as high as 2 would not be uncommon. A standard rule of thumb in factor analysis is that the data contain as many factors as there are eigenvalues greater than 1. Many studies have documented that this rule of thumb is biased toward the extraction of too many factors because eigenvalues often cluster just above 1, though it can also fail to identify factors of minor statistical but potentially important conceptual significance. Identifying the optimal number of dimensions in our matrix was not the point of this study. In this case, the best two-factor confirmatory solution fit the data much less well than the one-factor solution and showed even greater measurement variance than the one-factor solution we present. For the two-factor solution, even factor loadings could not be constrained equal and the median correlation between the two factors across the age groups was .82.

Table 1
Descriptive statistics of study variables.

	Age	Education	Feature Binding	Visual Pattern Span	Digit Span	Working Memory Span	Spatial Orientation
1. Ages 18–20 (N = 11,595, 64% female)							
Mean	18.93	3.72	4.71	5.53	53.95	27.48	8.64
SD	.83	.95	1.77	1.57	18.08	9.11	3.51
Skewness	.13	.34	−.53	−.02	−.15	−1.30	.03
2. Ages 21–25 (N = 19,228, 60% female)							
Mean	23.05	5.03	4.61	5.42	56.42	27.28	8.81
SD	1.41	1.31	1.84	1.56	18.29	9.22	3.53
Skewness	−.05	−.50	−.50	.01	−.34	−1.23	−.01
3. Ages 26–30 (N = 17,146, 58% female)							
Mean	27.88	5.47	4.46	5.20	57.00	26.92	8.92
SD	1.43	1.36	1.84	1.56	18.24	9.39	3.44
Skewness	.13	−.75	−.47	.06	−.39	−1.16	−.08
4. Ages 31–35 (N = 12,837, 56% female)							
Mean	32.95	5.33	4.33	5.00	57.32	26.43	8.87
SD	1.41	1.46	1.85	1.52	18.32	9.60	3.36
Skewness	.05	−.53	−.43	.07	−.43	−1.10	−.11
5. Ages 36–40 (N = 9,772, 56% female)							
Mean	37.88	5.09	4.16	4.75	56.52	26.04	8.68
SD	1.42	1.52	1.85	1.52	18.79	9.72	3.25
Skewness	.12	−.26	−.32	.06	−.45	−1.00	−.13
6. Ages 41–45 (N = 7,186, 59% female)							
Mean	42.93	5.00	3.89	4.45	56.02	25.38	8.33
SD	1.41	1.54	1.85	1.47	18.53	10.06	3.10
Skewness	.08	−.14	−.23	.13	−.38	−.88	−.10
7. Ages 46–50 (N = 6,094, 63% female)							
Mean	47.96	4.92	3.64	4.15	55.33	224.57	8.09
SD	1.42	1.53	1.87	1.48	18.57	10.34	3.01
Skewness	.05	−.07	−.14	.10	−.38	−.78	−.06
8. Ages 51–55 (N = 4,737, 65% female)							
Mean	52.88	4.90	3.37	3.91	54.54	23.94	7.95
SD	1.40	1.54	1.82	1.42	18.82	10.45	2.95
Skewness	.13	−.04	−.05	.08	−.33	−.67	−.01
9. Ages 56–60 (N = 3,166, 63% female)							
Mean	57.88	4.82	3.14	3.67	54.18	23.57	7.62
SD	1.38	1.57	1.85	1.39	18.78	10.54	2.88
Skewness	.11	.00	.09	.02	−.33	−.63	−.06
10. Ages 61–65 (N = 1,492, 62% female)							
Mean	62.72	4.74	2.79	3.42	52.43	22.17	7.18
SD	1.39	1.60	1.79	1.43	19.25	10.83	2.80
Skewness	.26	.09	.12	.17	−.27	−.52	.07
11. Ages 66–70 (N = 947, 59% female)							
Mean	67.82	4.46	2.34	3.04	48.72	19.71	6.52
SD	1.40	1.58	1.71	1.34	19.93	11.12	2.79
Skewness	.17	.29	.29	.02	−.23	−.24	.21
12. Ages 71–75 (N = 447, 52% female)							
Mean	72.87	4.55	2.08	2.91	46.06	17.46	6.32
SD	1.40	1.59	1.66	1.37	18.73	11.05	2.68
Skewness	.14	.18	.61	.04	−.03	.04	.04
13. Ages 76–80 (N = 210, 42% female)							
Mean	77.85	4.30	2.10	2.75	46.95	17.18	6.14
SD	1.45	1.75	1.69	1.58	17.49	11.58	2.60
Skewness	.19	.15	.51	.89	−.06	.08	−.04
N							
14. Ages 81 and Over (N = 104, 51% female)							
Mean	84.23	4.09	2.08	2.67	43.85	14.53	6.22
SD	2.89	1.82	1.64	1.57	20.51	11.92	2.96
Skewness	.65	.04	.75	.59	−.19	.44	.74
Total (N = 95,199, 60% female)							
Mean	33.55	4.97	4.09	4.90	55.87	26.15	8.57
SD	12.48	1.48	1.91	1.64	18.50	9.79	3.36
Skewness	.96	−.21	−.35	.04	−.35	−1.03	−.01

4. Results

4.1. Descriptive statistics

Table 1 and Fig. 1 show descriptive statistics for the study variables for each age group. Overall, 60% of participants were female. The female percentage ranged from a low of 42% in the 76–80 age group (the only age group without a female majority) to 65% in the 51–55 age group. The participants tended to be rather young (mean age was 33.55, 12.48 SD) and well educated (mean level of education was 4.97, 1.48 SD, indicating an average close to college educated). The largest age

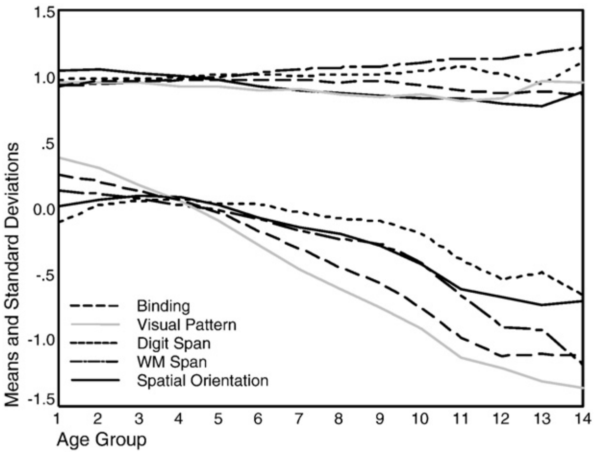

Fig. 1. Normalized test means and standard deviations by test and age group. The lines centered around 0 are means; those centered around 1 are standard deviations.

group was 26–30, and participation fell for each age group beyond that. Consistent with the years required to complete the highest levels of education, average education rose in the age groups until the 26–30 age group. Consistent with general demographic trends, average education fell with age after that. Consistent with many other studies, there were also substantial differences in mean scores on the tests across age groups. These were the subjects of the more extensive analyses that followed. Table 2 shows the correlations among the tests by age group. Though many of the correlations were quite similar across age groups, there were also important differences.

4.2. Tests of measurement invariance of working memory

Table 3 shows the step-by-step results of the tests of measurement invariance of the working memory construct across age groups. The simple one-factor model provided a reasonable description for each age group, establishing configural invariance. Given that BIC was lower (smaller) in the model with factor loadings constrained equal across groups and the other fit statistics gave indications consistent with this, it was also reasonable to consider that factor loadings could be constrained equal across groups, establishing metric invariance. All the fit statistics showed considerable deterioration when we constrained residual variances equal, however, and there was further marked deterioration when we constrained intercepts equal. Thus it was not possible to attain strict measurement invariance across age groups.

There were differences in sex and level of education across the age groups. Any failures of measurement invariance of the working memory construct across education and/or sex could contribute to the failures of measurement invariance with age because there were small correlations between education and test scores (.09 on average across the 5 tests) and between sex and test scores (−.05 on average across the 5 tests, with females coded as 1, males coded as 0). We therefore also tested invariance of measurement of the working memory construct across levels of education and sex. These results are shown in Table 2 as well. For level of education, it was possible to constrain both factor loadings and residual variances equal across groups, establishing metric invariance, but strict measurement invariance could not be established because intercepts could not be constrained equal. Results across sex were analogous to those for age groups: it was possible to constrain factors loadings equal, but not to constrain either residual variances or intercepts equal in males and females. Differences in levels of education and proportions of females across the age groups may thus have contributed to the failures of measurement invariance of a general working memory factor across age groups.

4.3. Evaluating the sources of measurement variance of working memory

Though we did not observe strict measurement invariance of working memory across age groups, it was possible to obtain strict measurement invariance across many different combinations of pairs of age groups, especially between those adjacent to each other, such as ages 20 and under and ages 21–25. It was not, however, possible to constrain either intercepts or residuals equal without loss of model fit over what might be considered the peak adult age range of 18–35. We thus thought it of greater interest to examine the patterns of freely estimated residual variances and intercepts across the age groups in order to develop ideas about how general working memory capacity and more specific abilities might be used at different ages to carry out tasks that were presumably reasonably novel to the participants. Table 4 shows these results.

General patterns were clearly evident. We begin with residual variances, reflecting the extent to which specific aspects of working memory and/or measurement error independent of the general construct accounted for individual differences in task performance. For Feature Binding, the residual variances generally rose gradually with age until ages 46–50 when the residual variance was .66, and then began a somewhat more rapid decline, reaching .44 by ages 81 and over. Visual Pattern Span showed the opposite pattern: its residual variance decreased gradually from .69 at ages 18–20 to .44 at ages 66–70, and then increased. Residual variances for Digit Span increased rather steadily from .71 to 1.01. Working Memory Span residual variances increased from .68 at the youngest ages to .80 in middle age, and then slowly decreased, while residual variances for Spatial Orientation decreased gradually from .92 at ages 20 and under to about .55 in old age. Factor intercepts, reflecting mean levels specific to the individual tests, also showed clear general patterns. Feature Binding intercepts showed continual decreases with age, as did those of Visual Pattern Span and Working Memory Span, though the slopes were obviously different. In contrast, Digit Span and Spatial Orientation intercepts increased until around age 30 before beginning continual decreases throughout the remainder of the lifespan.

All of the patterns of change in both residual variances and intercepts were clear enough that it made sense to fit regression lines to them. Because of the marked differences in sample size, we used regression equations weighted by age group n, thus giving greater weight to the younger age groups with the largest numbers of participants. We considered linear, quadratic, and cubic regression equations. In most cases, quadratic equations fit best. All regressions were highly significant. For residual variances, we used quadratic equations for all but

Table 2
Correlations among study variables.

	1	2	3	4	5	6	7
1. Ages 18–20							
1. Sex	1.00						
2. Education	.06	1.00					
3. Feature Binding	.00	.00	1.00				
4. Visual Pattern Span	−.10	−.01	.29	1.00			
5. Digit Span	−.07	−.01	.26	.23	1.00		
6. Working Mem. Span	.02	.02	.29	.24	.28	1.00	
7. Spatial Orientation	−.07	−.03	.22	.25	.20	.17	1.00
2. Ages 21–25							
1. Sex	1.00						
2. Education	.04	1.00					
3. Feature Binding	.01	.02	1.00				
4. Visual Pattern Span	−.10	.06	.29	1.00			
5. Digit Span	−.07	.09	.25	.21	1.00		
6. Working Mem. Span	.01	.09	.27	.23	.28	1.00	
7. Spatial Orientation	−.06	.02	.21	.25	.20	.18	1.00
3. Ages 26–30							
1. Sex	1.00						
2. Education	.00	1.00					
3. Feature Binding	.00	.06	1.00				
4. Visual Pattern Span	−.12	.08	.29	1.00			
5. Digit Span	−.08	.13	.27	.22	1.00		
6. Working Mem. Span	.02	.12	.28	.23	.30	1.00	
7. Spatial Orientation	−.06	.05	.24	.27	.21	.19	1.00
4. Ages 31–35							
1. Sex	1.00						
2. Education	−.04	1.00					
3. Feature Binding	−.03	.07	1.00				
4. Visual Pattern Span	−.13	.11	.31	1.00			
5. Digit Span	−.10	.12	.28	.23	1.00		
6. Working Mem. Span	−.01	.15	.28	.24	.31	1.00	
7. Spatial Orientation	−.07	.08	.24	.28	.22	.20	1.00
5. Ages 36–40							
1. Sex	1.00						
2. Education	−.03	1.00					
3. Feature Binding	−.05	.09	1.00				
4. Visual Pattern Span	−.11	.11	.31	1.00			
5. Digit recall	−.08	.10	.32	.24	1.00		
6. Working Mem. Span	.00	.14	.31	.24	.33	1.00	
7. Spatial Orientation	−.07	.07	.25	.28	.24	.20	1.00
6. Ages 41–45							
1. Sex	1.00						
2. Education	−.02	1.00					
3. Feature Binding	−.04	.09	1.00				
4. Visual Pattern Span	−.12	.13	.29	1.00			
5. Digit Span	−.07	.13	.28	.22	1.00		
6. Working Mem. Span	.01	.16	.30	.22	.32	1.00	
7. Spatial Orientation	−.07	.07	.22	.26	.24	.19	1.00
7. Ages 46–50							
1. Sex	1.00						
2. Education	−.03	1.00					
3. Feature Binding	−.04	.09	1.00				
4. Visual Pattern Span	−.11	.12	.31	1.00			
5. Digit Span	−.08	.09	.27	.21	1.00		
6. Working Mem. Span	.03	.15	.31	.24	.32	1.00	
7. Spatial Orientation	−.06	.09	.22	.25	.24	.19	1.00
8. Ages 51–55							
1. Sex	1.00						
2. Education	.01	1.00					
3. Feature Binding	−.06	.11	1.00				
4. Visual Pattern Span	−.11	.11	.33	1.00			
5. Digit Span	−.07	.09	.26	.21	1.00		
6. Working Mem. Span	.05	.14	.31	.23	.35	1.00	
7. Spatial Orientation	−.04	.08	.22	.25	.23	.20	1.00
9. Ages 56–60							
1. Sex	1.00						
2. Education	−.04	1.00					
3. Feature Binding	−.04	.15	1.00				
4. Visual Pattern Span	−.12	.13	.31	1.00			

Table 2 (*continued*)

	1	2	3	4	5	6	7
9. Ages 56–60							
5. Digit Span	−.08	.14	.28	.20	1.00		
6. Working Mem. Span	.02	.16	.30	.26	.34	1.00	
7. Spatial Orientation	−.06	.10	.24	.25	.23	.22	1.00
10. Ages 61–65							
1. Sex	1.00						
2. Education	−.04	1.00					
3. Feature Binding	.00	.16	1.00				
4. Visual Pattern Span	−.09	.13	.34	1.00			
5. Digit Span	−.05	.16	.26	.24	1.00		
6. Working Mem. Span	.07	.18	.29	.22	.33	1.00	
7. Spatial Orientation	−.04	.12	.21	.29	.21	.18	1.00
11. Ages 66–70							
1. Sex	1.00						
2. Education	−.12	1.00					
3. Feature Binding	−.01	.18	1.00				
4. Visual Pattern Span	−.12	.10	.34	1.00			
5. Digit Span	−.06	.15	.26	.32	1.00		
6. Working Mem. Span	.03	.15	.34	.30	.37	1.00	
7. Spatial Orientation	−.10	.12	.27	.27	.22	.17	1.00
12. Ages 71–75							
1. Sex	1.00						
2. Education	−.07	1.00					
3. Feature Binding	.07	.13	1.00				
4. Visual Pattern Span	−.11	.11	.28	1.00			
5. Digit Span	−.01	.09	.17	.23	1.00		
6. Working Mem. Span	.13	.13	.22	.29	.32	1.00	
7. Spatial Orientation	−.03	.13	.22	.26	.19	.17	1.00
13. Ages 76–80							
1. Sex	1.00						
2. Education	−.02	1.00					
3. Feature Binding	.07	.09	1.00				
4. Visual Pattern Span	−.22	.06	.45	1.00			
5. Digit Span	−.13	.13	.27	.24	1.00		
6. Working Mem. Span	.05	.21	.48	.35	.29	1.00	
7. Spatial Orientation	−.05	.12	.25	.16	.34	.22	1.00
14. Ages 81 & over							
1. Sex	1.00						
2. Education	.05	1.00					
3. Feature Binding	.08	−.09	1.00				
4. Visual Pattern Span	−.05	.01	.34	1.00			
5. Digit Span	.14	.08	.20	.26	1.00		
6. Working Mem. Span	.11	−.07	.43	.26	.30	1.00	
7. Spatial Orientation	.05	−.27	.24	.32	.13	.21	1.00
Total							
1. Sex	1.00						
2. Education	−.02	1.00					
3. Feature Binding	−.02	.06	1.00				
4. Visual Pattern Span	−.11	.07	.37	1.00			
5. Digit Span	−.08	.12	.27	.22	1.00		
6. Working Mem. Span	.01	.12	.31	.27	.31	1.00	
7. Spatial Orientation	−.06	.06	.25	.29	.22	.20	1.00

Visual Pattern Span, for which a linear equation was sufficient, accounting for 90% of the variance. For Digit Span, Working Memory Span, and Spatial Orientation, the quadratic regression equations accounted for 63%, 81%, and 98% of the variances, respectively. The n-weighted quadratic regression equation for Feature Binding accounted for only 42% of the variance, but an unweighted quadratic regression equation accounted for 88%. Fig. 2 shows the fitted weighted regression lines for the residual variances of the tests, which gives the clearest impression of the differences in the patterns with age among the tests. It does not give any indication, however, of the extent to which the weighted regression lines reproduced the data. Thus Fig. 3 shows both fitted and actual data for the tasks for which the regressions fit best and worst. The greater deviation between the fitted and actual data at the later ages reflected the smaller volumes of data in those age groups. For intercepts, we again used mostly quadratic equations. A linear equation was sufficient for Visual Pattern Span intercepts, accounting for 99% of the variance. Spatial Orientation intercepts required a cubic equation, but it accounted for 98% of the variance. The quadratic equations for Feature Binding, Digit Span, and Working Memory Span accounted for 99%, 86%, and 99% of the variance, respectively. Fig. 4 shows the fitted weighted regression lines for the tests' intercepts, and Fig. 5 shows both fitted and actual data for the tasks for which the regressions fit best and worst.

Table 5 shows the intercepts for the education groups, and the residual variances and intercepts for males and females. It

Table 3
Fit statistics for measurement invariance tests of *g* models.

Age groups	All parameters Free	Fix factor Loadings equal	Also fix Residuals equal	Also fix Intercepts equal
Chi-squared (*df*)	1456.81 (70)	1648.81 (122)	3013.11 (187)	10,383.93 (239)
Log likelihood (# par)	−638,943.70 (210)	−639,044.90 (158)	−639,746.26 (93)	−643,427.54 (41)
AIC	1,278,307.39	1,278,405.80	1,279,678.50	1,286,937.09
Sample size-adjusted BIC	1,279,627.39	1,279,398.94	1,280,263.07	1,287,194.80
		Ok	No	No
Level of education				
Chi-squared (*df*)	1832.85 (35)	1972.75 (59)	2115.98 (89)	3047.14 (113)
Log likelihood (# par)	−634,567.12 (105)	−634,617.69 (81)	−634,662.70 (51)	−635,129.55 (27)
AIC	1,269,344.23	1,269,397.37	1,269,427.42	1,270,313.10
Sample size-adjusted BIC	1,270,335.85	1,269,904.92	1,269,746.97	1,270,482.28
		OK	OK	No
Sex				
Chi-squared (*df*)	1876.78 (10)	1917.63 (14)	2019.93 (19)	3278.90 (23)
Log likelihood (# par)	−629,927.82 (30)	−629,944.21 (26)	−629,972.75 (21)	−630,597.05 (17)
AIC	1,259,915.64	1,259,940.42	1,259,987.51	1,261,228.09
Sample size-adjusted BIC	1,260,103.35	1,260,103.10	1,260,118.91	1,261,334.46
		Ok	No	No

would also have been possible to establish measurement invariance across some combinations of educational groups, but again we believed that it was more informative to show the freely estimated intercepts. For all tests, the lowest educational groups showed markedly poorer average performance than the higher educational groups, though sample sizes were much smaller. The highest overall performance was in college graduates with first degrees: their intercepts were highest for all the tests except Digit Span. Residual variances were generally higher in males than in females, as

Table 4
Residual variances and intercepts of variables from age group memory models.

Age group	*N*	Feature Binding	Visual Pattern Span	Digit Span	Working Memory Span	Spatial Orientation
			Residual Variances			
1. 20 & under	11,595	.591	.691	.712	.680	.922
2. 21–25	19,228	.664	.689	.739	.704	.925
3. 26–30	17,146	.649	.680	.720	.715	.857
4. 31–35	12,837	.645	.635	.715	.722	.806
5. 36–40	9,772	.626	.628	.728	.728	.759
6. 41–45	7,426	.661	.601	.732	.783	.686
7. 46–50	6,094	.669	.601	.745	.794	.647
8. 51–55	4,737	.623	.546	.768	.796	.615
9. 56–60	3,166	.647	.522	.766	.789	.575
10. 61–65	1,492	.604	.534	.815	.796	.551
11. 66–70	947	.528	.438	.851	.719	.545
12. 71–75	447	.565	.505	.810	.690	.508
13. 76–80	210	.434	.650	.658	.666	.482
14. 81 & over	104	.443	.644	1.006	.678	.622
				Intercepts		
1. 20 & under	11,595	.262	.383	−.104	.137	.021
2. 21–25	19,228	.207	.312	.030	.116	.072
3. 26–30	17,146	.132	.181	.062	.078	.104
4. 31–35	12,837	.064	.061	.079	.027	.089
5. 36–40	9,772	−.028	−.092	.036	−.016	.032
6. 41–45	7,426	−.169	−.273	.008	−.080	−.071
7. 46–50	6,094	−.300	−.460	−.029	−.162	−.143
8. 51–55	4,737	−.441	−.603	−.072	−.229	−.184
9. 56–60	3,166	−.561	−.748	−.091	−.267	−.282
10. 61–65	1,492	−.744	−.905	−.186	−.402	−.414
11. 66–70	947	−.981	−1.134	−.386	−.630	−.610
12. 71–75	447	−1.121	−1.213	−.532	−.832	−.671
13. 76–80	210	−1.105	−1.309	−.482	−.827	−.724
14. 81 & over	104	−1.119	−1.357	−.650	−1.007	−.699
Total	95,201					

Note: Test scores were standardized in the full sample. Residual variances and intercepts presented here were not further standardized within groups.

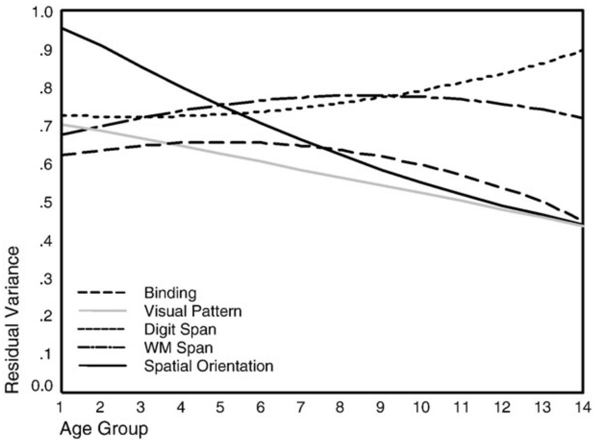

Fig. 2. Residual variances by test and age group, from fitted regression lines.

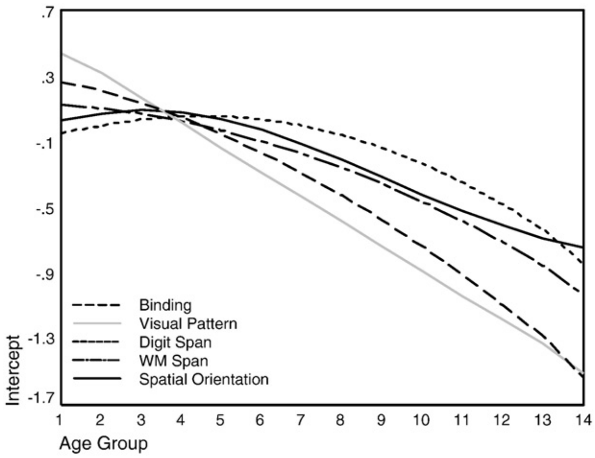

Fig. 4. Intercepts by test and age group, from fitted regression lines.

were intercepts, with the exception of the intercept for Working Memory Span.

5. Discussion

Our goals in this study were systematically to investigate the extent to which the working memory construct can be considered unitary, and to assess whether its structure can be measured consistently across the adult lifespan. This is important in understanding the extent to which age differences in working memory performance reflect differences in a general working memory capacity or differences in relatively modular working memory components that can function relatively independently. It is also important in understanding the ways in which the relative generality or modularity of working memory may change with age; that is, whether working memory may become less differentiated in later adulthood. Our results indicated, first, that a single latent general factor described the data for each age group. There was no evidence that the various aspects of working memory could be considered truly modular even when taking each age group on its own. At the same time, on the order of 75% of the variance in each individual working memory task was independent of the variance in the other tasks. While perhaps a third of this independent variance might be considered outright error variance, most of the tests used here have been shown previously within specific age groups to be robust and reliable measures of the cognitive abilities that they purport to measure, so the large amount of test-specific variance indicated that it would be difficult to consider the working memory construct to be truly unitary as well.

The single-factor structure of our tests of working memory could be considered invariant with age. That is, the individual tests were equivalently reliable indicators of that single factor in all age groups. There were, however, more subtle violations of measurement invariance that undermined the consistency of measurement with age, and in particular the ability to consider mean differences in test scores with age to be attributable to overall working memory. The existence of differences with age in residual variances indicated that there may be important differences with age in the extent to which people must rely on specific, relatively modular abilities or skills. For some tasks, performance reflected progressively greater reliance with age on

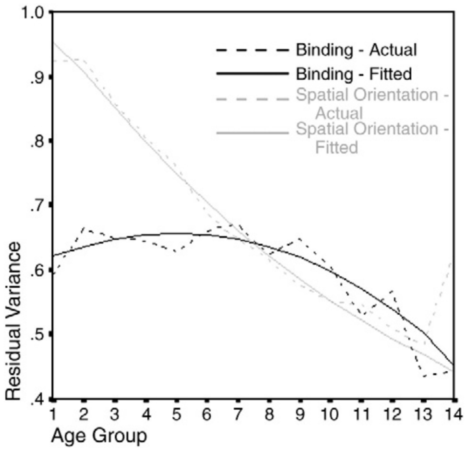

Fig. 3. Fitted and actual data for tasks for which regression lines fit residual variances best and worst.

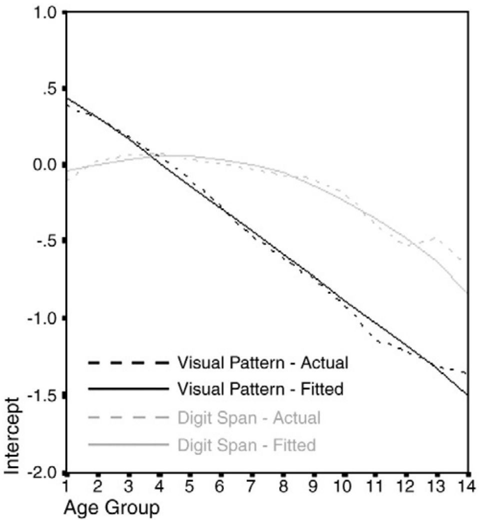

Fig. 5. Fitted and actual data for tasks for which regression lines fit residual variances best and worst.

Table 5
Residual variances and intercepts of variables from education and sex group memory models.

Education groups	N	Feature Binding	Visual Pattern Span	Digit Span	Working Memory Span	Spatial Orientation
				Intercepts		
None	488	−.73	−.65	−.72	−.76	−.53
Primary	563	−.46	−.38	−.48	−.51	−.29
Secondary	22,446	−.05	−.07	−.10	−.12	−.06
Technical college	12,168	−.11	−.13	−.14	−.14	−.07
Other college	16,168	.03	.02	−.07	.00	−.02
Graduate first degree	25,035	.09	.12	.12	.12	.09
Postgraduate degree	16,474	.00	−.01	.15	.11	.05
Total	93,342					
Sex				Residuals		
Male	37,246	.66	.69	.78	.73	.81
Female	55,273	.61	.66	.67	.73	.80
				Intercepts		
Male	37,246	.02	.13	.10	−.01	.08
Female	55,273	−.01	−.08	−.07	.01	−.06
Total	92,519					

Note: Test scores were standardized in the full sample. Intercepts presented here were not further standardized within groups. Reported education levels from people who reported ages too young to have earned them were considered missing. Many participants did not report sex.

the general factor, while for other tasks performance reflected progressively greater reliance on task-specific abilities. Just as importantly for our primary question of interest, these differences could also indicate differences with age in the abilities of the tests to measure their constructs of interest, including the general working memory factor (i.e., reliability). The existence of differences with age in intercepts indicated that age was associated with some difference(s) that affected mean individual task performance levels but not level of general working memory capacity. Possible sources of these differences include development of memory strategies (e.g. Logie, Della Sala, Laiacona, Chalmers, & Wynn, 1996), levels of background knowledge or problem-solving sophistication, and familiarity with task procedures.

5.1. Possible objections

Despite the clear advantage of its very large sample size, our study might be subject to three potential objections that should be acknowledged and evaluated before discussing its results in greater detail. The most obvious is the cross-sectional nature of our study's design. We discuss many of our results as if the attribution of effects to age is clear, but of course age and cohort effects (including the Flynn Effect; Flynn, 1994, 2007) were confounded in these data. Systematic longitudinal studies over the age range reported here and with a consistent battery of tests are impractical, but shorter time periods may be preferred when linking mean differences to aging. However, even those studies can understate age effects because of the existence of test practice effects (Rabbitt et al., 2008; Salthouse, 2009). If the results of longitudinal studies suggesting smaller and later cognitive decline with age than cross-sectional studies (Schaie, 1994) are correct, it may be possible that longitudinal samples would show invariance of measurement with age. In rebuttal, we note that analyses (manuscript in preparation) of three waves of data in the Lothian Birth Cohort 1921 (Deary, Whiteman, Whalley, Fox, & Starr, 2004) showed variance in measurement of general intelligence with age, particularly for the intercepts. Though it is conceivable that this occurred purely because of differential practice effects, it seems unlikely. The Lothian Birth Cohort data spanned the age range from 79 to 87, a much narrower range than in our study but completely contained in our oldest age group. The failure of measurement invariance there is a good indication that measurement invariance would also fail over a broader longitudinal age range, though it is possible that it would not fail for a younger longitundinal age range. It is also possible that measurement invariance could fail for some groups of cognitive tasks but not for others.

Second, our results indicated failure of measurement invariance across educational groups as well as age groups, indicating that some of the variance in measurement with age may be attributable to differences in education, not just to differences in educational credentials attained, but also to cohort differences in educational curricula. But some of the measurement variance may also be due to selection in the sample for general intelligence that increased with age; that is, participants chose to participate in a computer-administered cognitive test battery and the extent to which they were self-selected for above average age-adjusted general intelligence may have varied (most likely increased) with age. Similar comments can be made with respect to sex differences with age in participation, and the possibility of differences in socioeconomic status and other demographic factors that could impact Internet access and experience with computers. Evidence against these alternative accounts however, comes from the observation that there were variances of measurement between just, for example, the 26–30 and 31–35 age groups, for whom cohort differences of all kinds are less credible. Most importantly, whereas such cohort differences limit our ability to draw clear inferences about the age effects, they do not limit our ability to assess the failures of measurement invariance. All of these potential additional reasons for sampling differences that could have created the failures of measurement invariance would also be variables across which measurement of working memory capacity ideally should be invariant, and so these possible sampling differences do not undermine our conclusions in this regard.

Finally, the tasks in our study were self-administered on computers rather than under laboratory conditions. It is not clear what difference this might make, but one possibility involves the related lack of reliability information on the tasks. It is possible that it was error variance rather than dependence on specific abilities that differed with age in this data.

5.2. Patterns of residual variances with age

There were clear patterns of residual variance with age that have important implications for the theoretical conceptions of general and specific working memory abilities. Feature Binding showed some increases in residual variance until about age 40, followed by decreases after that. Mean performance on this test declined across the full age range, but increasingly steeply after age 40 (see also Brockmole et al., 2008; Brown & Brockmole, in press). One interpretation of this is that people were able to use test-specific abilities and skills to offset underlying decline in or ineffectiveness of general working memory capacity or the central executive for task purposes until about age 40, but that after that they relied increasingly but ineffectively on general working memory capacity. Table 6 shows the proportions of variance attributable to general working memory capacity for each test in each age group. It is unlikely that such a pattern would result from changes in test reliability, as the pattern would imply that test reliability increased as performance decreased. Similarly, Visual Pattern Span showed universal decreases in both residual variance and performance, possibly indicating that participants increasingly and ineffectively relied on general working memory capacity for task performance with age. Again, such a pattern would be unlikely to result from changes in test reliability. Digit Span showed universal increases in residual variance with age, and performance increased the longest of any of the tests, until the mid-40's. This suggested that participants were able to make successful use of test-specific abilities and skills to increase or maintain performance at least until middle age. After that, however, participants were apparently less likely to have appropriate test-specific skills and abilities. Increases in residual variance with age would be more likely to be associated with decreasing test reliability, but there is little reason to suspect changes with age in reliability of Digit Span when there is no indication of such changes for the other tests. Working Memory Span showed a pattern very similar to that of Feature Binding, but it extended further, until about age 55. The patterns for Digit Span and for Working Memory Span pattern were consistent with the results of many studies showing better preservation of abilities related to vocabulary (e.g., Cattell, 1971; Horn, 1986). Spatial Orientation showed a pattern very similar to that of Visual Pattern Span, except that presumed increasing reliance on general working memory capacity appeared to have some beneficial effect on performance at least until about age 30 for Spatial Orientation. Thus it is clear that variance in performance on the tests in our battery was not consistently attributable to general or specific memory capacities across the age groups.

Table 6
Proportions of variance attributable to memory factor by age.

Age	Feature Binding	Visual Pattern Span	Digit Span	Working Memory Span	Spatial Orientation
20 & under	.310	.243	.254	.250	.157
21–25	.288	.240	.244	.244	.161
26–30	.303	.246	.259	.251	.180
31–35	.314	.261	.270	.255	.192
36–40	.338	.265	.295	.266	.191
41–45	.300	.253	.270	.248	.195
46–50	.301	.258	.261	.248	.195
51–55	.316	.271	.257	.255	.201
56–60	.311	.269	.256	.263	.218
61–65	.311	.292	.247	.233	.206
66–70	.339	.339	.266	.275	.208
71–75	.251	.274	.208	.243	.201
76–80	.446	.293	.261	.331	.194
81 & over	.395	.285	.174	.315	.192

There was a marked difference between the patterns of residual variance for tests that differentially rely on verbal and visuospatial abilities. Residuals for Feature Binding, Visual Pattern Span, and Spatial Orientation showed decreases across the later lifespan whereas those for Digit Span and Working Memory Span either remained relatively stable or increased. Consistent with many other studies, the relative similarities in the patterns with age among the verbally oriented tasks and among the visuospatially oriented tasks, along with the differences between these two sets of patterns, suggest a basic verbal-visuospatial division that is consistent with working memory comprising multiple, domain-specific resources (e.g., Baddeley, 2007; Baddeley & Logie, 1999; Johnson & Bouchard, 2007a,b; Logie & van der Meulen, 2009; Saito, Logie, Morita, & Law, 2008), that also show different age-related trajectories (Logie & Maylor, 2009) rather than a domain-general system (e.g. Barrouillet et al., Cowan, 2005). A possible caveat is that these differential age trajectories could reflect cohort differences in experience with verbal and visuospatial kinds of problems such as would be the case if educational curricula and recreational activities have increasingly emphasized visuospatial tasks over the last 50 years or so. This latter explanation is, however, less convincing as an account of the very different patterns of change for Digit Span and Visual Pattern Span between the large participant groups that are relatively close in age and would have had very similar educational and recreational experiences such as between 21–25 and 26–30, or between 31–35 and 36–40.

Overall, the patterns of decreasing residual variance with age for most of the tests tended to support the dedifferentiation hypothesis in adulthood. Stronger support for this hypothesis, however, would have been provided by increases in test loadings on the general working memory factor with age. Moreover, in the very oldest age groups there was some evidence for increases in residual variances, contradicting the dedifferentiation hypothesis. These indications should be considered tentative because the sample sizes were smaller in those age groups than in the other groups. Nevertheless, the numbers in those older groups ($n = 1698$ for age >65 years) were larger than in many previous studies of cognitive changes with age.

5.3. Patterns of differences in intercepts with age

Digit Span and Spatial Orientation were the two tests that showed increases in performance at least until about age 30. Performance on Spatial Orientation was also the best maintained even once decline with age began. This might reflect the fact that this was the task that had the greatest probability of generating correct answers completely by chance, as participants had only to choose between left and right. Working Memory Span probably relied most on verbal knowledge and, after Digit Span and Spatial Orientation, performance on it was best maintained with age. In contrast, tasks that rely more on visual processing, namely Visual Pattern Span and Feature Binding, showed the sharpest decreases in performance with age. These might have been the tests with which participants overall were least likely to have some experience and the most likely to show cohort differences in experience. However, again this is not particularly convincing as the sole explanation because there was measurement variance between large groups adjacent in age, such as 26–30 and 31–35, and 45–50 and 51–55, for which cohort differences were less likely. The more dramatic declines in visual working memory tasks in this large sample lend support to lab-based studies suggesting that age-related working memory deficits for visuospatial material are more severe than those observed for verbal material (Jenkins, Myerson, Hale, & Fry, 1999; Jenkins, Myerson, Joerding, & Hale, 2000; Leonards, Ibanez, & Giannakopoulos, 2002; Myerson, Hale, Rhee, & Jenkins, 1999), and they counter observations to the contrary (Park et al., 2002; Salthouse, 1995). The results also are consistent with the suggestion that there is greater rather than lesser differentiation among cognitive abilities with age (e.g. Park et al., 2002).

5.4. Conclusions: implications for understanding working memory and its specific components and for future research

At all ages in our data, a single factor operated strongly across all five working memory tasks, indicating that some form of general capacity or central executive contributed to all of them, though to varying degrees. Moreover, the extent to which the five tasks either contributed to or were represented by this general factor *was* consistent at all ages. Though they should be replicated in other studies subject to different limitations, the failures of measurement invariance across age in this study tended to undermine rather than support the dedifferentiation hypothesis regarding cognitive changes with age: the general working memory factor did not appear to contribute consistently to test performance with age, as evidenced by the differences in residual variances and intercepts with age. These differences suggest that, not only do relatively specific abilities change in different ways with age, but people make use of their general and specific working memory capacities differently with age or developmental experiences. Exploration of these themes in future research could help to develop both more effective educational techniques and targeted interventions to help older adults cope with tasks that are most likely to be affected by cognitive decline.

Acknowledgements

Wendy Johnson holds a Research Council of the United Kingdom Fellowship. She and Robert Logie are members of the University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross-council Lifelong Health and Wellbeing Initiative (Grant #G0700704/84698). James Brockmole is an Honorary Fellow of the University of Edinburgh, and was a member of the above research center at the time this research was carried out. Funding of that Centre from the BBSRC, EPSRC, ESRC and MRC is gratefully acknowledged. The authors are grateful to the British Broadcasting Corporation (BBC) Science Division, especially Becky Palmer and Stephen Mather, for their collaboration on this project. Correspondence concerning this article should be addressed to Wendy Johnson or Robert Logie, Centre for Cognitive Ageing and Cognitive Epidemiology and Department of Psychology, 7 George Square, Edinburgh EH8 9JZ, UK. E-mail addresses: wendy.johnson@ed.ac.uk, rlogie@staffmail.ed.ac.uk.

References

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute*, *50*, 277−290.
Alloway, T. P., & Gathercole, S. E. (2006). *Working Memory and Neurodevelopmental Disorders.* Hove, UK: Psychology Press.
Anstey, K. J., Hofer, S. M., & Luscez, M. A. (2003). Cross-sectional and longitudinal patterns of dedifferentiation in later-life cognitive and sensory function: The effects of age, ability, attrition, and occasion of measurement. *Journal of Experimental Psychology: General*, *132*, 470−487.
Baddeley, A. D. (2007). *Working Memory in Thought and Action.* Oxford, UK: Oxford University Press.
Baddeley, A. D., Kopelman, M. D., & Wilson, B. A. (Eds.). (2002). *Handbook of Memory Disorders*. Chichester, UK: Wiley.
Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake, & P. Shah (Eds.), *Models of Working Memory* (pp. 28−61). New York: Cambridge University Press.
Baddeley, A. D., Logie, R. H., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, *24*, 119−131.
Balinsky, B. (1941). An analysis of the mental factors of various age groups from nine to sixty. *Genetic Psychology Monographs*, *23*, 191−234.
Baltes, P. B., & Lindenberger, U. (1997). Emergence of a powerful connection between snesory and cognitive functions across the adult lifespan: A new window to the study of cognitive aging? *Psychology and Aging*, *12*, 12−21.
Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and Cognitive Load in Working Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*, 570−585.
Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General*, *132*, 71−92.
Benson, A. J., & Gedye, J. L. (1963). Logical processes in the resolution of orientational conflict. *RAF Institute of Aviation Medicine Report 259, Farnborough, UK*.
Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the lifespan. *Intelligence*, *20*, 309−328.
Birnbaum, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, *55*, 803−832.
Brand, C. (1987). Bryter still and bryter? *Nature*, *328*, 110.
Brockmole, J. R., Parra, M. A., Della Sala, S., & Logie, R. H. (2008). Do binding deficits account for age-related decline in visual working memory? *Psychonomic Bulletin and Review*, *15*, 543−547.
Brown, L. A. & Brockmole, J. R. (in press). The role of attention in binding visual features in working memory: Evidence from cognitive ageing. *Quarterly Journal of Experimental Psychology.*
Cadman, C. (Producer). (2006, August 9). How to improve your memory (Television broadcast). London, UK: British Broadcasting Corporation.
Carter, R., & Woldstad, J. (1985). Repeated measurements of spatial ability with the Manikin test. *Human Factors*, *27*, 209−219.
Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.
Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547−552.
Conway, A. R. A., Jarrold, C., Kane, M. J., Miyake, A., & Towse, J. N. (2007). *Variation in Working Memory.* Oxford, UK: Oxford University Press.
Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.
Cutler, S. J., Hendricks, J., & Guyer, A. (2003). Age differences in home computer availabilty and use. *Journal of Gerontology Series B: Psychological Science and Social Sciences*, *58*, S271−S280.
Deary, I. J., Whiteman, M. C., Whalley, L. J., Fox, H. C., & Starr, J. M. (2004). The impact of childhood intelligence on later life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology*, *86*, 132−147.
DeFrias, C. M., Lovden, M., Lindenberger, U., & Nilsson, L. G. (2007). Revisiting the dedifferentiation hypothesis with longitudinal multi-cohort data. *Intelligence*, *35*, 381−392.

Della Sala, S., Darling, S., & Logie, R. H. (2010). Items on the left are better remembered. *Quarterly Journal of Experimental Psychology*, *63*, 848–855.

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. C., & Logie, R. H. (2001). Processing and storage in working memory span. *Quarterly Journal of Experimental Psychology*, *54A*, 31—48.

Flynn, J. R. (1994). IQ gains over time. In R. J. Sternberg (Ed.), *Encyclopedia of Human Intelligence* (pp. 617—623). New York: Macmillan.

Flynn, J. R. (2007). *What is intelligence?* Cambridge: Cambridge University Press.

Friedman, N. P., Miyake, A., Young, S. E., DeFries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of Experimental Psychology: General*, *137*, 201—225.

Garrett, H. E. (1938). Differentiable mental traits. *Psychological Record*, *2*, 259—298.

Gow, A., Johnson, W., Pattie, A., Whiteman, M., Whalley, L. J., Starr, J. M., et al. (2008). Mental ability in childhood and cognitive aging. *Gerontology*, *54*, 177—186.

Hofer, S. M., Horn, J. L., & Eber, H. W. (1997). A robust five-factor structure of the 16PF: Strong evidence from independent rotation and confirmatory factorial invariance procedures. *Personality and Individual Differences*, *23*, 247—269.

Horn, J. L. (1986). Intellectual ability concepts. In R. L. Sternberg (Ed.), *Advances in the Study of Human Intelligence* (pp. 25—77). Hillsdale, NJ: Erlbaum.

Hunt, W. L. (1949). The relative rates of decline of Wechsler-Bellevue “hold” and “don't hold” tests. *Journal of Consulting Psychology*, *13*, 440—443.

Jacobs, J. (1887). Experiments on “Prehesion”. *Mind*, *12*, 75—79.

Jenkins, L., Myerson, J., Hale, S., & Fry, A. F. (1999). Individual and developmental differences in working memory across the life span. *Psychonomic Bulletin & Review*, *6*, 28—40.

Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuo-spatial cognition is more age-sensitive than verbal cognition. *Psychology and Aging*, *15*, 157—175.

Jensen, A. R. (1998). *The g Factor. (1998)*. Wesport, CN: Praeger.

Johnson, W., & Bouchard, T. J. (2007a). Sex differences in mental ability: g masks the differences on which they lie. *Intelligence*, *35*, 25—39.

Johnson, W., & Bouchard, T. J. (2007b). Sex differences in mental ability: A proposed means to link them to brain structure and function. *Intelligence*, *35*, 197—209.

Jones, H. E., & Conrad, H. S. (1933). The growth and decline of intelligence: A study of homogeneous groups between the ages of ten and sixty. *Genetic Psychology Monographs*, *13*, 223—275.

Juan-Espinosa, M., Garcia, L. F., Escorial, S., Rebollo, I., Colom, R., & Abad, F. J. (2002). Age dedifferentiation hypothesis: Evidence from the WAIS III. *Intelligence*, *30*, 1—14.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin and Review*, *9*, 637—671.

Leonards, U., Ibanez, V., & Giannakopoulos, P. (2002). The role of stimulus type in age-related changes of visual working memory. *Experimental Brain Research*, *146*, 172—183.

Li, S. -C., Lindenberger, U., & Sikstrom, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, *5*, 479—486.

Logie, R. H. (2006, August 5-11). How to improve your memory. *Radio Times*, 20—25.

Logie, R. H., & Baddeley, A. D. (1983). A trimix saturation dive to 660 metres: Studies of cognitive performance, mood, and sleep quality. *Ergonomics*, *26*, 359—374.

Logie, R. H., Brockmole, J. R., & Vandenbroucke, A. (2009). Bound feature combinations in visual short-term memory are fragile but influence long-term learning. *Visual Cognition*, *17*, 160—179.

Logie, R. H., & Della Sala, S. (2005). Disorders of visuo-spatial working memory. In A. Miyake, & P. Shah (Eds.), *Handbook of Visuospatial Thinking* (pp. 81—120). New York: Cambridge University Press.

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wynn, V. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory and Cognition*, *24*(3), 305—321.

Logie, R. H., Della Sala, S., MacPherson, S. E., & Cooper, J. (2007). Dual task demands on encoding and retrieval processes: Evidence from healthy adult ageing. *Cortex*, *43*, 159—169.

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

Logie, R. H., & Pearson, D. G. (1997). The inner eye and the inner scribe of visuo-spatial working memory: Evidence from developmental fractionation. *European Journal of Cognitive Psychology*, *9*, 241—257.

Logie, R. H., & van der Meulen, M. (2009). Fragmenting and integrating visuo-spatial working memory. In J. R. Brockmole (Ed.), *Representing the Visual World in Memory* (pp. 1—32). Hove, UK: Psychology Press.

Lovden, M., Ghisletta, P., & Lindenberger, U. (2004). Cognition in the Berlin Aging Study (BASE): The first 10 years. *Aging Neuropsychology and Cognition*, *11*, 104—133.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenburgh, G. J. (2003). On the relationship between within- and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*, 533—566.

Markon, K. E., & Krueger, R. F. (2004). An empirical comparison of infomation-theoretic selection criteria for multivariate behavior genetic models. *Behavior Genetics*, *34*, 593—610.

Maylor, E. A. & Logie, R. H. (2010). A Large-Scale Comparison of Prospective and Retrospective Memory Development from Childhood to Middle-Age. *Quarterly Journal of Experimental Psychology*, *63*, 442–451.

Maylor, E. A., & Reimers, S. (2007). Cognitive aging research using the Internet: Four case studies. In Garriga-Trillo (Ed.), *Converging Research on Predictors of Cognitive Impairment and Neurodegenerative Diseases* (pp. 29—41). Madrid: Publidisa.

Maylor, E. A., Reimers, S., Choi, J., Collaer, M. L., Peters, M., & Silverman, I. (2007). Gender and sexual orientation differences in cognition across adulthood: Age is kinder to women than to men regardless of sexual orientation. *Archives of Sexual Behavior*, *36*, 235—249.

McGue, M., & Christensen, K. (2002). The heritability of level and rate-of-change in cognitive functioning in Danish twins aged 70 and older. *Experimental Aging Research*, *28*, 435—451.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika*, *58*, 528—543.

Mitchell, K. J., Johnson, M. K., D'Esposito, M., Raye, C. L., & Mather, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology and Aging*, *15*, 527—541.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49—100.

Molenaar, D., Dolan, C. V., & Wicherts, J. M. (2009). Power to detect sex differences in IQ test scores using multi-group covarariance and means structural analysis. *Intelligence*, *37*, 396–404.

Muthen, L. K., & Muthen, B. O. (1998-2006). *Mplus: Statistical Analysis with Latent Variables, User's Guide*, 4th Edition. Los Angeles: Muthen & Muthen.

Myerson, J., Hale, S., Rhee, S. H., & Jenkins, L. (1999). Selective interference with verbal and spatial working memory in young and older adults. *Journal of Gerontology*, *54B*, 161—164.

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.

Perfect, T., & Maylor, E. A. (2000). Rejecting the dull hypothesis. In T. J. Perfect, & E. A. Maylor (Eds.), *Models of Cognitive Aging*. Oxford, UK: Oxford Psychology Press.

Rabbitt, P., Lunn, M., Wong, D., & Cobain, M. (2008). Age and ability affect practice gains in longitudinal studies of cognitive change. *Journal of Gerontology: Psychological Sciences*, *63B*, P235—P240.

Reips, U. -D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, *49*, 243—256.

Robins, R. W., Trzesniewski, K. H., Gosling, S. D., & Potter, J. (423-434) (2002). Global self-esteem across the lifespan. *Psychology and Aging, 17*.

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with Kanji materials. *Journal of Memory and Language*, *59*, 1—17.

Salthouse, T. A. (1995). Differential age-related influences on memory for verbal-symbolic information and visual-spatial information. *Journal of Gerontology*, *50B*, 193—201.

Salthouse, T. A. (2006). Mental exercise and mental aging. *Perspectives on Psychological Science*, *1*, 68—87.

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

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, *49*, 304—313.

Schwartz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461—464.

Selwyn, N., Gorard, S., Furlong, J., & Madden, L. (2003). Older adults' use of information and communications technology in everyday life. *Ageing and Society*, *23*, 561—582.

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

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, *15*, 201—293.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review*, *38*, 406—427.

Treisman, A. M. (2006). Object tokens, binding, and visual memory. In H. D. Zimmer, A. Mecklinger, & U. Lindenberger (Eds.), *Handbook of binding* *and memory: Perspectives from cognitive neuroscience* (pp. 315−338). Oxford: Oxford University Press.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the lifespan. *Developmental Psychology*, *45*, 1097−1118.

Tucker-Drob, E. M., & Salthouse, T. A. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and Aging*, *23*, 453−460.

Turnage, J. J., & Kennedy, R. S. (1992). The development and use of a computerized human performance test battery for repeated-measures applications. *Human Performance*, *5*, 265−301.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635−654.

Vallar, G., & Shallice, T. (1990). *Neuropsychological Impairments of Short-term Memory*. Cambridge, UK: Cambridge University Press.

Wicherts, J. M. (2007). *Group Differences in Intelligence Test Performance.* Amsterdam: University of Amsterdam Dissertation.

Wicherts, J. M., Dolan, C. V., & Hesson, D. J. (2005). Stereotype threat and group differences in test performance. *Journal of Personality and Social Psychology*, *89*, 696−716.