



Test Review

Wechsler Memory Scale® – Fourth UK Edition

(WMS® IV UK)

The British Psychological Society © 2012. All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Permission has been granted to the distributor / publisher named above to distribute copies of this review in paper or PDF file format so long as such copies are not amended or changed in any way from the original version published by the BPS.

Wechsler Memory Scale ® – Fourth UK Edition (WMS® IV UK)

Reviewers: Joanna Horne
& Angus McDonald

Consulting Editor: L. A. Marshall

Senior Editor: Iain Coyne

GENERAL INFORMATION AND DESCRIPTION OF THE INSTRUMENT

Test Name: Wechsler Memory Scale ® – Fourth UK Edition

Authors of the original test:

David Wechsler

License for Distribution Issued To:

Pearson Assessment

Authors of the local adaptation:

David Wechsler

Local test distributor / publisher:

Pearson Assessment

Publisher of the original version of the test: Pearson Assessment (US)

Date of Publication of Current Review/Edition: 2011

Date of Publication of Adaptation for local use: 2010

Original Test Name: Wechsler Memory Scale ® – Fourth Edition

Date of Publication of the Original Test: 2009

Date of Current Review: 2011

ISBN: 9781854336064

General Description of Test: The Wechsler Memory Scale®–Fourth UK Edition (WMS–IVUK) is an individually administered battery designed to assess various memory and working memory abilities in individuals aged 16–90. In addition to the assessment of memory functioning, the WMS®–IVUK contains an optional Brief Cognitive Status Exam designed to screen for significant cognitive dysfunction by providing an indication of current cognitive status. Two batteries are contained within the WMS®–IVUK: an Adult battery for individuals aged 16–69, and a shorter Older Adult battery developed for use with individuals aged 65–90. The Older Adult battery has been designed to take less time and so reduce fatigue effects and consequently improve the psychometric properties of the tests with older populations.

WMS®–IVUK assesses aspects of memory relevant to clinical populations including those with neurological, psychiatric and developmental disorders.

The WMS®–IVUK contains a total of seven subtests: three subtests retained (and amended) from the WMS®–IIIUK (Logical Memory, Verbal Paired Associates and Visual Reproduction) and four new subtests (Brief Cognitive Status Exam, Designs, Spatial Addition and Symbol Span). The changes were designed to improve the administration, scoring and psychometric properties of the subtests. The subtests are as follows. Please note that the delayed condition subtests are administered 20–30 minutes after the first subtest.

Logical Memory I (LMI) assesses narrative memory under a free recall condition. Logical Memory II (LMII) is a delayed condition assessing long–term narrative memory with free recall and recognition tasks.

Verbal Paired Associates I (VPAI) assesses verbal memory for associated word pairs. Verbal Paired Associates II (VPAIL) is a delayed condition assessing long–term recall for verbally paired information with cued recall and recognition tasks and includes a free recall task.

Designs I (DEI) assesses spatial memory for unfamiliar visual material. Designs II (DEII) is a delayed condition assessing long–term spatial and visual memory with free recall and recognition tasks.

Visual Reproduction I (VPI) assesses memory for non–verbal visual stimuli. Visual Reproduction II (VPII) is a delayed condition assessing long–term visual–spatial memory with free recall and recognition tasks and includes a direct copy task.

Spatial Addition (SA) assesses visual–spatial working memory using a visual addition task.

Symbol Span (SSP) assesses visual working memory using novel visual stimuli.

The optional Brief Cognitive Status Exam (BCSE) assesses a variety of cognitive functions including orientation to time, mental control, clock drawing, incidental recall, automaticity and inhibitory control and verbal production.

The primary subtests: LM, VPA, DE, VR, SA and SSP are used to derive index scores. The five ‘index scores’ are Auditory Memory (AMI), Visual Memory (VMI), Visual Working Memory (VWMI), Immediate Memory (IMI) and Delayed Memory (DMI). VWMI is not available from the Older Adult battery.

As well as index scores, ‘contrast scores’ are also available. These scores allow users to compare performance on “higher– to lower–level cognitive functions” or to differentiate between different modalities of presentation. Finally, ‘process scores’ are designed to provide a more detailed level of analysis on subtest performance, such as a comparison between the recall of design content and the recall of spatial location in the DE subtest.

The WMS®–IVUK comes with ‘writer’ software which allows a range of reports to be produced:

The Score Report – a series of tables and graphs representing the scores. The tables and graphs are also included in the Interpretive Report.

Interpretive Report – a narrative interpretation of the WMS–IV scores, as well as relevant details from the examinee history screens.

Clinical Review – a narrative analysis of the background data.

Client/Caregiver Report – a narrative summary of the WMS®–IVUK scores.

Classification

Content Domains:

Memory

Intended or main area(s) of Use:

Psycho–clinical

Psycho–neurological

Forensic

Educational

Work and Occupational

Counselling, Advice, Guidance, and Career Choice

Intended mode of use (conditions under which the instrument was standardised and validated):

Supervised and controlled administration. Test administration under the control of a qualified administrator or proctor

Description of the populations for which the test is intended:

Adults from 16 to 90 years of age. There are two test batteries. One for Adults aged between 16 and 69 years and one for Older Adults aged between 65 and 90 years.

Number of scales and brief description of the variable or variables measured by the test:

In the Adult Battery there are 6 subtests within 5 indices:

Auditory Memory Index (AMI): Logical Memory (LM) I and II; Verbal Paired Associates (VPA) I and II.

Visual Memory Index (VMI): Designs (DE) I and II and Visual Reproduction (VR) I and II.

Visual Working Memory Index (VWMI): Spatial Addition (SA) and Symbol Span (SSP).

Immediate Memory Index (IMI): L M I, VPA I, DE I and VR I.

Delayed Memory Index (DMI): LM II, VPA II, DE II and VR II.

In the Older Adult Battery all subtests except for SA and both DE tests are included.

There is an optional Brief Cognitive Status Exam (BCSE) which assesses global cognitive functioning.

Items format:

Open

Number of test items:

LM I: 2 stories (25 items / story); LM II – recall 2 stories (25 items / story) and recognition 2 stories (15 items / story); VPA I – 4 lists (14 items / list); VPA II – recall 14 items, recognition 40 items; DE I – 4 items; DE II – recall 4 items, recognition 12 items; VR I – 5 items; VR II – recall 5 items, recognition 7 items, copy 5 items; SA – 24 items (max.); SSP – 26 items (max.); BCSE – 12 items.

Administration modes:

- Interactive individual administration

Response mode:

Oral interview

Manual operations

Paper and pencil

Time:

Preparation: 5 minutes

Administration: Information on the time required by the normative and special group samples is available in Table 2.2 of the Administration and Scoring Manual. This suggests that the Adult Battery takes between 80 and 116 minutes to administer which reduces to between 50 and 73 minutes for the shorter Older Adult Battery. The data suggests that times increase slightly for the special group samples.

Scoring: 5 minutes (using software).

Analysis: 10 minutes.

Feedback: 20 minutes.

Different forms of the test:

No

Measurement and Scoring

Scoring procedure for the test:

Computer scoring manual entry of responses from the paper response form
Complex manual scoring, requiring training in the scoring of the test

Scores:

A total raw score for each subtest is obtained by the sum of the item scores. These are converted into Scaled Scores (mean of 10 and standard deviation of 3) using the conversion table. These are summed for each of the five Indices. Each Index Score is, in turn, converted into a Percentile Rank, with 90% or 95% Confidence Intervals. In addition process scores are available for some subtests where more specific aspects of performance are examined. Finally, contrast scale scores are calculated in which a candidate's score on one scale is adjusted based on performance on another scale.

Score transformation for standard scores:

- Normalised – scores obtained by use of normalisation look-up table

Scales Used:

Percentile-based scores:

- Centiles

Standard scores:

- Subtest scaled scores with a mean of 10 and a standard deviation of 3. Index scores have a mean of 100 and a standard deviation of 15.

Computer-Generated Reports

Are computer generated reports available with the instrument?:

- Yes

Number of Computer Generated Reports available:

2

Do distributors offer a service to correct and/or develop computer generated reports?:

- No

Report Name:

The Score Report

Brief description of Reports..	
Media:	• Integrated text and graphics
Complexity:	• Medium
Report Structure:	• Scale based
Sensitivity to context:	• One version for all contexts
Clinical–actuarial:	• Based on empirical/actuarial relationships
Modifiability:	• Limited modification
Degree of 'finish':	• Publication quality
Transparency:	• Clear linkage between constructs, scores and text
Style and tone:	• Guidance
Intended recipients:	• Qualified test users • Third parties

Report Name:

The Interpretive Report

Brief description of Reports..	
Media:	• Integrated text and graphics
Complexity:	• Medium
Report Structure:	• Construct based
Sensitivity to context:	• One version for all contexts
Clinical–actuarial:	• Based on clinical judgement of group of experts
Modifiability:	• Limited modification
Degree of 'finish':	• Publication quality
Transparency:	• Clear linkage between constructs, scores and text
Style and tone:	• Guidance
Intended recipients:	• Qualified test users • Third parties

Supply Condition and Costs

Documentation provided by the distributor as part of the test package:

User manual
Technical (psychometric) manual

Methods of publication:

Paper
PC – CD/ROM

Start-up costs:

A complete kit costs £804 (£884.83 inc VAT) This includes:

1 Administration Manual

Stimulus Book 1

Stimulus Book 2

1 pack of 25 response booklets

1 pack of 25 Adult battery record forms

1 pack of 25 Older Adult battery record forms

WMS-IV US Technical Manual

Block Design Set

WAIS/WMS online training

1 Memory Grid

1 Scoring Template in envelope

1 Design and Spatial Addition cards

WAIS-IV/WMS-IV UK Scoring Software and Report Writer: £557 (£668.40 inc VAT)

Recurrent costs:

Response Booklets: pack of 25 £42 (£50.40 inc VAT)

Adult Battery Record Forms: pack of 25 £74.50 (£89.40 inc VAT)

Older Adult Battery Record Forms: pack of 25 £57 (£68.40 inc VAT).

Administration and Scoring Manual: £119.50 (£119.50 inc VAT).

Prices for a report generated by user installed software:

Not applicable

Prices for reports:

Not applicable

Prices for a report generated by the Internet services:

Not applicable

Prices for other bureau services:

Not applicable

Test-related qualifications required by the supplier of the test:

- Accreditation in general achievement testing: measures of maximum performance in attainment
- Accreditation in general ability and aptitude testing: measures of maximum performance in relation to potential for attainment

Professional qualifications required for use of the test:

- Practitioner psychologist with qualification in the relevant area of application
- Use of this test requires professional post graduate qualification and training in Clinical or Educational Psychology. They are also available to Chartered Occupational Psychologists and individuals with a relevant PhD.

Evaluation of Test Materials

Key to symbols:

★	Inadequate
★★	No longer used
★★★	Adequate/Reasonable
★★★★	Good
★★★★★	Excellent
[N.r.i.o.r]	(for updates only) Item was not rated in original review

Quality of the explanation of the rationale, the presentation and the quality of information provided:

★★★★★

Overall rating of the Quality of the explanation of the rationale:	★★★★★
i) Theoretical foundations of the constructs:	★★★★★
ii) Test development procedure:	★★★★★
iii) Thoroughness of the item analyses and item analysis model:	★★★★★
iv) Explanation of content validity:	★★★★★
v) Summary of relevant research:	★★★★★
Adequacy of documentation available to the user (user and technical manuals, norm supplements etc):	★★★★★
i) Rationale:	★★★★★
ii) Development:	★★★★★
iii) Standardisation:	★★★★★
iv) Norms:	★★★★★
v) Reliability:	★★★★★
vi) Validity:	★★★★★
Quality of the Procedural instructions provided for the user:	★★★★★
i) For test administration:	★★★★★
ii) For test scoring, norming etc:	★★★★★

iii) For interpretation and reporting:	★★★★★
iv) For providing feedback and debriefing test takers and others:	0 stars
v) For providing good practice issues on fairness and bias:	★★★★
vi) Restrictions on use:	★★★★★
vii) References and supporting materials:	★★★★★
Quality of the materials:	★★★★★
i) General quality of test materials (test booklets, answer sheets, test objects, software, etc):	★★★★★
ii) Test quality of the local adaptation (if the test has been translated and adapted into the local language):	★★★★★
iii) Ease with which the test taker can understand the task:	★★★★★
iv) Ease with which responses or answers can be made by the test taker:	★★★★★
v) Quality of the items:	★★★★★

Reviewer's comments on the documentation (comment on rationale, design, test development and acceptability):

The documentation provided is extremely comprehensive, clearly presented and laid out in a logical, easy-to-follow way. The Technical and Interpretive Manual begins with a good explanation of the concepts of memory and learning. The rationale, test development procedure, content validity and relevant research are clearly explained. Presentation of this material aids understanding of the rationale behind the tests and how they were structured into the Adult and Older Adult batteries. The decisions taken during the development of the test and the reasons for these are also clearly given, allowing users to understand the reasoning behind the choice of tasks included in the subtests.

There is clear and detailed information regarding the standardization process, the norms and evidence of reliability and validity. The technical manual provides information regarding the items added and removed from each subtest and gives details of the new subtests. However, specific information regarding the item analysis for each subtest is not given. For the UK version of the test, modifications were made to 14 items within three of the subtests in order to change American words and phrases to ones more appropriate for British examinees.

The Administration and Scoring Manual contains very clear and detailed step-by-step guidelines, although there is no explicit advice on dealing with candidates' questions or problem situations. Only the visual working memory subtests require the use of start points, reverse and discontinue rules which make administration of the remaining tests more straightforward as all examinees start with the same item and complete all items.

Clear and detailed information is provided regarding scoring including extensive guidance on how to score the open-response questions, and this should ensure reliable assessment of these subtests. Information is also provided on using norm tables as well as interpreting and reporting the results. However, the manual does not contain illustrative examples or case studies or advice on providing feedback or debriefing test takers.

There is no information regarding group differences, for example, gender and ethnicity, although the manual does discuss suitability and fairness issues with regard to testing 65–69 year olds, re-testing examinees and testing examinees with special needs, for example, physical, language or sensory

limitations.

The quality of the test materials is excellent being well designed and very professionally presented. The UK item modifications are appropriate and the clear instructions and use of sample items make the test easy for test takers to understand and respond to. The manuals are both spiral bound which makes them easy to use and the Administrative and Scoring Manual comes with a robust cover, presumably designed to withstand regular use.

However it would be helpful if more prominence was given to UK validation study. Currently it is the penultimate Appendix in the Administration and Scoring Manual whereas other technical information is in the separate Technical and Interpretive Manual. This is problematic for the UK user who will want to evaluate the appropriateness of using this US developed test with UK populations. It would be useful for this important information to be clearly signposted in the introduction to the test in both manuals.

Norms, Validity & Reliability

Evaluation of technical information:



Norms or reference group information:

Overall adequacy:	★★★★★
i) Appropriateness for local use, whether local or international norms:	★★★★★
ii) Appropriateness for intended applications:	★★★★★
iii) Sample sizes:	★★★★★
iv) Procedures used in sample selection:	Representative of the US population
v) Quality of information provided about minority/protected group differences, effects of age, gender etc:	0 stars

Reviewers' comments about the norms:

The normative sample for the WMS®-IVUK is based on data collected as part of the standardization sample in the United States. Adults aged 16 to 90 were sampled using a stratified sampling procedure according to sex, race/ethnicity, education level and geographic region. 100 adults were sampled for each age band, giving a total sample size of 900 for the Adult battery and 500 for the Older Adult battery. Testing was continued to ensure that in each age band there was a mean score of 100 on the General Ability Index. Extensive quality control measures were used to ensure that test administration and scoring were carried out consistently and accurately and comprehensive exclusion criteria were used to reduce atypical scores. Obtained samples were compared to 2005 US Census data with results indicating that the sampling procedures resulted in a highly accurate reflection of the population in terms of the demographic characteristics chosen.

Norms were produced using the method of inferential norming (Zachary and Gorsuch, 1985) which uses statistical methods to ensure that information from the whole sample informs the conversion between raw score and scaled scores as contained in each age band table. This was originally done to overcome problems when a small change in age could lead to a large difference in scaled score because a different norm table was required for the new age band. More recent research supports the view that such techniques enable the creation of norm tables using relatively small sample sizes within age bands that are comparable with tables created using traditional methods using larger norm samples (Chen and Zhu, 2011).

A UK research study was conducted on a broadly representative sample of 235 adults aged 16 to 90, representativeness being judged using comparative figures from the 2001 UK Census. As before the data was collected using quality control measures to support consistent test administration and scoring by the range of testers involved. Of the sample of 235 people, 208 completed the Adult Battery and 27 the Older Adult Battery, consequently much of the analysis presented is based on the Adult Battery. In addition to this difficulty there was an over-sampling of more educated individuals and a corresponding under-sampling of less educated people leading to slightly higher scores than the US sample. However summary statistics that compare the UK data for the sub tests and index scores of the Adult Battery with the US data show that any differences are small and become smaller when the UK sample are reduced slightly to more closely match with the census by educational level.

This UK research study was not intended to produce normative data, but to evaluate the consistency of results obtained from the WMS®-IV between the US and UK. This analysis provides sufficient evidence to support the use of the US norms in the UK for the Adult Battery. However, in the absence of data for the Older Adult Battery it could be argued that this conclusion is not appropriate to support its use in the UK. There is no information provided regarding group differences on the tests based on gender or minority group status.

Validity:

Overall adequacy:	★★★★★
Construct Validity (overall adequacy):	★★★★★
i) Designs used:	Confirmatory Factor Analysis Subtest intercorrelations Correlations with other instruments
ii) Sample sizes:	★★★★★
iii) Procedure of sample selection:	Varied; some representative of US population.
iv) Median and range of the correlations between the test and other similar tests:	★★★
v) Quality of instruments as criteria or markers:	★★★★★
vi) Differential Item Functioning (DIF) analyses:	N/A
Criterion-related validity: overall adequacy:	★★★
i) Description of the criteria used and characteristics of the populations:	Opportunity samples for a range of clinical conditions with known memory deficits.
ii) Sample sizes:	★
iii) Procedure of Sample selection:	Data collected by independent examiners and submitted to test developers. Examinees needed to meet specified

	inclusion criteria.
iv) Median and range of the correlations between the test and criteria:	N/A

Reviewers' comments about validity:

The Technical Manual provides an extensive range of information on the content and construct validity of WMS®-IV. Content validity is derived from the subtests drawing heavily on relevant academic research which, in turn, leads to the rationale for the content of the subtests and their item content. Observation of the response processes of respondents during the test development is also argued to provide further evidence of content validity.

Evidence in support of construct validity is primarily based on US data. The inter-correlations between the subtests are cited as evidence of the internal structure of the test. An examination of a priori hypotheses regarding the relationship between WMS®-IV subtests for both the Adult and the Older Adult batteries supports the rational structure of the batteries at the level of subtest scores and index scores in that subtests measuring auditory or visual memory functions correlate more highly with other subtests measuring similar rather than the different types of memory functioning.

Confirmatory factor analysis was used to test two possible underlying models, the first a two, correlated factor model based on Auditory and Visual Memory and the second a three, correlated factor model which includes visual working memory. Both models performed well in the analysis with the three, correlated factor model preferred. This was based not only on the analysis but also in support of the important conceptual distinction introduced by working memory and because it is recognised that visual memory and visual working memory are disassociated in clinical samples.

This Confirmatory factor analysis was replicated in the UK Research Study using data from the Adult Battery and included evaluation of a third model based on a single memory factor. Each of the three models performed well with the three, correlated factor model having the best fit. This evidence of the equivalence of WMS®-IVUK with the US edition enables the UK user to have confidence that the extensive body of validity evidence presented for the WMS®-IV is also applicable for the UK edition. A total of 15 other measures have been used to assess what is described in the Technical Manual as the concurrent criterion validity of the WMS®-IV. However using the EFPA review criteria, relationships between WMS®-IV and other measures of similar or related constructs are classified as evidence of construct validity and therefore this evidence is rated accordingly. In each study a prediction was made of the expected relationships between scores.

A study was carried out using the WMS®-IV Adult Battery and WMS®-III battery for 224 examinees. The main changes between the versions are in the visual memory and visual working memory subtests and here results were in line with predictions, reflecting such changes while also supporting the contention that similar constructs are being measured. The correlations between Index scores were 0.50 for the Visual Immediate Memory Index, where there were the most changes in the subtests, and 0.81 for the Auditory Immediate Memory Index, where there were the least changes.

Correlations for 197 examinees between the WMS®-IV Older Adult Battery and WMS®-III Abbreviated scores of Immediate Memory Composite, Delayed Memory Composite and Total Memory Composite were used to demonstrate the construct validity for the Older Adult Battery. Results for Index scores ranged from 0.73 and 0.76 supporting the claim that both measures are assessing similar constructs.

Correlations between the WMS®-IV indices and Californian Verbal Learning Test version II (CVLT-II) scores for 380 examinees showed the expected relationships with the highest correlations between CVLT-II and the auditory memory scales and the lowest with the visual memory scales.

Correlations between the WMS®-IV indices and the Children's Memory Scale (CMS) measure of learning and memory for children and adolescents, administered to 38 examinees aged 16 years, range from 0.25 between the AMI and the Visual Immediate memory scales to 0.74 between WMS®-IV Immediate Memory Index and the CMS General memory scale.

Correlations with other tests of general cognitive functioning were used in the expectation that the results would show that although scores are likely to be associated because both intelligence and memory rely on the same underlying component processes, they would also provide evidence that memory is a separate

construct. More specifically it was predicted that the WMS®–IV Auditory Memory Index would be more closely associated with the WAIS Verbal Comprehension Index (VCI), the WMS®–IV Visual Memory Index with the WAIS Perceptual Reasoning Index (PRI) and the two working memory indexes with each other. Relationships were calculated from the standardization samples for both tests based on 1250 examinees. As predicted the AMI correlated most highly with the VCI (0.53) and the VMI with the PRI (0.62) and the two working memory indexes showed an association of $r = 0.62$.

A much smaller study of 52 examinees aged 16 found that the correlations between the WMS®–IV Index scores and Wechsler Intelligence Scale for Children (WISC®)–IV full scale IQ range from 0.49 to 0.68, with a median value of 0.66. AMI correlated most highly with the Verbal Comprehension Index (0.51) while VMI showed a stronger relationship with the Perceptual Reasoning Index (0.63).

Evidence of the relationships between WMS®–IV and measures of both broad and more specific neuropsychological domains is provided firstly from a study of 100 examinees who completed the WMS®–IV scales and also the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). Results were broadly in line with expectations in that indices within the same domains correlated most highly with each other. Secondly, the Delis–Kaplan Executive Function System (D–KEFS) subtests measure a wide range of verbal and non–verbal executive functions and 90 examinees took both tests. It was predicted that low to moderate correlations would be found. This was the case with correlations between the WMS–IV indices and the D–KEFS Trail Making and Verbal Fluency tests ranging from 0.22 to 0.54, with a median of 0.35. For 30 examinees with Traumatic Brain Injury the correlations between measures showed a similar pattern but were stronger ranging from 0.03 to 0.76, with a median of 0.52. This was predicted given that executive function and memory deficits are both associated with damage in areas affected by traumatic brain injury.

Evidence of the relationships between the memory scales and measures of both academic and everyday living skills are also presented. Academic achievement was measured using WIAT–II with 54 examinees aged between 16 to 19 years. The strongest relationships were between AMI and VWMI and the WIAT–II Total composite score while VMI, IMI and DMI correlated most highly with the Oral Language Composite. The results are interpreted as showing the relevance of memory functioning to academic achievement. Relationships between Older Adult scale scores and scores on the Independent Living Scales (ILS) were predicted to be low on the grounds that the 74 normal examinees aged between 65 and 89 should not have any difficulties living independently so that their ILS scores would have a very restricted range. This proved to be the case with correlations ranging from 0.08 to 0.51. A further study was carried out using a group of 30 examinees with Mild Cognitive Impairment (MCI) when the same pattern of results was expected on the basis that MCI does not lead to significant problems in daily living skills. However lower average scores were expected on all measures and this was found to be the case. A further study used the ABAS–II measure of adaptive functioning with a normally aging group of 132 examinees and again found no relationship between adaptive behaviour and memory functioning in this group. This study was repeated with 33 examinees with Mild Cognitive Impairment and while lower average scores were found there was little relationship between the scales.

Further validity evidence is provided in terms of the scores obtained by various clinical groups on the WMS®–IV, and shows the test’s ability to discriminate meaningfully between different samples. Such evidence has the potential to provide what the EFPA criteria refers to as criterion–related validity as test scores would in theory be able to predict or classify an individual’s diagnostic category. However the authors emphasize that such classification should never be made on the basis of test scores alone and also describe the limitations of the studies as some are based on small sample sizes, that the samples themselves may not be representative of the condition and finally that only group performance is reported as compared to matched control groups.

These Special Group Studies were carried out to demonstrate the ‘clinical utility’ of the scales and the special groups were chosen because of their high incidence or because of the memory deficits associated with them. The clinical conditions include: Probable Dementia of the Alzheimer’s Type–Mild Severity, Mild Cognitive Impairment, Traumatic Brain Injury, Right Temporal Lobectomy, Left Temporal Lobectomy, mental illnesses such as Schizophrenia, Major Depressive Disorder, Anxiety Disorder, and developmental conditions such as Mild and Moderate Intellectual Disability, Autism and Asperger’s Disorders, Reading and Mathematics Disorders and ADHD.

Reliability:

Overall adequacy:	★★★★★
i) Data provided about reliability:	Standard error of measurement given for a number of different groups
Internal consistency:	
i) Sample size:	★★★★★
ii) Median of coefficients:	★★★★★
Test retest stability:	
i) Sample size:	★★★★
ii) Median of coefficients:	★★★★
Equivalence reliability:	
i) Sample size:	N/A
ii) Median of coefficients:	N/A

Reviewers' comments on Reliability (comment on confidence intervals for reliability coefficients and provide Spearman Brown equivalents for a 30-item scale):

A good range of reliability evidence is presented for the WMS®-IV. Internal consistency data is provided for each subtest (or an estimate based on an alternate method: test retest or decision consistency as appropriate to the sub-test) and index scores obtained from both the Adult and Older Adult versions of the test. This is further broken down by age band.

Data from the US standardization sample show average subtest coefficients ranging from $r = 0.74$ to $r = 0.97$ with a median value of $r = 0.85$. The average composite coefficients for the five indices range from 0.92 to 0.97 with a median of 0.95.

These values translate into average SEMs that range from 0.50 to 1.55 for the subtests in the Adult version of the WMS and from 0.60 to 1.53 in the Older Adult version (presented on scales with $SD=3$). For the index scores, SEMs range from 3.04 to 4.12 in the Adult version and 2.77 to 4.29 in the Older Adult version (presented on a scale with $SD=15$).

The UK research study ($n = 235$) provides comparative coefficients for the subtests and index scores with internal consistency estimates ranging from 0.83 to 0.97. Data was not available in the required form for the Visual Reproduction subtests and data from the US standardization was substituted to estimate the reliability of the derived Index scores.

Test-retest reliability was evaluated on 173 respondents for the Adult battery and a smaller sample of 71 for the Older Adult Battery, with both samples being from the United States. Test-retest intervals ranged from 14 to 84 days with a mean of 23 days. Corrected and uncorrected correlations are reported, as are effect sizes for the differences for subtests and index scores. Corrected test-retest correlations range from 0.59 to 0.77 for subtests in the Adult battery and 0.69 to 0.81 in the Older Adult battery. For Index scores, corrected test-retest correlations range from 0.81 to 0.83 for the Adult battery and 0.80 to 0.87 for the Older Adult battery.

Evidence of inter-rater reliability is also presented and shows this to be high, even for subtests that place greater interpretive demands on raters (e.g. Clock Drawing and Visual Reproduction) where values of 96 to

97% were obtained.

It is noted by the authors that high levels of reliability have been achieved despite the limited variability on some subtests in non-clinical samples.

A further study was carried out to provide evidence of the reliability of the tests with special groups. This was based on 555 examinees with a range of disorders. The study found average internal consistency subtest coefficients for the 15 special groups (including Reading Disorder, Mathematics Disorder, ADHD, ASD and Aspergers) ranged from $r = 0.86$ to 0.97 with a median of 0.92 . Average internal consistency index coefficients for the special groups ranged from 0.93 to 0.98 with a median of 0.96 .

Quality of Computer-Generated Reports

Report Name:

The Score Reprt

Overall adequacy of computer-generated reports:	★★★★★
i) Scope or coverage:	★★★★★
ii) Reliability:	★★★★★
iii) Relevance or Validity:	★★★★★
iv) Fairness, or freedom from systematic bias:	★★★★★
v) Acceptability:	★★★★★
vi) Practicality:	★★★★★
vii) Length-number of printed pages:	5
viii) Length index (number of pages (including composite and derived scales) divided by number of scales), multiplied by 10 and expressed as an integer:	1

Report Name:

The Interpretive Report

Overall adequacy of computer-generated reports:	★★★★★
i) Scope or coverage:	★★★★★
ii) Reliability:	★★★★★
iii) Relevance or Validity:	★★★★★

iv) Fairness, or freedom from systematic bias:	★★★★★
v) Acceptability:	★★★★★
vi) Practicality:	★★★★★
vii) Length–number of printed pages:	12
viii) Length index (number of pages (including composite and derived scales) divided by number of scales), multiplied by 10 and expressed as an integer:	4

Reviewers' comments on the quality of computer generated reports:

The report writer software allows respondent data to be managed from a PC and detailed reports to be produced. The software generates the reports based on raw scores, saving the user a lot of time in the scoring process as well as in the production of a report. The two reports reviewed contain a lot of detailed information, though are well presented.

The Score Report includes all of the subtest scores and Index scores using a series of tables and graphs to represent the scores. The Interpretive Report, in addition to the information in the Score Report, provides a narrative interpretation of the WMS®–IVUK, as well as relevant details from the examinee history screens. This report describes the tests and explains the scores as well as giving background information and recommendations. The report also gives tables of scores, graphical profiles and discrepancies. The report states 95% confidence intervals for the Index scores.

Where narrative text is included, the links between the scores obtained from the WMS®–IVUK and the reports are clear and suitable cautions are attached to report interpretation. Reports are editable, if users wish to do this.

The reports are generally appropriate for the end user, if that is the test user or related professional.

FINAL EVALUATION

Evaluative report of the test:

The WMS®–IVUK is the fourth UK edition of a memory test that has a long history of continual development in the US. One of the specifications of WMS®–IV was to produce an instrument that was culturally universal so that it could be used with minimal changes across the world. It offers a comprehensive, psychometric assessment of memory with norm referenced interpretation based on a carefully chosen sample of normal adults from the US. The WMS®–IV contains two batteries, one for adults aged between 16 and 65 and a second shorter battery for older adults between 65 and 90 years. Both are individually administered tests. It is most likely to be used within the clinical, neurological, forensic and educational fields as well as a work rehabilitation setting.

The WMS®–IVUK is a complex test battery and is likely to require a fair amount of experience before new users are comfortable with it and able to make full use of its potential. The user qualification required by the publisher would seem to be at the right level, given the expertise required in interpreting the test results. Users are required to have a postgraduate qualification and training in clinical or educational psychology or a relevant PhD or be a chartered occupational psychologist.

The two batteries use norms based on the inferential norming method which enables information contained in the whole sample size to be utilized within each age band norm group. This mitigates against the instability that might otherwise occur in norm groups based on only 100 people even though they are carefully chosen to be representative samples.

The UK validation study provides some reassurance that it is appropriate to use the US norms with UK subjects although the absence of analysis of the equivalence of the Older Adult Battery is of concern. This UK study also provides supporting evidence of the reliability of the scales and of the similarity of the factor structure of the measures in the UK sample. This latter evidence of the equivalence of the UK edition provides support that the results of the extensive validity studies carried out in the US are applicable to the UK. Even so, a more comprehensive UK standardisation study including both batteries would be preferable.

The information presented on the rationale, background research and test development are comprehensive. The separate Administrative and Scoring Manual is well laid out, providing an overview of the test before giving detailed instructions for administration and scoring. The extensive technical information contained in the Technical and Interpretive Manual will give users confidence in the robustness of the test development process and resulting scores.

The inclusion of report writing software greatly aids with the generation of reports given the complex nature of the WMS®UK-IV.

The Wechsler Memory Scale® – Fourth UK Edition is a very well-developed psychometric assessment of memory in adults that provides rich information on memory functioning.

Conclusions:

The WMS®-IVUK is an improvement on the WMS-IIIUK in terms of its subtest modifications (to fit with current theoretical perspectives) and its psychometric properties. The UK validation study goes some way to reassure the UK user that the extensive research underpinning the original US fourth edition can be relied upon. In this case it is fair to conclude that the WMS®-IVUK is an excellent memory scale, with good evidence of its reliability and validity.

Recommendations:

- Suitable for use in the area(s) of application defined by the distributor, by test users who meet the distributor's specific qualifications requirements

SUMMARY EVALUATION OF THE TEST

Content Domains:

Memory

Intended or main area(s) of Use:

Psycho-clinical
 Psycho-neurological
 Forensic
 Educational
 Work and Occupational
 Counselling, Advice, Guidance, and Career Choice

Intended mode of use (conditions under which the instrument was standardised and validated):

Supervised and controlled administration. Test administration under the control of a qualified administrator or proctor

Test Description:

Test Name:	Wechsler Memory Scale ® – Fourth UK Edition
Local test distributor / publisher:	

	Pearson Assessment
Date of Current Review:	2011
Date of Publication of Current Review/Edition:	2011
Constructs Measured:	Auditory memory, visual memory, visual working memory, immediate memory and delayed memory.
Administration Mode:	Interactive individual administration
Response Mode:	Oral interview Manual operations Paper and pencil

Instrument Evaluation:

Characteristics	Evaluation
Quality of Documentation	★★★★☆
Quality of Materials	★★★★★
Norms and reference groups	★★★★★
Construct validity	★★★★★
Criterion-related validity	★★★
Reliability-overall	★★★★☆
Number of Computer-Generated Reports	2

The British Psychological Society © 2012. All rights reserved. No part of this publication may be reproduced or transmitted, in any form or by any means, without permission.

Debug mode: **0 (none), 1 (basic), 2 (advanced), 999 (edit)