WMS-R and MAS Correlations in a Neuropsychological Population

Charles J. Golden
Nova Southeastern University

Laura White, Tanya Combs, Michelle Morgan, and Dorothy McLane
Chesapeake Rehabilitation Hospital

The present study was an attempt to examine the relationship between the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987) and the Memory Assessment Scale (MAS; Williams, 1991). The sample consisted of 51 patients referred for neuropsychological examination. Average age was 55.29 (SD = 20.94). Average education was 11.94 years (SD = 3.18). Average Full Scale IQ was 94.75 (SD = 15.44). The results showed that no MAS index correlated higher than .60 with any of the WMS-R indices. There was minimal specific prediction between instruments across each memory domain. Agreement between the tests clinically was low as well. Corrections for such factors as age, education, intelligence or diagnosis lowered the relationships between the tests although not significantly. Corrections to the correlations for the inherent reliability of each test increased correlations between the tests but still resulted in a maximum common variance of 56% (for the visual measures) down to 40% (for General Memory). These results are consistent with the argument that “general memory” is not as useful a construct as that of “general intelligence” and that these tests of general memory measure different underlying constructs. This has important impact on how we conceive and report memory test results. © 1999 National Academy of Neuropsychology. Published by Elsevier Science Ltd

Although the Wechsler Memory Scale-Revised (Wechsler, 1987) was itself developed because of unhappiness with the original Wechsler Memory Scale, the scale has been subject to extensive criticism. This in turn has spawned many attempts at the development of additional memory instruments, some of which are more ambitious attempts to develop new batteries of general memory function. One of these is the Memory Assessment Scale (Williams, 1991).

Zielinski (1993) published a theoretical comparison of both the batteries. He suggested that the MAS is a better normed test that has less of the verbal loading of the WMS-R, as well as being more psychometrically sophisticated. Although differences exist between individual subtests, the general memory quotients would still be expected to
be reasonably correlated with one another as both tests sample the same general domain of memory skills.

Several studies have attempted to empirically study the relationship of the two tests. Bowler and Khazanov (1994) presented a comparison of the WMS-R and the MAS in 34 clients with exposure to neurotoxicants. These clients did poorly on both batteries when compared to age and education norms. The authors correlated “similar” indices on the two tests, finding only weak to moderate correlations. The authors concluded that while the two batteries may be similar in terms of measuring attention and concentration but not in other areas.

Hilsabeck, Dunn, and Lees-Haley (1996) evaluated the two tests using a sample of 30 clients referred for neuropsychological examination because of a traumatic brain injury. The authors found that MAS scores were significantly below WMS-R scores on all global measures.

Lazarus, Small, and Williams (1994) examined a sample of 52 patients referred for neuropsychological evaluation. These were general cases with a wide variety of problems. Average IQ for the group was 97 ($SD = 15$) suggesting relatively mild brain injury overall. The average age was 41 ($SD = 16$). Chronicity was not indicated. In this study, mean scores were statistically identical. The authors reported correlations among the respective verbal, visual, and global indices of .56, .30, and .52, all of which were statistically significant. The highest correlation was between attention and short-term memory (.73). The data showed that short-term memory from the MAS generally correlated as high with each of the WMS-R indices as did the corresponding scale of the MAS. For example, the two general memory indices correlated .52, and general memory of the WMS-R correlated .50 with the MAS short-term memory.

In the largest study to date, Putnam et al. (1994) compared the tests on 103 patients referred for assessment because of a head injury. The sample averaged 24 ($SD = 20$) months postinjury and had a mean age of 36.2 ($SD = 12$). The cases were classified as 80% mild head injury. The authors found correlations between the verbal, visual, and global indices of .60, .64 and .64 respectively. The correlation between the attentional and short-term memory indices was .79. Correlations among the verbal and visual subtests generally ranged from .23 to .66 with most of the correlations being in the .40 to .60 range. The authors concluded that the tests were correlated but had considerable unique variance as well. The authors did not report mean scores or any differences between mean scores.

These studies have raised significant questions about whether these tests are interchangeable. Although there are statistically significant correlations among these tests, the magnitudes of the correlations suggest that the tests may be measuring very different specific factors. However, the presence of low correlations alone does not mean that these tests would yield different clinical conclusions about each client. If clients are classified in a similar manner by both tests, then the degree of correlation and differences between the tests become less important. The present study was an attempt to further investigate the relationships between the general scores of each test to determine whether clinical classifications between the tests are similar.

**METHOD**

The sample consisted of 51 patients selected from 65 patients all referred for neuropsychological examination secondary to inpatient rehabilitation. Patients were included in the final sample if they had taken all parts of the WMS-R, MAS, and the WAIS-R. In-
pection of the clients not included showed no significant differences on demographic variables between these clients and the clients included. In each case, the client simply had failed to take all of the tests included in the current study usually due to time demands of their treatment or of the neuropsychology department.

Final diagnoses included Head Injury (19), Stroke (18), Tumor (2), Anoxia (1), Other Brain Injury (e.g., MS, Parkinson’s, Dementias, etc.) (4), and Undetermined (7). This latter category included individuals referred for brain injury but for whom no final diagnosis could be identified neurologically or radiologically. All diagnoses were made based on the medical rather than the neuropsychological data. All of the head injured clients showed clear neuroradiological evidence of focal damage in addition to any more diffuse problems caused by the trauma.

Average age of the sample was 55.29 (SD = 20.94). Average education was 11.94 years (SD = 3.18). There were 30 males and 21 females in the sample. Five of the patients were African-American. Average Full scale IQ was 94.75 (SD = 15.44), with a mean Verbal IQ of 99.43 (SD = 14.29) and a mean Performance IQ of 88.74 (SD = 16.58). The population was generally acute with an average time since injury at the initiation of testing of 18.3 days (SD = 22.1). Patients were administered the WAIS-R, WMS-R, and MAS in random order, generally in more than one session over 2–7 days when their condition was considered stabilized. All clients had adequate verbal skills to complete the three tests and understand their instructions.

RESULTS

Table 1 presents the mean scores for each of the MAS summary indices and each of the WMS-R summary indices. Scores for the normal population on all measures average 100 (SD = 15). As can be seen, scores varied on the memory indices from a low of 83.73 on Delayed Memory of the WMS-R to a high of 92.04 on Visual Memory of the WMS-R. Mean scores on the Wechsler were higher than mean scores on the MAS with the exception of short-term memory/attention.

Table 1 also reports the average difference scores between similar indices on each scale and the associated standard deviation. The average differences between matched scales were generally low ranging from 3 to 5 points overall. However, the standard deviations for the difference scores ranged from 15.22 to 21.54, suggesting substantial discrepancies between individual scores for specific clients.

<table>
<thead>
<tr>
<th>Scale</th>
<th>WMS-R M</th>
<th>WMS-R SD</th>
<th>MAS M</th>
<th>MAS SD</th>
<th>Difference M</th>
<th>Difference SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verbal</td>
<td>90.39</td>
<td>19.58</td>
<td>86.16</td>
<td>16.14</td>
<td>4.23</td>
<td>15.22</td>
</tr>
<tr>
<td>Visual</td>
<td>92.04</td>
<td>22.76</td>
<td>87.96</td>
<td>20.58</td>
<td>4.08</td>
<td>19.89</td>
</tr>
<tr>
<td>General</td>
<td>91.59</td>
<td>22.10</td>
<td>86.26</td>
<td>17.51</td>
<td>5.33</td>
<td>19.11</td>
</tr>
<tr>
<td>Attention/ST</td>
<td>85.47</td>
<td>20.69</td>
<td>88.84</td>
<td>19.28</td>
<td>-3.37</td>
<td>21.54</td>
</tr>
<tr>
<td>Delayed</td>
<td>83.73</td>
<td>18.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. WMS-R = Wechsler Memory Scales-Revised; MAS = Memory Assessment Scale; ST = Short Term.
For the two verbal scores, 24/51 (47%) scores were within 10 points of one another and 36/51 (71%) were within 15 points. For the two visual scores, 20/51 (39%) were within 10 points of one another and 26/51 (51%) were within 15 points of one another. For the short-term memory/attention variables, 18/51 (35%) were within 10 points and 25/51 (49%) were within 15 points. For General Memory, 22/51 (43%) were within 10 points and 30/51 (59%) were within 15 points.

Using a simple diagnostic criteria of classifying a case as impaired if it fell more than 1 standard deviation below the population mean (\(<85\)) and as normal if it fell above that point (\(>84\)), there was agreement on the general memory score of only 33/51 cases (65%). Results were slightly better for the Verbal indices (69%) and slightly worse for the Visual indices (59%).

Table 2 presents the intercorrelations among the two tests. All correlations were significant at \(p < .05\) \((r = .27)\). Fourteen of 20 correlations were at or above .43 \((p < .001)\) but none exceeded .59. The highest correlation was between MAS Short-term Memory and WMS-R Visual Memory. The highest correlation for the two Verbal scales were with one another (.57), but in all other cases the highest correlation was not with the score that would have been expected by the scale titles.

Because correlations between any two test scores reflect not only the underlying relationships between what the tests measure but also the inherent reliability of each test, such correlations are generally underestimates of the actual correlations between the underlying constructs. This can be addressed statistically by correcting the correlations by dividing each correlation by the square root of the product of the reliabilities for each score. The resultant correlation is nearly always higher (and may in some extreme cases exceed 1.0) and may better reflect the correlations between the constructs which are being measured. These corrected correlations are reported in Table 3. As can be seen, the pattern of the correlations remains generally the same but the specific correlations are increased up to a maximum of .75.

To study the influence of such variables as age, education, intelligence, gender, and diagnosis on these correlations, partial correlations were calculated using each of these variables individually as well as together. In the case of intelligence, partial correlations were conducted for Full Scale, Verbal, and Performance IQ separately. In the case of diagnosis, partial correlations were conducted for two variables. The first variable was coded one if the client had a head injury, but was otherwise coded zero. The second variable was coded one if the client had a stroke, but was otherwise coded zero. For the overall partial correlation, all variables were included except Full Scale IQ because of its close relationship with Verbal and Performance IQ.
In every case, the partial correlations were less than the original correlations, but in no case was the partial correlation more than .10 below the original correlation (which occurred when using Full Scale IQ as a covariate). The influence of each of these variables was minimal and in no case was the partial correlation significantly different from the original correlation.

**DISCUSSION**

The results clearly suggest that these two memory tests generate distinctly different results. Even when correcting for the reliability of the tests themselves, the degree of shared variance across scales with similar names ranges from 40% to 56%. Although each generally correlated best with its matching index, there were also substantial correlations with the other indices. Combined with the relatively low correlations between actual test scores, this suggests that these scores are not interchangeable and cannot be used to predict performance from one test to the other.

It is common that correlations between individual specific memory subtests may produce low correlations because such subtests may differ in content, scoring procedures, time delays, and other important factors. However, such low relationships are more of a concern when comparing overall summary measures. These are expected to be more consistent as they theoretically reflect an underlying general competency in memory tasks rather than specific subtest variations. The findings here indicate that the failure to find relationships between the tests cannot be attributed to any of the demographic or diagnostic factors, but (as suggested by the past research) is a function of the tests themselves.

An analysis of the specific patterns of intercorrelations using the reliability corrected data produces some interesting observations. The WMS-R Verbal Index shows good specificity with much stronger correlations for the MAS Verbal than the Visual Index, while the WMS-R Visual Index shows strong Verbal and Visual correlations with the MAS. This emphasizes the role of verbal skills in the WMS-R Visual index. The WMS-R Attentional measure best relates to the MAS Verbal index, but the two General Indices show surprisingly low correlations with one another. These correlations emphasize the differences between these two memory test batteries.

The concept of general memory was adopted for memory tests in the same way psychology adopted the idea of general intelligence. Such constructs were based on the recognition that individual tests of intellectual or memory skills may vary because of con-
tent and specific skills, but that individual tests could be combined to form more stable general indices (verbal, nonverbal, general). The present results, along with the past findings in this area, could be used to challenge a concept of general memory or any other specific summary indices of memory skills. Although subtests within these two batteries differ, they clearly sample the same domain of memory behavior. While the failure to find a relationship could be ascribed to defects in one of the tests, the data suggests it can best be interpreted to suggest that these summary indices are inefficient and perhaps misleading ways to present memory data.

Thus, although the concept of general intelligence appears to be useful even in wide varieties of brain-injured clients, these results raise the issue of whether a concept of a general memory index is a viable construct or simply one which obscures wide variation within and between individual brain-injured clients and which depends substantially on the specific content of the precise test battery employed. Reliance on the assumption that the results of any one memory test or battery will generalize to any other memory test (or the real environment) is questionable at best.

The clinical consequences of these findings and the previous studies are clear. Although both tests are routinely used to estimate general memory functioning, there may be widespread differences in the magnitude of the obtained scores. In our sample, more than one third of the possible pairs of scores demonstrated differences of greater than 15 points (1 SD), with one client showing a difference of over 40 points. There was agreement as to whether a score was normal or abnormal in less than two thirds of the cases.

These differences did not result, however, from one test simply generating consistently higher scores than the other. While the WMS-R scores tended to be slightly higher overall, more than one third of the cases showed higher MAS scores. In addition, if we used the Verbal-Visual difference score within each battery to suggest laterality of a lesion, the tests would have disagreed in 22 out of the 51 cases (43%; compared to a chance level of 50%).

It may be that there is no “general memory” or “verbal memory” but rather strengths and weaknesses in very specific types of memory that may vary not only with materials and content but with method of administration and scoring as well. This would have substantial implications for attempting to make clinical “predictions” and “prognosis” from limited samples of memory skills that are used to form a composite overall quotient.

Clinically, it may be both necessary and advantageous to use parts of both batteries (as well as other tests) to get a good picture of a wide sample of memory functions. After gathering such a sample, interpretation should focus on the pattern of results across specific memory subtests rather than attempting to reach conclusions on the basis of global values.

Additional research on the characteristics of the batteries (as well as other tests) and their use in specific populations and situations in predicting functional impairment is clearly needed. Caution should also be expressed in terms of the effect of any specific sample on results such as these, although the partial correlations suggest in our sample that such variables as age, education, intelligence, or diagnosis are not prominent intervening variables.

However, we must recognize that neuropsychological samples can differ widely in the types and severity of injuries and that these differences can substantially effect test results. These tests may be more similar in some populations than others (e.g., homogeneous groups like normals or left posterior temporal injuries). As a result, replication in more homogeneous brain-injured samples would be helpful in fully understanding the relationships of these two tests and the broader theoretical issues which may underlie these findings. The current results and those of past studies clearly suggest that these
general memory quotients should not be interchanged or considered to be equivalent in brain-injured clients.

REFERENCES


