



The Wechsler memory scale-revised, or the Wechsler memory scale-revisited?

David W. Loring

To cite this article: David W. Loring (1989) The Wechsler memory scale-revised, or the Wechsler memory scale-revisited?, *The Clinical Neuropsychologist*, 3:1, 59-69, DOI: [10.1080/13854048908404077](https://doi.org/10.1080/13854048908404077)

To link to this article: <https://doi.org/10.1080/13854048908404077>



Published online: 08 Nov 2007.



Submit your article to this journal [↗](#)



Article views: 173



View related articles [↗](#)



Citing articles: 46 View citing articles [↗](#)

FORUM

The Wechsler Memory Scale-Revised, or The Wechsler Memory Scale-Revisited?

David W. Loring
Department of Neurology
Medical College of Georgia

ABSTRACT

The Wechsler Memory Scale-Revised has addressed many of its predecessor's shortcomings. Due to its relatively strong psychometric grounding and representative normative sampling, the WMS-R will likely obtain a prominent position in many neuropsychology batteries. The WMS-R contains multiple memory measures, has included a Delayed Recall Index, and separates attention/concentration factors from verbal, visual, and general memory function. Consequently, it is hoped that the information obtainable will have clinical, as well as psychometric, significance. This paper reviews the revised scale, and presents some preliminary concerns regarding the test's construction prior to its widespread implementation.

The Wechsler Memory Scale-Revised (WMS-R), due to its relatively strong psychometric development and representative normative sampling, will likely obtain a prominent position in many neuropsychology batteries. Unlike the original WMS, normative information is available from age 16 to 74 years. The test has been updated to include multiple measures of verbal and visual memory, and examines the retention of information over time. Consequently, it is hoped that the WMS-R will have clinical, as well as psychometric, significance.

The suggestions of Gregory P. Lee, Roy C. Martin, and Kimford J. Meador are appreciated. I thank William O. Thompson for his comments on sampling estimation, and Patricia A. Downs for manuscript preparation.

Requests for reprints should be addressed to David W. Loring, Ph.D., Section of Behavioral Neurology, Department of Neurology, Medical College of Georgia, Augusta, GA, 30912-3275, USA.

Accepted for publication: July 15, 1988.

According to the manual, five major changes from the original scale are now included (p. 2):

1. Provision of norms stratified at nine age levels.
2. Replacement of a single global summary score (the Memory Quotient) with five composite scores.
3. Addition of new subtests measuring figural and spatial memory.
4. Addition of measures of delayed recall.
5. Revision of the scoring procedures for several subtests to improve scoring accuracy.

In a recent issue of *The Clinical Neuropsychologist* (Vol. 2, No. 2, 1988), the initial validity studies of the WMS-R suggest its potential clinical utility. However, these reports were largely descriptive, with only one paper directly comparing the WMS-R to an independent memory measure (i.e., Delis, Cullum, Butters, Cairns, & Prifitera, 1988). Similarly, with the exception of Fischer (1988), the reports did not directly examine individual patient classification based upon the available normative information provided by the WMS-R manual. Informal discussions at the Sixteenth Annual Meeting of the International Neuropsychological Society revealed residual concern regarding whether the test has been improved, or has merely been repackaged in an expanded format.

I. NORMATIVE DATA

One must temper criticism of normative data by acknowledging that any systematically collected database attempting to adequately sample the population of interest is of value, particularly when considering memory tests. With respect to the WMS-R, normative data are available from ages 16-74. Approximately 50 subjects were included in each age range tested and, according to the manual, this is considered an adequate sample size for deriving stable population estimates. This contrasts with 200 subjects per decade for the Wide Range Achievement Test-Revised (Jastak & Wilkinson, 1984) and Peabody Picture Vocabulary Test-Revised (Dunn & Dunn, 1981). Although not spanning the same age ranges, the Stanford-Binet Intelligence Scale employed 200-300 subjects per age group (Thorndike, Hagen, & Sattler, 1986). For the Wechsler Intelligence Scale-Revised, 200-300 subjects were included through age 54, with 160 subjects per group from 55-74. By including 50 subjects per group rather than 200, the estimated population means are half as stable; the standard error of the mean increases by a factor of two.

To compound the problem of relatively few subjects, all age groups were not included in the standardization sample and only estimated performances are available for certain age ranges (18-19, 25-34, 45-54). To the extent that there is a linear decline in memory function, this may be an acceptable technique to reduce the number of subjects sampled in order to obtain a population estimate.

However, examination of Table 8 from the manual reveals that this is likely not to be the case. Performance estimates for ages 35-44 are based upon a sample of 54 subjects; performance levels for the two decades on either side (25-34, 45-54) are derived from interpolated estimates. The difference in weighted raw score composites from 35-44 to 55-64, the next age group for which genuine performance data are available, is 17.3 points. This contrasts sharply with the 2.3 point difference observed from ages 20-24 to 35-44. The assumption of a linear decline is untenable unless one argues that this is an artifact of relatively unstable estimates of the mean. The inclusion of actual performance levels for this important age at which many patients present with memory difficulty would be reassuring, rather than having to rely upon interpolated estimates based upon data from relatively few subjects.

Data presented by Butters et al. (1988) underscore the above concerns regarding interpolated performance estimation. Because each composite memory index is standardized with a mean of 100 and standard deviation of 15, the expected value for the difference between any two indexes in healthy subjects is 0. Yet, the mean difference between General Memory and Attention/Concentration Indexes reported in a healthy group of controls (mean age = 51 years) was 17 points (Butters et al., 1988). This age corresponds to 45- to 54-year-old range in which the WMS-R normative tables have been derived by interpolation.

Because of the scale's emphasis on representative normative information, importance must be placed upon individual patient prediction analogous to the approach employed in a clinical setting (e.g., Loring, Lee, Martin, & Meador, 1988). Using the approach of Fischer (1988), who examined individual classification based upon both the WMS-R norms and her control group performance, somewhat different patterns of memory performance in multiple sclerosis patients emerge. When comparing classification rates for the two samples using a one standard deviation below the mean criterion on the General Memory Index, approximately three times as many patients were classified as memory impaired using her local control group compared to the WMS-R tables (64% vs. 22%). Although differences in education between the two normative groups is likely contributing to the classification discrepancy, it is essential that, for measures designed primarily for clinical assessment such as the WMS-R, clinical as well as statistical approaches to performance differences are examined.

II. COMPOSITE MEMORY SCORES

The Memory Quotient. The use of summary measures to characterize patient performance in neuropsychology has been criticized as insensitive to the variety of performances displayed by patients with brain dysfunction (Lezak, 1988). Specifically, the smoothing necessitated by collapsing scores obscures patterns that may be clinically relevant. In this vein, the Memory Quotient from the original WMS was inadequate to represent the diversity of recent memory

functioning displayed by both healthy individuals and by patient populations. To deal with this shortcoming, the WMS-R contains five index scores to more fully characterize an individual's memory functioning. However, the WMS-R still provides an overall General Memory Index, analogous to MQ, which in many contexts will be treated as the MQ-R. Although the General Memory Index no longer consists of orientation and attention/concentration subtests, the index remains a composite of both verbal and visual learning performance.

Subtest Composition. The original WMS was criticized by Prigatano (1977) as being primarily a test of verbal memory, with only Visual Reproduction assessing visual memory. The maximum number of points for the three subtests used most frequently to assess memory (i.e., Logical Memory, Visual Reproduction, and Paired Associate Learning) totaled 50. Since the two verbal memory tests' sum is 36, a greater contribution of verbal memory over visual memory was present (72% vs. 28%). Consequently, in its revision, two additional subtests were developed which, in conjunction with Visual Reproduction Memory, are components of the Visual Memory Index (Figural Memory, Visual Paired-Associate Memory).

Unfortunately, the General Memory Index is still more heavily dependent upon verbal memory functioning. When calculating the Indexes (General, Verbal, and Visual), the number of points still favors the verbal subtests. Of the possible 193 points after weighting, 124 are from the verbal subtests, and 69 are from the visual subtests. Although the verbal/visual discrepancy (64% vs. 36%) is not as extreme as with the original scale, the General Memory Index is clearly more dependent on verbal memory than on visual memory functioning. However, because it can be argued that the number of subtest items is only one of several factors that contribute to a summary score of several subtests (e.g., item reliability, mean item difficulty, and item homogeneity), it is illustrative to explore the following example.

Consider two patients with Verbal and Visual Memory Indexes of 100 and 74 using the Table C-1 from the WMS-R manual, p. 128, for 55- to 64-year-old subjects. When the higher index is verbal and the lower index is visual, a composite index of 91 (27th centile) is obtained. In contrast, when the opposite pattern is observed (verbal=74, visual=100), a composite index of 84 (14th centile) is derived. This 7-point difference exceeds the standard error of measurement for the General Memory Index, and somewhat different inferences might be made based upon FSIQ-General Memory Index differences (e.g., with a FSIQ of 100). Perhaps, in our culture, the concept of recent memory should be more heavily weighted in favor of verbal learning. However, cognitive theory should determine the relative contributions of material-specific forms of memory, rather than developing memory composites based solely on psychometric considerations designed only to maximize composite score reliability.

III. NEW SUBTESTS

The WMS-R contains two new subtests, Figural Memory and Visual Paired Associates, to examine more fully visual memory functions. Figural Memory requires the patient to study modular designs containing white, grey, and black components for 5 s each, and then to identify the same figure from an array in a recognition format. Both single designs (5 s) and sets of three designs (15 s) are presented.

For Visual Paired Associates, an analogue of Verbal Paired Associates, the patient is presented abstract designs paired with different colors. Following paired presentation, only the design is displayed and the subject indicates the associated color from a response array of differed colored squares. "In order to minimize the role of verbal mediation in memorizing and responding to the figure-color pairs, the color names are not used either in presenting the items or in responding to them" (Wechsler, 1987, p. 5).

By including additional visual memory measures, a greater diversity of visual memory functions is sampled. However, the net effect may have been to make this a less material-specific memory measure. Remarkably little information is presented on the rationale for inclusion of the two new visual tests. During test development, the overriding concern was apparently "to clarify the directions for administration" (Wechsler, 1987, p. 43).

Cursory inspection reveals that these two subtests lack the necessary face validity to assess visual memory and learning. Figural Memory appears to be a measure of higher-order visual attention span that does not examine retention of information over time. Figural Memory loads substantially, although not exclusively, on the attention/concentration factor in a mixed clinical sample in the preliminary factor analytic study offered in the manual (p. 76). However, these findings are preliminary, without marker variables designed to provide sufficient covariation for establishment of specific factors of neuropsychologic interest (e.g., Larrabee, Kane, Schuck, & Francis, 1985). Bornstein and Chelune (1988) presented data that suggest Figural Memory loads more heavily on a nonverbal memory factor than either verbal memory or attention when immediate and delayed recall memory trials are included in the analysis. However, this finding must also be interpreted cautiously since by including both immediate and delayed recall performance in a single factor-analysis, the tests themselves will cluster into separate factors based upon the high immediate/delay performance correlations (Fletcher, Taylor, Morris, & Satz, 1982). As with any factor-analytic approach to data reduction in heterogenous populations, cautions regarding correlative analyses exclusively to define theoretical constructs are germane (see Gould, 1981).

Visual Paired Associates contains a significant verbal component. Although "the color names are not used" to minimize verbal mediation, our experience suggests the almost all patients spontaneously employ verbal labeling during performance. Reversing stimuli and responses and having the patients select the

designs when presented with a color would lessen the contribution of verbal mediation.

However, Visual Paired Associates is not a selectively sensitive measure of visual learning. Brown, Sawyer, Nathan, and Shatz (1987) reported that paired-associate learning of geometric figures that could not be easily verbalized was not selectively sensitive to right cerebral dysfunction. A tendency exists to develop analogues of "left hemisphere" tests containing spatial rather than verbal stimuli. Unfortunately, since the cerebral hemispheres are specialized for differing task performances, there is no *a priori* reason to expect that simply changing from words to figures will create a sensitive "right hemisphere" measure. This error is similar to the situation described by Meehl (1975) in which matching subjects in an *ex post facto* design may produce subject inclusion that is not representative of the population of interest. By developing a visual-spatial extension of paired-associate word learning without appropriate pilot research, the possibility clearly exists that this test may not selectively reflect the normal (or abnormal) memory functions of the right cerebral hemisphere (e.g., Brown et al., 1987). The WMS revision would have been better served by selecting tests that had been previously demonstrated to be selectively sensitive to memory deficits associated with right cerebral dysfunction (e.g., facial recognition; Milner, 1968; Warrington, 1984).

The summary indexes are reportedly unable to differentiate patients with left temporal lobe epilepsy from patients whose seizures originate from the right temporal lobe (Wechsler, 1987, pp. 84-85). In our experience with patients who have previously undergone unilateral anterior temporal lobe resection, not only do the summary indexes fail to correctly predict unilateral temporal dysfunction, but may also incorrectly suggest lateralized dysfunction (Loring, Lee, Martin, & Meador, 1988). For example, using a 15-point or greater Verbal/Visual Memory discrepancy, only 2/10 right temporal lobectomy patients were correctly classified, whereas 8/10 right temporal lobectomy patients were incorrectly classified (i.e., Visual Memory Indexes greater than Verbal Memory Indexes).

Using patients with lesions confined to a single hemisphere, Chelune and Bornstein (1988) reported left/right group difference for only the verbal memory. Further, the discrepancy between indexes was sufficiently small (8.5 points) relative to their variability (pooled standard deviation estimate=17.4 points), that it is clear that statistical significance was achieved with relatively large sample sizes. That Verbal and Visual Indexes may falsely lateralize unilateral mesial temporal lobe dysfunction, and that a relatively small left/right group effect exists for only the Verbal Memory Index in patients with unilateral lesions, suggests that these material-specific summary scores frequently will not be useful for individual patient prediction.

The failure of the summary indexes to identify lateralized dysfunction, of course, does not preclude certain derived scores such as percent retention from being selectively sensitive to unilateral lesions, since the Verbal and Visual Indexes are based solely on immediate recall trials. However, Butters et al. (1988)

report that of the four tests with both immediate and delayed recall measures, "Logical Memory and Visual Reproduction tests were far superior to those from the two paired-associates tasks for differentiating patients from normal subjects" (pp. 145-146). Thus, no major content improvement compared to the original scale is suggested by the addition of Visual Paired Associates, or the formation of Verbal and Visual Memory Indexes.

IV. DELAYED RECALL

Since patients with temporal-lobe dysfunction reportedly forget newly learned information more rapidly than do patients with diencephalic lesions or healthy controls (e.g., Huppert & Piercy, 1979), the inclusion of delayed-recall performance is of potential clinical significance. The Delayed Recall Index is composed of delayed-recall performance for Logical Memory, Visual Paired Associates, Verbal Paired Associates, and Visual Reproduction. By comparing the General Memory Index to the Delayed Recall Index, an estimate can be obtained whether there is disproportionate decline in the ability to retain newly learned information (e.g., Butters et al., 1988).

Careful examination of the Delayed Recall Index derivation reveals some interesting inconsistencies. As previously stated, there is no delay condition for the Figural Memory subtest; in consequence, a somewhat different composition of tests constitute the various indexes. In addition, the subtests are differentially weighted for the General Memory and Delayed Recall Indexes. Because learning trials are not administered, a different total of possible points exists for delayed recall and for the relative contributions of each subtest change.

A preferable way to examine memory decay is to present difference scores for each of the subtests. In this way, a common metric could be applied. For Logical Memory and Visual Reproduction, the same number of possible points exists since learning over repeated trials is not conducted. For both paired-associate tests, the delay should be contrasted with the final trial administered during acquisition.

The WMS-R has partially addressed the issue of prompting if the patient fails the free-recall trial. Unfortunately, prompting is employed only for Logical Memory and not for Visual Reproduction. The potential exists for significant variability in performance by not allowing a prompt. For example, if the subject is unable to recall freely the final design, consisting of two major elements, 18/41 (43%) possible points are lost. If able to recall the design following a cue, a somewhat different picture of delayed-recall memory would emerge. Multiple-choice allows examination of memory for designs over time without confounding retention with free-recall failure. Unfortunately, this option was not implemented in the WMS revision.

V. SCORING CRITERIA

The scoring criteria have been improved significantly, and this clarification will undoubtedly produce greater scoring reliability. However, some disagreements between our own personal biases and the official criteria are present. For example, on the Robert Miller paragraph, the phrase "tractor trailer" is frequently given for truck, but receives no credit. "His truck skidded off the road into a ditch," will provide three points if appropriately reproduced. However, when recalling this, many patients indicate that there was simply an accident or wreck, for which not a single point can be awarded. Further, no description is provided regarding the basis for determining what constituted a memory unit. I doubt whether a research paper on prose passage memory would be accepted for publication if the same degree of description in memory unit definition were included.

MISCELLANEOUS COMMENTS

(1) Although the WMS-R is designed to assess the ability to learn new information and material, adequate attention and orientation is necessary. This is recognized explicitly in the WMS-R, and formal assessment of these functions is part of the scale. Digit span has a long history, both in clinical neurology and clinical psychology. In a neuropsychological context, this subtest is better treated as two separate tests, forward and backward span (Lezak, 1983; Milberg, Hebben, & Kaplan, 1986). In this vein, the WMS-R now provides separate normative tables for these measures. However, a digits forward-digits backward discrepancy score would be a valuable measure to assess higher-order mental tracking.

(2) The likelihood is great that FSIQ and General Memory Index differences will be used to measure relative impairment in memory functioning, given the history of IQ-MQ deviation scores reports (e.g., Bornstein, Pakalnis, Drake, & Suga, 1988). Since FSIQs were obtained (or estimated) in the standardization sample, and used in certain statistical analyses, the standard error of FSIQ-General Memory Index differences should have been included with the other difference measures. Because diminution of memory is a frequent sequela to any brain injury or insult, it is unlikely that the WMS-R by itself will be effective in "evaluating the pattern and localization of organic brain damage" (Wechsler, 1987, pp. 6-7).

(3) The sample stratification should have selected subjects from distinct community sizes with clearly identified targets in small towns and rural areas. Although the WMS-R manual indicates geographic stratification based upon the four major regions used by the Bureau of Census, the size continuum from rural areas to cities exceeding 1,000,000 may be more appropriate. For example, it is easier to generalize from performances in Atlanta to Pittsburgh, than from

Atlanta to rural Georgia. Further, although education level is typically considered by whether or not an individual was graduated from high school, a subgrouping of those not finishing high school should also be considered. Many individuals in the rural South who are referred for neuropsychological assessment have limited education. The discrepancy between those patients with a grade school education only, and those with 7-11 years of schooling, suggests that this classification is as important as the high school/no high school distinction. Hopefully, subsequent research will examine this issue.

CONCLUDING REMARKS

Of the problems discussed, some difficulties could be addressed easily, although other flaws are more fundamental and without simple solution. For example, it is essential to have immediate/delay difference estimates for individual subtests since all measures will not be equally sensitive to memory dysfunction, and different relative contributions of the subtests comprise the General and Delayed Memory Indexes. These difference scores may require transformation in order to account for distribution differences, or perhaps a multivariate combination could be formed that takes advantage of the existing correlations. However, information regarding retention over time would be an important contribution to neuropsychological assessment. Digit span difference scores (forward vs. backward) would be similarly informative.

The WMS-R still appears to be more a test of verbal learning. However, when attempting to include additional nonverbal measures, new subtests were implemented that do not appear to be pure measures of visual learning/memory. It is likely that many WMS-R measures may be sensitive to generalized, nonlateralized, memory dysfunction (e.g., Butters et al., 1988). However, it is premature to use the Verbal and Visual Memory Indexes to infer lateralized temporal-lobe dysfunction.

Whether the WMS-R will provide a more sensitive and reliable index of memory function/dysfunction will require many years of research examining individual subtests in a variety of clinical populations, and comparing WMS-R results with other measures of memory. It is unfortunate, however, that the advancements made over the past several decades in cognitive and experimental/clinical psychology were largely ignored in the revision. There exists unquestioned improvement in the test's "surface structure." However, the test's "deep structure," the area of more theoretical importance and interest, remains essentially unchanged.

REFERENCES

- Brown, G.G., Sawyer, J.D., Nathan, A., & Shatz, M.W. (1987). Effects of lateralized cerebral dysfunction on the continuous paired-associate test. *Journal of Clinical and Experimental Neuropsychology*, *9*, 680-698.
- Bornstein, R.A., & Chelune, G.J. (1988). Factor structure of the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, *2*, 107-115.
- Bornstein, R.A., Pakalnis, A., Drake, M.E., Jr., Suga, L.A. (1988). Effects of seizure type and waveform abnormality on memory and attention. *Archives of Neurology*, *45*, 884-887.
- Butters, N., Salmon, D.P., Cullum, C.M., Cairns, P., Tröster, A.I., Jacobs, D., Moss, M., & Cermak, L.S. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale-Revised. *The Clinical Neuropsychologist*, *2*, 133-148.
- Chelune, G.J., & Bornstein, R.A. (1988). WMS-R patterns among patients with unilateral brain lesions. *The Clinical Neuropsychologist*, *2*, 121-132.
- Delis, D., Cullum, C.M., Butters, N., & Cairns, P., & Prifitera, A. (1988). Wechsler Memory Scale-Revised and the California Verbal Learning Test: Convergence and Divergence. *The Clinical Neuropsychologist*, *2*, 188-196.
- Dunn, L.M. & Dunn, L.M. (1981). *Manual for the Peabody Picture Vocabulary Test-Revised*. Circle Pines, MN: American Guidance Company.
- Fischer, J.S. (1988). Using the Wechsler Memory Scale-Revised to detect and characterize memory deficits in multiple sclerosis. *The Clinical Neuropsychologist*, *2*, 149-172.
- Fletcher, J.M., Taylor, H.G., Morris, R., & Satz, P. (1982). Finger recognition skills and reading achievement: A developmental neuropsychological analysis. *Developmental Psychology*, *18*, 124-132.
- Gould, S.J. (1981). *The mismeasure of man* (pp. 234-320). New York: W.W. Norton.
- Huppert, F.A., & Piercy, M. (1979). Normal and abnormal forgetting in organic amnesia: Effect of locus of lesion. *Cortex*, *15*, 385-390.
- Jastak, S., & Wilkinson, G.S. (1984). *Manual for the Wide Range Achievement Test-Revised*. Wilmington, DE: Jastak Associates, Inc.
- Larrabee, G.J., Kane, R.L., Schuck, J.R., & Francis, D.J. (1985). Construct validity of various memory procedures. *Journal of Clinical and Experimental Neuropsychology*, *7*, 239-250.
- Lezak, M.D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.
- Lezak, M.D. (1988). IQ:RIP. *Journal of Clinical and Experimental Neuropsychology*, *10*, 351-361.
- Loring, D.W., Lee, G.P., Martin, R.C., & Meador, K.J. (1988). Material-specific learning in patients with partial complex seizures of temporal lobe origin: Convergent validation of memory constructs. *Journal of Epilepsy*, *1*, 53-59.
- Loring, D.W., Martin, R.C., Meador, K.J., & Lee, G.P. (1988). The WMS-R incorrectly predicts lateralized temporal lobe dysfunction. *The Clinical Neuropsychologist*, *2*, 284 (Abstract).
- Meehl, P.E. (1975). Nuisance variables and the ex post facto design. In M. Radner & S. Winokur (Eds.), *Analyses of theories and methods of physics and psychology* (pp. 373-387). Minneapolis: University of Minnesota Press.
- Milberg, W.P., Hebben, N., & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment. In I. Grant & K.M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 65-86). New York: Oxford University Press.
- Milner, B. (1968). Visual recognition and recall after right temporal lobe excision in man.

Neuropsychologia, 6, 191-209.

- Prigatano, G.P. (1977). The Wechsler Memory Scale: A selective review of the literature. *Journal of Clinical Psychology*, 34, 816-832.
- Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1986). *Technical Manual for the Stanford-Binet Intelligence Scale: Fourth Edition*. Chicago: The Riverside Publishing Company.
- Warrington, E.K. (1984). *Manual for the Recognition Memory Test*. Windsor, UK: NFER-NELSON Publishing Company, Ltd.
- Wechsler, D. (1987). *Manual for the Wechsler Memory Scale-Revised*. New York: The Psychological Corporation.