Chapter 6

WECHSLER MEMORY SCALES

The Wechsler memory scales are individually administered, composite batteries designed to enable the user to better understand various components of a patient's memory. Now in its fourth edition (WMS-IV), it has been conormed with the WAIS-IV. Another major feature is that it provides a full range of memory functioning and has been carefully designed according to current theories of memory. As a result of these features, it typically is considered to be a core component of any thorough cognitive assessment (Rabin et al., 2005), which is reflected in its being ranked as the ninth most frequently used test by clinical psychologists (and third by neuropsychologists; Camara et al., 2000).

Memory complaints are extremely prevalent among client populations. These complaints are associated with anxiety, schizophrenia, depression, head injuries, stroke, learning disabilities, and neurotoxic exposure. For example, the impact of alcohol and other drugs on memory might need to be evaluated carefully. In occupational contexts, one similarly might need to evaluate the impact of exposure to industrial agents (lead, mercury, organic solvents) that can potentially result in impaired memory function. The increasingly aging population means that distinguishing normal memory loss from the early expression of dementia will become progressively more important. One crucial differential diagnosis is to distinguish between pseudodementia resulting from depression and Alzheimer's disease. As various drugs are developed for treating cognitive difficulties, it also will become increasingly important to monitor client improvement with a particular emphasis on memory functions. This array of symptoms suggests a developmental perspective; children are most likely to experience memory complaints related to learning disabilities, adults typically experience difficulties because of neurotoxic exposure or head injuries, and older populations have memory problems related to dementing conditions.

Many of the early conceptualizations of memory considered it a unitary process. From a practical assessment perspective it was not necessary to have a composite battery that assessed various components of memory. In contrast, more recent conceptualizations consider memory to have various components (see Lezak et al., 2004). One major distinction is between short-term and long-term memory (sometimes described as primary and secondary memory storage, respectively). For memory to be stored effectively, there also needs to be some active engagement on the part of the person. Thus, "working memory" was conceptualized as containing an executive component that initiated, monitored, and evaluated information. It also included an attentional component that had a limited capacity. A further well-supported distinction is between memory that is conscious and reflected in verbal reports of facts, events, and experiences (*declarative, explicit*, or *episodic* memory) versus memory that is more unconscious and measured implicitly by changes in performance (*procedural* or *implicit* memory). Finally, memory can involve various sensory components, particularly visual and auditory modes of processing.

HISTORY AND DEVELOPMENT

In some ways, the development of the Wechsler memory scales have paralleled the development of knowledge on memory. Each of the four editions has increasingly incorporated advances in the theoretical understanding of memory. The original Wechsler Memory Scale (WMS; Wechsler, 1945) reflected the earlier nonspecific conceptualizations of memory. It was composed of brief procedures on memory for number sequences, recalling a story, simple visual designs, and paired words. The advantage of using a variety of procedures was that a client might have intact memory for visual information but not auditory information or vice versa. Despite the fact that the early WMS procedures could be *logically* divided into visuospatial versus auditory tasks, the overall scoring was a composite Memory Quotient that, similar to the Wechsler intelligence scale Intelligence Quotients (IQs), had a mean of 100 and a standard deviation of 15. This was extremely valuable information for clinicians because they could easily compare a client's IQ with his or her Memory Quotient. Any large discrepancy could be investigated further. The WMS was also quite popular as it was a relatively brief procedure, typically taking about 15 minutes to complete. Because retesting a client would be likely to result in practice effects, it had the further advantage of having a parallel form. As a result of these advantages, it became a ubiquitous procedure among clinicians.

The WMS had surprising longevity, considering that a formal new version did not become available until 1987 (a 42-year interval). The WMS was limited, however, because it included unsophisticated methods of scoring the various procedures. In addition, the algorithms to determine the Memory Quotient were overly simple because they did not consider a sufficient number of client variables. The norms were derived from a small sample of 200 patients between the ages of 25 and 50 at the Bellevue Hospital in New York. Scores for either older or younger persons were extrapolated from this sample but were not based on actual participants. In addition, the alternate form was rarely used, and the research supporting it was quite limited. Finally, it did not reflect advances in knowledge related to memory processes.

One early attempt to correct for the deficiencies of the WMS was Russell's (1975, 1988) adaptation in which he administered two of the subtests (Logical Memory and Visual Reproduction) in an immediate format combined with a delay of 30 minutes. This allowed comparisons to be made between short-term and long-term memory. Research on Russell's WMS supported the predicted difference between left-hemisphere (relatively lowered auditory recall based on Logical Memory) and right-hemisphere (relatively lowered visual reproduction based on Visual Reproduction) lesions. Despite these advantages, the psychometrics were weak, and the test was poorly standardized. Unfortunately, it was titled the Wechsler Memory Scale-Revised (WMS-R). This could create confusion because The Psychological Corporation developed a full revision of the WMS that was also titled the *Wechsler Memory Scale-Revised*. Subsequent publications have attempted to clarify the two versions by referring to them as either Russell's WMS-R *or the* WMS-R.

The 1987 revision (Wechsler Memory Scales-Revised or WMS-R) was a significant improvement over the WMS. It had age-related norms for nine different age groups ranging between 16 and 17 years for the youngest group and 70 to 74 years for the oldest group. The standardization sample was composed of 316 persons, who had characteristics that closely approximated 1980 census data. There were approximately 50 subjects in each of the age groups. Whereas the WMS had only 1 composite Memory Quotient, the WMS-R had 12 subtests from which these five composite scores could be derived: General Memory, Attention-Concentration, Verbal Memory, Visual Memory, and Delayed Recall. Each of the index scores had a mean of 100 and a standard deviation of 15. This division into index scores is consistent with theories that have divided memory into short term versus long term (note the Delayed Recall used to assess long-term memory) and verbal/auditory versus visual (note the Verbal Memory and Visual Memory indexes).

Reliability of the WMS-R was generally low to adequate (internal consistency ranged from .44 to .88 and test-retest ranged from .51 to .60) The index standard error of measure ranged from a high of 8.47 for the Visual Memory Index to a low of 4.86 for the Attention-Concentration Index (Wechsler, 1987). Similar to studies on reliability, the validity of the WMS-R was good to adequate. Factor-analytic studies supported either a two- (Bornstein & Chelune, 1988; Roid, Prifitera, & Ledbetter, 1988; Wechsler, 1987) or three-factor solution (Bornstein & Chelune, 1988). A wide range of studies supported the ability of the WMS-R to distinguish between normal and clinical groups (A. Hawkins, Sullivan, & Choi 1997; Reid & Kelly, 1993; Wechsler, 1987), distinguish the relative severity of deficits based on subjective complaints (Gass & Apple, 1997), provide an index that related to client ratings of level of everyday memory (Reid & Kelly, 1993), and predict the degree of brain atrophy (Gale, Johnson, Bigler, & Blatter, 1995). In addition, the Attention-Concentration Index was found to be one of the most sensitive measures in identifying cognitive impairment (M. Schmidt, Trueblood, Merwin, & Durham, 1994). Despite a conceptual basis for believing that visual and verbal memory would relate to brain laterality of deficits, research on this has produced inconsistent results (Chelune & Bornstein, 1988; Loring, Lee, Martin, & Meador, 1989).

The WMS-R had clear advantages over the WMS because it had a far better normative base, was validated on diverse populations, had quite extensive studies performed on it, and divided memory into various indexes, thereby allowing the possibility for measuring various aspects of memory. Nevertheless, its weaknesses resulted in a revision within a relatively short period. One of the most serious limitations of the WMS-R has been the relatively low reliabilities of the subtests and indexes (Elwood, 1991), which likely significantly reduced the accuracy of measurements. In addition, the different indexes are probably not very good measures of specific components of memory. This is not to say they are not sensitive to both *general* cognitive impairment and the *degree* of that impairment. However, the specific nature of the impairment cannot be determined accurately by referring to the specific indexes despite the fact that the index names suggest that this differentiation can be made. Finally, current theories of memory were not used in the design of the WMS-R (Lichtenberger, Kaufman, & Lai, 2002).

The Memory Scale-III was published just 10 years after the release of the WMS-R (Wechsler, 1997a, 2002a). The new revision was designed not merely as a face-lift of the WMS-R but rather a "state of the art assessment instrument that comprehensively addresse[d] the complexity of brain/behavior relationships involved in learning and memory" (Edith Kaplan in the foreword to the WMS-III manual, p. iii). To accomplish this goal, new subtests were added, scoring procedures were made more sophisticated, stimulus materials were changed, and new index configurations were developed. This resulted in 6 primary and 5 optional subtests that were organized into index scores. Whereas the manual stated that it is possible to administer the 6 primary subtests in 30 to 35 minutes, research with a clinical population indicated that it took 42 minutes to administer the 11 primary subtests (see Axelrod, 2001). An abbreviated version was published in 2002 that reduced the administration time to 15 to 20 minutes (Wechsler Memory Scale-III Abbreviated; Wechsler 2002b).

One of the most important aspects of the WMS-III is that it was developed simultaneously with the WAIS-III. This enabled the two tests not only to share two subtests but also to be conormed. The normative sample consisted of 1,250 adults ranging between 16 and 89 years. Instead of 9 groups, as in the WMS-R, the WMS-III had 13 different groups. These groups not only had more subjects (50 in each group for the WMS-R versus 100 for the first 11 groups of the WMS-III) but also extended to a far higher age range (74 for the WMS-R versus 89 for the WMS-III). This is appropriate because one of the more important functions of memory assessment is to evaluate older clients.

The WMS-III had better reliability than its predecessor. The *WAIS-III/WMS-III Technical Manual* indicated that internal consistency for the primary subtest scores ranges between .74 and .93 for all age groups. As would be expected, the primary indexes had better internal consistencies of .82 or higher. Test-retest reliabilities for all age groups over a 2- to 12-week interval mostly ranged between .62 and .82 for the individual subtests and between .75 and .88 for the indexes. The

Technical Manual stated that even those subtests requiring the most judgment (Logical Memory I and II, Family Pictures I and II, Visual Reproduction I and II) had interscorer reliabilities above .90.

Factor analyses reported in the 1997 *WMS-III Technical Manual* concluded that a three-factor model composed of Working Memory, Visual Memory, and Auditory Memory most closely fit the data. In contrast, the 2002 *Technical Manual* found that a five-factor model composed of Working Memory, Auditory Immediate Memory, Visual Immediate Memory, Auditory Delayed Memory, and Visual Delayed Memory fit the age groups from 30 to 64 and 65 to 89. For ages 30 to 89, this closely corresponded to five of the eight index scores. The change in factor structure between the younger and older age groups is also consistent with findings that the components of memory become more clearly distinguishable ("dissociated") with age (Dolman, Roy, Dimeck, & Hall, 2000). Thus, the index scores might become more meaningful with older populations. However, it should be noted that for most populations, there was a correlation of .98 between Immediate Memory and General Memory (K. Hawkins, 1998; Weiss & Price, 2002). This finding suggests that much of the time, these indexes were redundant. Other research has found a four-factor model (Auditory, Visual, Working Memory, Learning; Price, Tulsky, Millis, & Weiss, 2002) and a two-factor model (Wilde et al. 2003).

There was ample evidence that the WMS-III effectively differentiated between clinical and normal populations. Various clinical groups (Alzheimer's, Huntington's, and Parkinson's disease; multiple sclerosis; chronic alcohol abuse; temporal lobe epilepsy; schizophrenia) consistently scored lower than the standardization sample (D. Fisher et al., 2000; K. A. Hawkins, 1998; Psychological Corporation, 1997). For example, Korsakoff's patients typically have severe difficulties with encoding and storing new information, but their attention and working memory are normal. This finding was reflected on the WMS-III index performances wherein Working Memory was in the normal range but all other index scores were in the impaired range (Psychological Corporation, 1997). In addition, patients with mild Alzheimer's disease scored in the 60 to 71 range for most of the primary indexes except for a mean score of 80 for Working Memory (Psychological Corporation, 1997) Fisher et al. found that patients with moderate to severe traumatic brain injury scored low on all indexes. This finding is also consistent with the finding that clinician ratings of the severity of brain injury were accurately reflected on scores on the WMS-III (Makatura, Lam, Leahy, Castillo, & Kalpakjian, 1999). The previous sampling of studies indicates that many of the predicted theoretical and clinical patterns of performance were supported.

Despite the relative success of the WMS-III, there were still a number of limitations. Some of the most important of these limitations were the equivocal factor structure, long testing time for older adults, subtest overlap with the WAIS-III, and problems with some of the subtests (Faces, Family Pictures, Verbal Paired Associates). In order to correct for these limitation in the WMS-III and refine memory assessment further, the Wechsler Memory Scale-IV was developed and published in 2009 (Pearson, 2009a, 2009b). To counteract the WMS-III limitations, the WMS-IV organized the indexes according to a clear factor structure, tried to reduce the administration time (especially for older adults), eliminated the subtest overlap with the WAIS-IV (deleted Digit Span and Letter-Number Sequencing), and eliminated some of the subtests specific to the WMS-III (Faces, Family Pictures, Spatial Span, Word Lists). Additional innovations were increasing the score ranges (ceiling and floor), focusing on *visual* working memory tasks (versus the WAIS-IV *auditory* working memory tasks), an additional new subtest (Spatial Addition), clarifying/simplifying some of the scoring procedures (Visual Reproduction), and modifying some of the WMS-III subtests (Logical Memory, Verbal Paired Associates, Symbol Span, Designs).

These efforts resulted in six subtests plus an additional optional Brief Cognitive Screen (see Table 6.1). Most of the subtests are administered a first time, then readministered following a 20- to 30-minute delay. Theoretical and factor analyses clustered the subtests according to five indexes (see Table 6.2). This is in contrast to the seven indexes for the WMS-III. Another major change was the development of one battery for adults (*WMS-IV: Adult Battery*; ages 16 to 69) and a slightly modified battery for older adults (*WMS-IV: Older Adult Battery*; ages 65 to 90, see Table 6.3). The advantage of the Older Adult Battery is that administration time is shorter but only four indexes can be calculated. In order to assist with interpretation, a series of "contrast scores" were included that determine whether differences between subtests were large enough to be interpretable. For example, a Visual Memory Index that was significantly higher than a client's Auditory Memory Index suggests that visual ability is a relative strength. Interpretation of these differences could assist with diagnosis and guide the development of treatment recommendations.

Table 6.1. WMS-IV subtest names, descriptions, and abilities measured

Source: Adapted from Table 1.1 from the WMS-IV Administration and Scoring Manual by Pearson, 2009, Pearson, Inc.

Subtest title	Description
Brief Cognitive Status (Optional)	Basic tasks include orientation (year, date, etc.), mental control (i.e. counting backwards), drawing a clock, recall of objects that had been named previously, inhibition of responses, and verbal production. A total score is derived to provide an estimate of any major cognitive impairment (*gross cognitive impairment*[b]).
Logical Memory (Ages 16–90)	I[b]: Two short stories are read and examinees are requested to repeat as many details as possible. Older adults (65–90) are presented with only one story presented twice (*short-term auditory–verbal memory*).
	II: Examinees are again asked to recall as many of the details as possible. (*long term auditory-verbal recall*) They are then asked yes/no questions about details in the stories (*long-term auditory–verbal recognition*).

Verbal Paired Associates (Ages 16–90)	I: A list of pairs of words are read (i.e. "dark . . . light"). The first word of the pair is read again and the examinee is asked to remember what the second word of the pair is (i.e. (dark . . .?") (*short-term auditory learning*).
	II: The examinee is again read the first word in the list and requested to recall the correct paired word (i.e. "light . . .?"). (*long-term auditory memory*) They are then read a list of paired words and asked which pairs were/were not read to them in condition I. (*long-term auditory recognition*). They are then asked (optional task) to say as many of the word pairs as they can recall (*long-term auditory recall information*).
Designs (Ages 16–69)	I: The examinee is shown a series of designs placed on a grid (10 seconds). The grid is removed and a new grid is presented along with a set of designs. Examinees must then identify where on the grid the original designs belong. (*short-term spatial memory*).
	II: Examinees are shown designs and grids and asked to reproduce the original placement of the designs on the grid (long-term visuospatial memory). They are then shown grids with designs on them and asked to recognize which designs are the same as in the immediate (I) condition (*long-term visual recognition*).
Visual Reproduction (Ages 16–90)	I: Five designs are shown to the examinee for 10 seconds. He or she must then draw the designs from memory (*short-term visual memory*).
	II: Examinees are requested to draw the original designs from memory (free recall task) (*long-term visual memory*). Next the examinee is asked to identify which of six designs on a page is the same as the design shown in condition I (*long-term visual recall*). Finally (optional task), examinees are shown the original designs and requested to draw them (copy phase; *visuospatial construction*).
Spatial Addition (Ages 16–90)	Examinees are show two grids with blue and red circles. They are then asked to add or subtract the location of the circles but are guided by a set of rules (*visual-spatial working memory*).
Symbol Span (Ages 16–90)	Examinees are first shown a page with a series of abstract symbols. They are then shown a different array of symbols and are asked to identify the correct order that they were presented on the original page (*visual-sequencing working memory).*

[a] I indicates that the procedure was administered and then their memory for the activity was assessed immediately afterward ("immediate" condition). In contrast, II indicates that a variation on the original procedure (I) occurred 20–30 minutes later ("delayed" condition).

[b] The phrases in italics indicate the type of memory function that is measured by the subtest.

Table 6.2. WMS-IV: Adult Battery (ages 16–69) indexes, primary subtests

Indexes	Subtests used to calculate indexes
Auditory Memory	Logical Memory I, Verbal Paired Associates I, Logical Memory II, Verbal Paired Associates II
Immediate Memory	Logical Memory I, Verbal Paired Associates I, Designs I, Visual Reproduction I
Delayed Memory	Logical Memory II, Verbal Paired Associates II, Designs II, Visual Reproduction II
Visual Memory	Designs I, Visual Reproduction I, Designs II, Visual Reproduction II
Visual Working Memory	Spatial Addition, Symbol Span

Table 6.3. WMS-IV: Older Adult Battery (ages 65–90) indexes and primary subtests

Indexes	Subtests used to calculate indexes
Auditory Memory	Logical Memory I, Verbal Paired Associates I, Logical Memory II, Verbal Paired Associates II
Immediate Memory	Logical Memory I, Verbal Paired Associates I, Visual Reproduction I
Delayed Memory	Logical Memory II, Verbal Paired Associates II, Visual Reproduction II
Visual Memory	Visual Reproduction I, Visual Reproduction II

The standardization sample for the WMS-IV was representative of the 2005 U.S. census of persons between the ages of 16 and 90. As such, the sample reflects the U.S. population based on age, sex, race/ethnicity, education level, and geographic region. A total of 1,400 examinees were included with 100 in each of 14 age bands. A wide variety of exclusion criteria were used to make sure that inappropriate persons were not included (i.e., persons with dementia, psychosis, medication that might impair their performance). The WMS-IV was conormed with the WAIS-IV, thereby making comparisons between the two instruments more appropriate.

RELIABILITY AND VALIDITY

The WMS-IV has good to excellent reliability. Subtest internal consistency among the normative groups was highest for

Visual Reproduction II (.97) and Verbal Paired Associates (.94) and lowest for Visual Reproduction (.74) and Verbal Paired Associates (.76; Pearson, 2009a). Internal consistency among a wide variety of clinical groups (i.e., Alzheimer's disease, traumatic brain injury) was even higher. As would be expected, the index score internal consistencies were all excellent, ranging from a high of .98 for Visual Memory (SEM = 3.04) to a low of .93 for Visual Working Memory (SEM = 3.71 Pearson, 2009a). Subtest test-retest reliabilities over a 14- to 84-day interval (*M* = 23 days) ranged between .77 for Spatial Addition and a low of .59 for Designs I (Spatial scoring category). Since the index scores have a larger number of items/subtests, it would be expected that their test-retest reliabilities would be higher than for the individual subtests. This finding was supported in that the index test-retest reliabilities ranged from a high of .83 (Auditory Memory and Visual Working Memory) to a low of .81 (Visual Memory and Immediate Memory).

Extensive and quite supportive evidence for the validity of the WMS-IV is presented in the *WMS-IV Technical and Interpretive Manual* (Pearson, 2009a), which can be organized according to content validity, correlations among the WMS-IV subtests/indexes themselves, factor analyses, correlations with other measures, and relationships with special groups (i.e., traumatic brain injury, intellectual disabilities, Alzheimer's disease). Content validity was based on a combination of research on previous versions of the Wechsler memory scales, expert review, client feedback, and research on the cognitive processes clients underwent when responding to the test items. Based on this information, the test items were modified and eventually evolved into the current WMS-IV. As a result, considerable efforts have been made to develop and refine their content validity.

The WMS-IV technical manual presents additional information related to correlations among the various subtest/index scores. Ideally, subtests/indexes on a test such as the WMS-IV would be expected to have positive correlations with similar tests (convergent validity) and low or nonexistent correlations with tests that do not seem similar (discriminant validity). Among the WMS-IV subtests/indexes themselves, they would all be expected to have at least some correlation with each other since each of the subtests measure, to some extent, the common variable of memory. This was indeed the case. In addition, similar abilities would be expected to have somewhat higher correlations than those with dissimilar abilities. Thus, a verbal subtest such as Logical Memory I was found to have a moderate correlation with Verbal Paired Associates (.44). In contrast, a lower correlation was found between Verbal Paired Associates (a verbal subtest) and the more visual subtest of Spatial Addition (.31). However, the correlation between the Delayed Memory and the Immediate Memory indexes was quite high, .87. This finding is high enough to suggests that the subtests may be measuring quite similar constructs. In contrast, the Auditory Memory and Visual Memory indexes were only moderately correlated (.64), suggesting that the auditory and visual memory components of the WMS-IV are adequately differentiated.

Previous factor analysis of earlier editions of the Wechsler Memory Scales resulted in inconsistent findings, which created considerable debate regarding the true structure of the instruments and called into question the accuracy of some WMS R/WMS-III index groupings. As a result, the WMS-IV closely adhered to a factor-analytically supported three-factor model comprised of Auditory Memory, Visual Memory, and Working Memory. Immediate Memory and Delayed Memory indexes were also included even though they were highly correlated. The rationale for including these last two indexes was that they have been found to be clinically useful constructs among some clinical groups where short-term acquisition of memory occurs but then decays over time (Millis, Malina, Bowers, & Ricker, 1999).

A wide number of correlations with other similar measures support the concurrent validity of the WMS-IV. For example, the correlation between the California Verbal Learning Test-II (learning trials 1–5) and the WMS-IV Auditory Memory Index was .63 (Pearson, 2009b). Similarly, correlations between the Children's Memory Scales (CMS; for 16-year-olds) and the WMS-IV ranged between a high of .74 (for Immediate Memory-CMS General Memory) to a low of .25 (for Auditory Memory-CMS Visual Immediate). In addition to correlations with specific measures of memory, the WMS-IV technical manual provides numerous correlations with more general ability measures. For example, index correlations with the WAIS-IV ranged from .71 (between Full Scale IQ and Visual Working Memory) to a low of .40 (between Auditory Memory and Processing Speed). As would be expected, the highest subtest correlations were between WMS-IV and WAIS-IV spatial measures (i.e., Spatial Addition-Block Design *r* = .51). One of the most important functions of a psychological test is to make accurate predictions related to everyday behavior. In support of this, the WMS-IV demonstrated positive correlations with a measure of independent living (the Independent Living Scales; i.e., Immediate Memory Index and Independent Living Scale Full Scale score *r* = .51). Finally, measures of achievement were moderately correlated with various scores on the Wechsler Individual Achievement Test-II (i.e., Visual Working Memory Index and Mathematics score *r* = .77).

The WMS-IV technical manual provides results for a wide variety of special groups. A sampling of some of these results is provided here for persons with mild intellectual disabilities, Alzheimer's disease, traumatic brain injury, and schizophrenia. Again, data has been highly supportive of the validity of the WMS-IV. As would be expected, scores for moderately intellectually disabled persons ranged between a low of 49 on Immediate Memory to a high of 54 for Auditory Memory. Patients in the early stages of Alzheimer's disease typically report memory as being their primary complaint. As a result, it would be expected that their WMS-IV scores would be lower than their WAIS-IV scores. This expectation was supported; mean WMS-IV scores for mild Alzheimer's patients ranged between 64 for Delayed Memory and 72 for Immediate Memory. In contrast, their mean WAIS-IV General Ability Index was a significantly higher 87. The subtests that were the most difficult for patients with mild Alzheimer's were Logical Memory (Scaled score *M* = 2.20) and Verbal Paired Associates (Scaled score *M* = 2.05). It was found that patients with mild to severe traumatic brain injury had WMS-IV scores that were significantly lower than the standardization group, ranging from a high of 86 for Visual Working

Memory to a low of 78 for Delayed Memory. Schizophrenics also had lower WMS-IV scores, ranging between a high of 82 for Visual Memory to a low of 77 for Immediate Memory.

Research reported in the *WMS-IV Technical and Interpretive Manual* amply support differentiating between normal and clinical groups. However, what is particularly crucial for the practicing clinician is to determine whether the individual indexes can accurately measure subcomponents of memory. Factor-analytic studies and determining whether patterns of scores match theories of memory are particularly important. As noted previously, the WMS-IV indexes were carefully organized according to the results of factor analysis. A further area of investigation is to see whether expected index patterns occur among specific types of clinical populations. For example, it would be hoped that the WMS-IV visual and auditory index scores would reliably differentiate patients with right-hemisphere brain damage (lower visual memory scores) from those with left-hemisphere brain damage (lower verbal/auditory scores). Some support for this was found among right temporal lobe epilepsy patients, who had lower Visual Memory scores (M = 86) compared to scores on Auditory Memory (M = 95). In contrast, left temporal lobe epilepsy patients had, as expected, lower Auditory Memory scores (M = 78) than Visual Memory scores (M = 98). Additional future research will no doubt investigate the extent to which the WMS-IV can differentiate discrete components of memory.

ASSETS AND LIMITATIONS

The WMS-IV is generally an excellent instrument capable of measuring a wide range of memory functioning. It has been based on theoretical research into the processes of memory, it has excellent standardization, and most research indicates solid empirical support. Subtests found to be problematic on the WMS-III were eliminated or modified, and a new subtest was added. There are only five indexes (four for the Older Adult Battery), and these are consistent with theories of memory, have generally good empirical support, and should make interpretation easier than the seven indexes developed for the WMS-III. The WMS-IV has been conormed with the WAIS-IV, which allows practitioners to make realistic comparisons between performance on the two instruments. In addition, the shorter format for older adults (ages 65–90) has advantage of making the Older Adult Battery more user friendly for this population. The WMS-IV is clearly an improvement over previous editions.

The scoring and administration of the WMS-IV is, for the most part, clearly described in the manual. The artwork is also clear, as is the Record Form. However, Logical Memory does not present guidelines regarding the speed at which the stories should be read. It also does not have guidelines for intonations, pauses, or inflections. Examiner variation in each of these areas may, therefore, result in the potential for error. For the WMS-III, Lichtenberger et al. (2002) recommended that the test developers introduce an audiotaped administration. This might be considered for the WMS-IV as well. A further issue with both Logical Memory I and II is its high degree of cultural loading; therefore, persons whose first language is not English may be disadvantaged on this subtest.

The original WMS had the advantage of taking only 15 minutes to administer. The WMS-III increased the administration time to an average of 42 minutes, but it may have actually taken up to 100 minutes for some clinical populations (Lichtenberger et al., 2002). Since the WMS-IV reduced the number of subtests and resulting indexes, it would seem reasonable that administration times would be shorter than for the WMS-III. However, administration times reported in the *WMS-IV Administration and Scoring Manual* (p. 14) indicated that for most participants, the total time for administering the WMS-IV Adult Battery was 75 to 77 minutes. The WMS-IV Older Adult Battery administration times were shorter, between 35 to 41 minutes for most participants. Currently these administration times should be considered quite tentative since the data was derived from inexperienced examiners (James Holdnack, personal communication, January 1, 2008). It is fair to assume that administration times would become faster with greater experience. Future studies will more precisely determine administration times among experienced examiners and for various clinical populations.

In the past, practitioners concerned with time efficiency used short forms of the WMS-III/WMS-R. For example, a three-subtest WMS-III short form consisted of Logical Memory, Verbal Paired Associates, and either Faces or Family Pictures and correlated at a .97 level with General Memory (and Immediate Memory; Axelrod, Ryan, & Woodward, 2001). A two-subtest short form composed of Logical Memory and Verbal Paired Associates had a quite similar correlation of .96 with General Memory (and Immediate Memory). These two short forms accounted for 95% and 87% of the variance in General Memory and Immediate Memory, respectively (Axelrod & Woodward, 2000). Concerns with developing a formal short form resulted in the *WMS-III Abbreviated* (Wechsler, 2002b), which used four WMS-III subtests that could be used to calculate visual and auditory memory indexes. Clearly clinicians are concerned with time efficiency and probably will use various short forms of WMS-IV. One feature that would have been good to include in the WMS-IV resources would have been guidance on how this could be best accomplished. Future research probably will establish these guidelines.

The WMS-IV was designed to be a fairly comprehensive measure of memory as reflected in the five indexes (four for older adults). There are also numerous methods for displaying and analyzing various combinations of scores, including index scores, scaled scores, percentiles, confidence intervals, graphical displays, subtest variability within indexes, contrast scores between subtests, contrast scores between indexes, and comparisons between the WAIS-IV and WMS-IV. The numerous options for displaying and organizing scores is clearly an advantage in that it allows clinicians to extend the possible meanings of test scores. For example, clinicians can determine whether an examinee's long-term (delayed) memory is significantly higher/lower than his or her short-term (immediate) memory. The *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009c) will provide additional strategies for analysis, including forensic applications, considerations for older populations, demographically corrected norms, and information on whether changes in scores on repeat testing represent reliable change. Difficulties with the sheer number of options are the extensive time required for training and an increased possibility that clerical errors will occur for scoring (Hopwood & Richard, 2005; Ryan & Schnakenberg-Ott, 2003). In addition, the large number of comparisons increases the likelihood that seemingly meaningful differences will occur simply due to chance ("random" significance). Both scoring errors and random significance may result in incorrect interpretations, leading to poor patient care. Clinicians must take particular care to make sure their interpretations are accurate.

An important unanswered question with the WMS-IV is the extent to which it actually measures the various components of memory. Its divisions (and corresponding indexes) into visual, auditory, and working memories are well supported. However, the distinction between immediate and delayed memory is not as well supported. This difficulty was also noted on the WMS-III. As a result, clinicians who note differences between the Immediate and Delayed Indexes should seek further support based on other measures and relevant history. A related and important issue is that the various components of memory (and corresponding indexes) are likely to perform differently across various clinical populations and age groups. A final unanswered question in need of further exploration is the extent to which the WMS-IV indexes relate to aspects of everyday memory. This question is often crucial for clinicians; many referral questions ask such things how much supervision the patient might need or whether the client can return to work. Given the considerable research that resulted from the WMS-III, these and many additional questions will no doubt be answered over the next few years.

USE WITH DIVERSE GROUPS

Since the WMS-IV is a measure of cognitive abilities, many of the considerations that apply to the Wechsler intelligence scales also apply to the WMS-IV, including level of acculturation, language proficiency, ensuring rapport, encouraging optimal effort, paying particular attention to nontest information, and caution interpreting the meanings of verbal (auditory) versus nonverbal (visual) comparisons. However, an important contrast between the Wechsler intelligence scales and the WMS-IV is that general ability measures, such as the Wechsler intelligence scales, typically are used to determine functional level compared to the general population. This is often the goal for psychoeducational assessments, assessing intellectual deficiency, and vocational assessments. In these situations, demographically adjusted norms are not recommended (see "Use with Diverse Groups" in Chapter 5) and the norms provided in the WMS-IV manuals should be adequate. In contrast, the WMS-IV is much more frequently used to determine neuropsychological diagnosis and level of impairment. In these cases, using demographically corrected norms are recommended (Heaton et al., 2003; Strauss et al., 2006; Wechsler, 2002a). The main rationale here is that, rather than normative comparisons being made, comparisons are typically made between a client's current status and a presumed premorbid level of functioning. Demographically corrected norms are more likely to give a more accurate estimate of premorbid level.

Research on the WMS-III found that the highest scores occurred among European Americans followed by Hispanics; the lowest scores were found among African Americans (Heaton, Taylor, & Manly, 2003; Heaton, Miller, Taylor, & Grant, 2004). Thus, using the norms provided in the *WMS-III Administration and Scoring Manual* may result in overestimating the number of African Americans and other minorities who are "impaired." Normative adjustment for age, education, gender and ethnicity were prepared for the WMS-III (Heaton et al., 2003; Heaton et al., 2004), and these authors are likely to develop demographically corrected norms for the WMS-IV. Demographic corrections will also be available in the upcoming *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009c). Until these norms are available, clinicians should proceed with caution when assessing persons from various ethnic groups. Clinicians should also take a careful history to make sure that ethnicity is indeed the actual variable that needs to be corrected. Other possibilities that might lower performance include quality of education, quality of the home environment, socioeconomic status, level/persistence of poverty, and health/nutritional status.

Some clients with physical, sensory, or language limitations might need special consideration with test administration and interpretation (see guidelines in Sattler, 2008, and Strauss et al., 2006). For example, it might be advisable to administer only the Auditory Memory and Symbol Span subtests to persons with physical difficulties (Pearson, 2009b). In contrast, it might be appropriate to give only the visual subtests and not the auditory (verbal) subtests to persons with language difficulties. If a client is not fluent in English, it might be advisable to administer the WMS-IV in a client's native language. However, the advantages of greater comprehension should be balanced with the reduction in test validity resulting from nonstandardized administration. Any modification of test administration should be noted on the Record Form and in the psychological report.

INTERPRETATION PROCEDURE

The WMS-IV measures a wide range of different functions. As a result, interpretation can be complex. The strategies to be described focus on index scores and comparisons between various combinations of index scores. Since referral questions frequently ask how a patient's memory compares with his or her overall ability, a section has been included on various relationships between scores on the WAIS-IV and various WMS-IV scores. This interpretive approach is designed to focus on the most important dimensions of the WMS-IV. More detailed information on a wider range of interpretive strategies can be found in the *WMS-IV Technical and Interpretive Manual* and will be available when a new edition of *Essentials of WMS-III Assessment* (Lichtenberger et al., 2002) becomes available with WMS-IV updates.

Often psychological reports include quite technical interpretative phrases. For example, a clinician might write something like "Ms. Memory's auditory immediate memory was statistically higher than her visual immediate memory." This may be an accurate interpretation, but, at the same time, it is likely to be understood by a relatively narrow group of readers. Typically a much wider audience will read the reports. As a result, clinicians may wish to use an interpretive phrase, such as "Ms. Memory's ability to recall information that has been spoken to her was much better (top 50% of the population) than her ability to remember information she has seen (lower 2% of the population)." This statement might be linked to actual test behavior, such as "She had a difficult time recalling details of designs she had been shown and then requested to draw." Another option would be to link test scores to examples of everyday behavior, such as "This suggests she would have problems remembering whom she had met previously or how she had gotten from one place to the next."

Prior to administering the WMS-IV primary indexes, clinicians may choose to give the optional Brief Cognitive Screening Exam (see Table 6.1). This example presents clients with a series of fairly basic tasks, such as recalling the date/day/month, counting backward, drawing a clock, or naming objects they had been previously shown. A total score can be used to obtain a general sense of any major cognitive difficulties. Thus this exam serves a similar function as a mental status examination (see Chapter 3). Scores are converted into classification labels for Average (25–100%), Low Average (10–24%), Borderline (5–9%), Low (2–4%), and Very Low (<2%). As can be seen, scores are not so much geared toward high and superior levels of functioning but more toward distinguishing various levels of poor functioning. If a patient obtains a low or very low score, the clinician may even decide not to proceed with the more demanding tasks of the primary WMS-IV subtests.

Prior to interpreting the WMS-IV, clinicians should thoroughly understand these essential principles:

- The WMS-IV index and subtest scores are arranged in the same format as the WAIS-IV. The indexes have a mean of 100 with a standard deviation of 15. The range (floor and ceiling) extends from 4 standard deviations above the mean (160) to 4 standard deviations below the mean (40). Percentile ranks are calculated as part of the standard scoring procedure. Subtest scores have a mean of 10 and a standard deviation of 3 (range is 1–19).
- Whereas the index and subtest scores provide information on how the patient performs in relation to their age-related peers, "contrast scores" measure the differences between two scores. One ability score is referred to as the "control," since it becomes the basis of comparison; the second one is referred to as the "dependent" measure. For example, a clinician might wish to note whether patients' memory for information they have seen (based on their Visual Memory Index) is significantly lower than for information they have heard (based on their Auditory Memory Index). In this case, the Auditory Memory Index is the control measure; Visual Memory is being contrasted with it and thus is the dependent measure.
- Memory is a complex function that can be influenced by factors other than memory itself. These factors include poor hearing, language impairment, visual difficulties, poor attention, general intellectual impairment, and impaired executive functioning. The possibility of comorbid conditions should always be considered when trying to determine the reasons for difficulties in memory (i.e., traumatic brain injury being comorbid with a learning disability or the side effects of medication). It is incumbent on clinicians to identify whether low scores on the WMS-IV are due to specific problems with memory or are secondary to one or more of the factors just listed.
- Patterns of WMS-IV scores cannot be used to diagnose specific clinical conditions or to make diagnoses. In other words, there is no WMS-IV score "fingerprint" that is specific to a given condition. However, when combined with other information, WMS-IV can be a potentially crucial source of information to help with diagnosis.
- When the WMS-IV is used with diverse groups, clinicians should carefully consider clients' acculturation and language facility. Proficiency with American English often allows examinees to understand the directions more easily and to encode, consolidate, and retrieve the information. This is likely to be more important for auditory information when compared to more visually oriented tasks.
- Clinicians should be careful not to overinterpret WMS-IV scores. Sometimes overinterpretation can occur from relying on a single low subtest score. In contrast, it is not unusual for average healthy persons to have one low score (Brooks, Iverson, Holdnack, & Feldman, 2008). Another potential source of overinterpretation is to confuse statistical significance with clinical significance. In other words, just because a formal calculation has found that a low score is statistically significant does not mean that this indicates an "impairment," "deficit," or "pathology." A closer inspection of cumulative percentages may reveal that many differences occur relatively frequently in the normal population. Thus normal patterns of individual differences in memory abilities should not be confused with cognitive impairments. Finally, overinterpretation potentially can occur when a large number of scores are analyzed such that some of the comparisons might be mistaken as being clinically significant when in reality they might merely be a random event ("random spurious significance").
- Different clinicians may vary in their determinations of whether a score is "impaired." As a general rule of thumb

WMS-IV index scores of 70 (2 standard deviations below the norm or lower second percentile) occurred in the most impaired clinical groups used in the validation studies (Alzheimer's disease, mild/moderate intellectual disability). In contrast, borderline to low average scores (70–85) occurred among clinical groups that were less impaired (schizophrenics, patients who had had their temporal lobes removed, moderate to severe traumatic brain-injured patients). However, this impairment should also be considered within the context of a client's overall abilities and occupation. For example, attorneys who rely heavily on auditory/verbal skills may have considerable difficulty functioning in the profession if their auditory/verbal memory performance has been lowered into the average to low-average range.

- The main focus of interpretation should be on the index scores, which represent robust, psychometrically sound measures. In contrast, subtest scores are not as psychometrically sound. As a result, there is a minimal emphasis on subtest interpretation in the next section. Instead, subtests should be used to develop tentative hypotheses in need of further support (seeTable 6.1). Subtests can also be used to make qualitative descriptions that can assist report readers to understand the types of behaviors on which the interpretations have been based (i.e., "For example, Mr. Memory did poorly at recalling details of a brief short story that was read to him").
- When minimal variability among the subtests compromises an index, the index itself can be interpreted with a high level of confidence. In contrast, high subtest variability suggests that the unity of the index might be compromised due to the disparate abilities that may be present. This does not invalidate the index, but it does challenge clinicians to determine why there was less consistency in performance.

INTERPRETING PATTERNS OF INDEX SCORES

The purpose of interpreting patterns of index scores is to better understand a person's memory-related strengths and weaknesses. Initially, clinicians might do this by noting the absolute values of the index scores. These values will provide comparisons with the standardization group. For example, a relatively low score on Visual Memory might indicate a relative weakness in this modality compared to the examinee's age-related peers. In contrast, a low score on Auditory Memory might suggest difficulties with recalling verbally meaningful information. However, clinicians should also be aware that fluctuations can occur for a number of different reasons. It is up to each clinician to evaluate these various possibilities by carefully integrating additional relevant information. Therefore, the index "interpretations" listed in this section should be considered tentative.

Another strategy is to compare various combinations of index scores. Instead of making normative comparisons, this level of interpretation compares clients with their own relative strengths and weaknesses (so-called ipsative analysis). The comparisons included here are based on those distinctions that are both most clinically useful and have received empirical and theoretical support. Thus a clinician may wish to know if visual or auditory modalities are relatively stronger or weaker. A second issue relates to whether a low score on visual working memory is really due to poor *working* memory or is rather the result of poor visual memory in general. A final assessment issue relates to differences between immediate (short-term) and long-term (delayed) memory. Thus, a delayed memory that is significantly lower than immediate memory suggests that there is a decay (forgetting) of memory over time. These three comparisons are described in the next section and can be formally determined by calculating and referring to the "Index-level contrast scaled scores" on the WMS-IV Record Form (and calculated by using conversion tables in Appendix G, Table G.12, of the *WMS-IV Administration and Scoring Manual*). Knowledge related to each of these components of memory has relevance for diagnosis and treatment planning as well as for understanding normal levels of strengths and weaknesses.

The next section briefly describes the index or contrast scores, then summarizes what a high or low score means along with some examples of how the abilities measured by the indexes might occur in everyday life. Finally, consideration is given to understanding more in-depth aspects of the index, especially when there is wide variability or scores among the subtests.

Auditory Memory Index

The Auditory Memory Index (AMI) requires people to attend to information that has been presented to them orally. They then must comprehend the information and repeat it immediately after it has been presented. They must later recall the information again after a 20- to 30-minute delay. One subtest of the Auditory Memory Index (Logical Memory) requires examinees to repeat a brief story that has been read to them. The second subtest (Verbal Paired Associates) requests examinees to learn pairs of words that belong together (i.e., "dark . . . *light*"; see Table 6.1).

High scores on the Auditory Verbal Index indicate that the person has excellent abilities attending to and recalling information that he or she has heard. In contrast, low scores suggest the person will have difficulties attending to and recalling information that he or she has heard. Everyday examples might include recalling material presented in lectures, oral directions, remembering conversations a few days later, recalling shopping items without the help of a list, or phone numbers that the person has been told. Persons with low scores might need to write down oral information. Assuming their visual memory is intact, they might also learn to translate the information into visual cues (visual reminders or "mind maps" of more complex information).

Although the WMS-IV norms provided in the administration/scoring and technical/interpretive manuals do not take into account sex differences, females typically perform better than males on Verbal Paired Associates. Analysis of the 1997 WMS-III found that females had a mean of 10.58 versus a mean of 8.46 for males on total recall scaled scores (Basso Harrington, Matson, & Lowery, 2000). This effect was moderately strong (approximately 3 subscale points) and should therefore be considered when making interpretations. Although formal analysis has not yet been done on the WMS-IV, this same sex difference is likely to occur and will be included as part of the demographically corrected norms in the upcoming *WAIS-IV/WMS-IV Advanced Clinical Solutions*.

One potentially useful behavioral observation is that excessive embellishment of stories on Logical Memory I and II may be a maneuver to compensate for or cover up difficulty remembering accurate information. Such embellishment may result in coherent elaboration or more illogical confabulations. A further behavioral observation is to note whether a client remembers primarily the global gist of the story as opposed to quite specific linear details. This result may suggest a global, holistic mode of processing as opposed to a more linear approach.

If all the scores for the subtests comprising Auditory Memory are relatively close together, interpretation of the Auditory Memory Index is fairly straightforward. In some situations, there might be fluctuations among the different scores (see Table F.1 in the *WMS-V Administration and Scoring Manual* and Record Form for "Subtest-level differences within indexes"), which would then require careful consideration of why these scores were discrepant. This can best be done by considering the differences and similarities of the subtests. The Logical Memory tasks require examinees to recall information that has been read in a short story format. In contrast, Verbal Associates requires examinees to learn pairs of words (i.e., "high . . . *low*") over four consecutive learning trials. These two subtests are similar in that the information is not only for auditory recall but for auditory *verbal* information (rather than musical or other types of sounds). However, they are different in that Logical Memory requires examinees to learn more complex, verbally relevant information whereas Verbal Paired Associates is for simple pairs of words and involves a prompt (one word is used as a prompt for the person to repeat the second/paired word). A discrepancy between these subtests might be explained by understanding the differences between them. For example, a relatively higher Logical Memory suggests that the person is better able to attend to and consolidate more complex, verbally meaningful information.

Another consideration when parsing discrepancies might be to note whether the delayed portions of the subtests (Logical Memory II, Verbal Paired Associates II) are significantly higher or lower than the immediate versions (Logical Memory I, Verbal Paired Associates I). In other words, are examinees' short-term memory better or worse than their long-term memory? For example, if the delayed versions were significantly lower, it suggests that examinees forget over time information what they have heard and initially learned. In order to help understand this distinction further, clinicians should note the Immediate and Delayed Memory indexes. In addition, they should obtain information from the client and informants to see if there are noteworthy examples of material that initially has been learned but seems to have been forgotten a short time later.

Visual Memory Index

The tasks on the Visual Memory Index require examinees to recall designs from memory and either to draw them or to place them in the correct spatial location (see Tables 6.1, 6.2, and 6.3). As a result, it measures their memory for both visual details and where visual information should be located. Since examinees must respond to information both immediately after it has been presented and after a 20- to 30-minute delay, the Visual Memory Index is a measure of both short-term and long-term visual memory.

High scores on the Visual Memory Index suggest that examinees have good abilities in recalling the details and location of information they have seen. In contrast, low scores indicate problems with remembering the details and location of information they have seen. Everyday examples might include remembering whom they had seen earlier in the day, where something has been left in the house, how they had gotten from one place to the next, or finding where their car was parked in a parking lot. Patients with low scores might compensate by writing events that have occurred in a diary or writing down directions in a verbal form.

If all the subtest scores in the Visual Memory Index are similar, index interpretation can be fairly straightforward because the abilities that have been measured are more likely to be unitary. Interpretation is more complicated in cases where there is wide variability among the subtest scores, suggesting that there may be quite specific visual memory difficulties (see Table F.1 in the *WMS-V Administration and Scoring Manual* and Record Form for "Subtest-level differences within indexes"). Visual Reproduction (I and II) requires examinees to look at a design and then draw it. Not only must they recall the design itself, but they also must go through the perceptual process of internally reconstructing it along with the external motor task of actually drawing the design. In contrast, Designs (I and II) requires examinees to look at the location of objects on a grid and later recall where the designs belonged on the grid. Thus the Designs subtest has more of a spatial component than Visual Reproduction. In contrast, Visual Reproduction has more of a psychomotor reconstructive component.

Discrepancies between scores on these subtests might be explained by understanding the differences in the tasks. For example, a significantly lower Visual Reproduction score compared with Designs might be due to examinees having difficulties with the task of having to draw the design. It should be noted that the Older Adult Battery does not include Designs I and II.

A further comparison might be made between the immediate and delayed portions of the subtests on Visual Memory. If scores for delayed visual memory tasks were significantly lower than those for immediate tasks, it suggests that the visual information that has been learned has been forgotten over time. Support for this, and other related inferences, should be obtained by noting performance on the Immediate and Delayed Indexes as well as obtaining information on the client's everyday behavior. For example, do other people in the client's life describe noteworthy instances in which the client seems to have rapidly forgotten information that has been seen (i.e., recalling who was at a meeting)?

Auditory Memory Index versus Visual Memory Contrast Scaled Score

One of the basic distinctions supported by WMS-IV factor analysis is between auditory and visual memory. The difference between these modalities (and the indexes that measure them) thus can be used to hypothesize relative auditory versus visual strengths and weaknesses. Thus it answers the question "Is this client stronger, weaker, or the same when auditory and visual memory abilities are compared?" A significant difference can indicate either lifelong patterns related to differences in abilities or acquired deficits in these modalities. The WMS-IV converts differences between index scores into scaled scores with means of 10 and standard deviations of 3 (see *WMS-IV Administration and Scoring Manual* Appendix G, Table G.12, and Record Form for "Index-level contrast scaled scores"). A score of 7 (16th percentile) indicates that the examinee's visual memory is in the low-average range compared with his or her auditory memory. Lower scores exaggerate this difference and suggest visual memory impairment. In contrast, a score of 13 or greater (84th percentile) suggests that the examinee's auditory memory is a relative strength compared with his or her visual memory.

Some research has found hemispheric laterality differences in patients with visual versus auditory memory impairments. Specifically laterality differences have been noted previously; patients with unilateral left-hemisphere damage have been found to do more poorly for verbal-auditory information than for visual information (K. Hawkins, 1998; Pearson, 2009a). For example, they would be expected to have particular difficulty when given verbal directions. In contrast, they might perform far better when shown a visual map of how to get from one place to the next. In contrast, patients with unilateral right-hemisphere damage would be expected to do more poorly on visual memory tasks. Thus, they would be expected to benefit more from auditory-verbal directions than from directions that were visually presented. However, visual memory performance was found to be the most sensitive to any type of brain damage, and patients with both unilateral right- and left-hemisphere damage performed poorly on visual memory types of tasks (Pearson, 2009a). If one modality was found to be relatively stronger than another, this stronger modality might be used to maximize learning. For example, if a person achieved a low score on auditory memory tasks, he or she might use learning strategies that capitalized on visual modes (or vice versa).

Visual Working Memory

The Visual Working Memory Index assesses a person's ability to temporarily attend to, organize, and manipulate visuospatial information (see Table 6.1). Note that it is not included on the Older Adult Battery. Visual Working Memory is similar to the WAIS-IV Working Memory Index in that both indexes evaluate the degree to which a person can hold and manipulate information for a short period of time. However, the WAIS-IV is specific to auditory-verbal material. It includes tests that require examinees to repeat a series of numbers, perform arithmetic problems that have been orally presented, and reorganize numbers and letters. In contrast, the WMS-IV has developed a quite different measure of working memory that is specific to visual information. The WMS-IV Visual Working Memory subtests require examinees to add/subtract visual information (Spatial Addition) and to arrange visual information into the correct sequence.

High scores on Visual Working Memory suggest that the person has excellent abilities holding and manipulating visual information. In contrast, low scores indicate the person has difficulties with visual information. Everyday examples might include being able to concentrate on a visual task without being distracted, staying focused on reorganizing furniture in a house, reorganizing the sequences of images on a computer screen, or tracking cards that have been seen in a card game.

Interpretation of the Visual Working Memory Index is made easier when the subtest scores are all within the same range, which indicates that the ability is more unitary. In contrast, subtest scores that are quite variable suggest that the ability may be due to more specific aspects of visual working memory (see Table F.1 in the *WMS-IV Administration and Scoring Manual* and Record Form for "Subtest-level differences within indexes"). As with previous indexes, it is incumbent on the clinician to parse these abilities in order to better understand the meaning of the index score. The Spatial Addition subtest requires examinees to look at two grids with different color circles. They must then add or subtract the location of the circles by following a set of rules. It is thus a spatial equivalent to the WAIS-IV Arithmetic subtest. Symbol Span shows examinees a series of abstract symbols on a page. They are then shown a different page with an array of symbols, including some from the previous page. They must identify which symbols had been shown to them previously and then indicate the order in which they were presented on the original page. The task is thus a visual analog to the WAIS-IV Digit Span subtest. Whereas Spatial Addition seems to be more of a visual "arithmetic" subtest (addition of the symbols is involved) Symbol Span involves more visual sequencing (items must be placed in the correct order). Clinicians should take these differences into account when understanding discrepant scores between the subtests. For example, a much lower Symbol Span subtest suggests that visual sequencing may be a particular difficulty for the examinee.

Visual Working Memory Index versus Visual Memory Index Contrast Scaled Score

An important consideration in understanding an examinee's performance on Visual Working Memory is whether it is due to poor working memory itself or rather to poor visual memory in general. In other words, is the problem a visual memory impairment beyond merely difficulties with working memory? Clinicians can determine the answer by checking to see if there is a significant difference between Visual Working Memory and Visual Memory. The WMS-IV converts differences between index scores into scaled scores with means of 10 and standard deviations of 3 (see *WMS-IV Administration and Scoring Manual* Appendix G, Table G.12, and Record Form for "Index-level contrast scaled scores"). A score of 7 (16th percentile) indicates that examinees' visual memory is in the low-average range compared to their relatively higher visual *working* memory. This finding suggests that their visual memory is probably the reason why their visual *working* abilities are low. For example, patients who have had their right temporal lobes removed were found to have contrast scores of 7.7 (Pearson, 2009a), which reflects poor visual memories. This somewhat low score is what would be expected, given that the right temporal lobe processes information related to visual memory. In contrast, a scaled score of 13 (84th percentile) suggests that clients have higher visual memory relative to their visual *working* memory. Thus it can be concluded that visual memory was not the reason why their visual *working* memory was low. In such cases, more faith can be placed in the interpretation that it was visual *working* memory itself that caused the low performance.

Sometimes scores on both the Visual Working Memory and Visual Memory indexes are low. In these cases, the contrast score is also likely to be low. This situation likely is caused by poor visual memory processes in general. It may also reflect interference on the tasks due to impaired visual perception.

Immediate Memory Index

Short-term (immediate) and long-term (delayed) memory are two of the crucial distinctions related to understanding memory. The WMS-IV Immediate Memory Index assesses how well examinees can recall both verbal and visual information immediately after the information has been presented. It includes tasks that require examinees to recall a story that has been read to them, learn words that are paired, draw designs from memory, and recall the correct location where designs should be placed on a grid (see Tables 6.1, 6.2, and 6.3).

High scores suggest that a client has good short-term memory for recalling information that they have heard (auditory-verbal) and information that they have seen (visual). In contrast, low scores indicate that a client has difficulty with these abilities. Everyday examples might include being able to recall a license plate they have seen, a phone number they have been told to remember, or where items belong in a cupboard. However, the above examples all relate to information that they can recall on a short-term (immediate) basis. It does not necessarily imply that they will be able to recall it over a longer duration.

Interpretation of the Immediate Memory Index is relatively clear when all the subtests scores are fairly even. This means that the ability being measured is unitary. In contrast, variation among the subtests means that the score might have occurred due to more specific factors (see Table F.1 in the *WMS-IV Administration and Scoring Manual* and Record Form for "Subtest-level differences within indexes"). The most obvious factor might have been due to differences in auditory as opposed to visual abilities. Thus clinicians might note scores on the Auditory and Visual Memory Indexes as well as the Auditory Memory Index versus Visual Memory Contrast Scaled Score. For example, if the Immediate Memory Index score was low but the Visual score was much lower than the score for the Auditory Index, it suggests that the relatively poor performance on visual material was mainly responsible for the poor performance on the Immediate Memory Index.

Delayed Memory Index

In addition to measuring short-term (immediate) memory, the WMS-IV also measures the extent to which examinees retain the information. This is measured by requesting examinees to recall details of the information that has been presented to them in each of the primary subtests following a 20- to 30-minute delay. Clients must first attend to the information and then encode, consolidate, retrieve, and provide the correct answer after the delay.

Persons with high scores on Delayed Memory can be expected to be good at retaining and retrieving information they have learned. In contrast, persons with low scores can be expected to have difficulty retaining and retrieving information. Everyday examples might include long-term recall of instructions, times of meetings, where things should be placed in cupboards, and repeating jokes or stories they have heard.

The Delayed Memory Index is comprised of many memory components since it requires that clients first accurately encode and consolidate short-term ("immediate") information and then must recall it at a later time. It involves both visual and auditory information. As a result, it can be conceptualized as a measure of global memory (similar to the General Memory Index on the 1997 WMS-III; James Holdnack, personal communication, January 6, 2009).

As with the previous indexes, subtest scores on the Delayed Memory Index that are similar mean that the ability is a unitary construct. As a result, the index can be interpreted with confidence. In contrast, a high level of variability challenges the clinician to determine if there are more specific abilities that resulted in the subtest variability (see Table F.1 in the *WMS-IV Administration and Scoring Manual* and Record Form for "Subtest-level differences within indexes"). Doing this can be particularly challenging since the Delayed Memory score represents the end product of a potentially wide number of processes (attending to the test material, good visual/auditory perception, encoding, consolidation, retrieval, expressing the response). As a result, a wide variety of reasons can disrupt performance on Delayed Memory. One potentially important distinction is between visual and auditory-verbal modalities. Clinicians might inspect possible differences in subtest scores to determine if the visual subtest scores are higher/lower than the auditory-verbal scores. They might also check the Visual and Auditory Index scores and note especially the Auditory Memory Index versus Visual Memory contrast scaled score. If, for example, the auditory scores were significantly lower than the visual scores, it suggests that the client's auditory memory might be the reason for the low score on the Delayed Memory Index.

One possible explanation for a low Delayed Memory score might be that people had a difficult time retrieving the information even though they had learned it successfully. Thus they might not have been able to "recall" the correct answers, but, if given a chance, they might be able to "recognize" the correct answers. This finding can be parsed by having administered the recognition procedures (see WMS-IV Record Form "Process Score Conversion" section). They might have scored poorly in the standard scores on the primary subtests, but their recognition scores might have been quite good. Everyday examples might be persons who have extensive tip-of-the-tongue struggles or who say "I know I know the answer but I just can't remember it," or can get the correct answer with minor prompting.

Immediate Memory Index versus Delayed Memory Index Contrast Scaled Score

The distinction between short-term (immediate) and long-term (delayed) memory is often important for practicing clinicians. Thus a referral question might be "Does this patient have impairments in forgetting material she has previously learned?" In contrast, the memory of some examinees actually might improve over time since they need the extra time to consolidate the information. A question relating to this issue might be: "Does the examinee have an improvement in memory over time?" If delayed memory is considerably lower than immediate memory (see *WMS-IV Administration and Scoring Manual* Appendix G, Table G.12, and Record Form for "Index-level contrast scaled scores"), it suggests that the person can learn material initially but that the information decays over a period of time. It should be stressed in this regard that performance on immediate memory becomes the benchmark for how much information has been lost. In other words, unless a person has learned something initially, there is nothing to lose. The exception might be that a person has acquired information but then may not be able to recall it due to poor retrieval. However, recognizing the information is generally a much easier task; the person might be able to recognize information accurately even though he or she may not be able to recall/retrieve that information. A number of procedures are available on the WMS-IV to contrast a person's recall with recognition (see Record Form for relevant "Process Score Conversions" and "Subtest-Level Contrast Scaled Scores" scores).

One issue is that factor analysis of the immediate/delayed distinction on the WMS-IV may not be as strong as would be optimal (Pearson, 2009). This issue is consistent with the finding that there was a quite high (.87) correlation between the Immediate Memory and Delayed Memory Indexes, as was also the case for the WMS-III (K. A. Hawkins, 1998; Millis et al., 1999; Weiss & Price, 2002). Clinically this means that most of the time, the index scores will not reveal a significant difference between the two abilities. Despite these findings, it was decided to include Immediate and Delayed Memory Indexes on the WMS-IV, since they can still provide potentially useful clinical information (Millis et al., 1999). In other words, there may be some populations (i.e., Korsakoff's disease, older populations) who can repeat information they have just seen or heard but forget it a short time later. This situation can be suspected in cases where informants state that clients seem to understand and can repeat information but cannot say what they saw or heard the next day.

COMPARING SCORES ON THE WAIS-IV AND THE WMS-IV

One of the most important referral questions is whether a client's memory is low compared to his or her other, more general abilities. The question may be phrased in this way: "Is this client's memory consistent with his general level of cognitive functioning?" A comparison between performance on the WMS-IV and performance on the WAIS-IV allows a clinician to answer this question. It thus places memory performance within a larger context. Thus general ability (WAIS-IV) provides a baseline or comparison point for evaluating the extent that memory (on the WMS-IV) has declined. This assumes, of course, that the more general abilities measured on the WAIS-IV are relatively stable. In contrast, memory is usually considered to be more sensitive to decline, a finding consistent with the fact that memory often is patients' main concern. For example, memory is usually the main complaint reported by patients with traumatic head injury or Alzheimer's disease. In contrast, their other more general abilities tend to be more stable.

Previous versions of the Wechsler memory scales used a total or general score that could be compared with a person's Full Scale IQ. The original WMS allowed practitioners to calculate a "Memory Quotient," and the 1997 WMS-III had a General Memory Index. Differences between the general ability and the memory scores were fairly easy to explain to referral sources or family members. Clinicians could use a phrase such as "Joe's overall mental abilities were in the average range (50th percentile) but, in contrast, his memory was much lower since he was in the bottom 5% of the population." Instead, the WMS-IV makes comparisons between each of its index scores and the WAIS-IV General Ability Index. The General Ability Index was used since it is comprised of verbal (Verbal Comprehension Index) and performance/nonverbal (Perceptual Reasoning Index) abilities. Both these abilities/indexes tend to be fairly resistant to the impact of most clinical disorders. As such, they are quite stable. In contrast, speed (Processing Speed Index) and attention/manipulation (Working Memory Index) are quite sensitive to a variety of clinical conditions. Thus many of the conditions that would be likely to lower memory also would be likely to lower speed and attention (Processing Speed and Working Memory indexes). The General Ability Index is likely to be a more stable benchmark for comparison than the Full Scale IQ, which includes all four WAIS-IV indexes (including measures of speed and attention). In other words, the General Ability Index–WMS-IV index comparisons are likely to be more sensitive to difficulties with memory when compared to using Full Scale IQ–WMS-IV index comparisons.

The Record Form allows for completion of a wide number of ability-memory comparisons (see the Record Form "Ability-Memory Analysis" section that uses Tables B.1–B.16 on pp. 200–218 of the *WMS-IV Technical and Interpretive Manual*). Most clinicians focus primarily on comparisons between the General Ability Index and the WMS-IV index scores. For that reason, we describe only those scores in this chapter. Some examiners may want to make more detailed comparisons between additional combinations of the WMS-IV index and WAIS-IV index scores. All differences are converted to contrast scaled scores with a mean of 10 and a standard deviation of 3. A score of 7 (16th percentile) indicate that the memory index is unexpectedly low (a relative weakness) compared with the General Ability Index. Lower scores exaggerate this difference and suggest the possibility of memory impairment specific to the index. In contrast, a score of 1 or greater (84th percentile) suggests that the examinee's memory index is unexpectedly high (a relative strength) compared with his or her General Ability Index. One caution is that, with so many potential comparisons, the possibility of random spurious significant differences increases. In other words, some of the "significant" differences may not actually be clinically accurate descriptions of the client. As a result, clinicians should be careful not to overinterpret the difference scores.

The next descriptions are quite brief. Clinicians who want more detailed interpretations can read information under each of the WMS-IV indexes; that material includes a description of the index, a listing of the types of tasks involved, a bri interpretation of the meaning of high/low scores, and everyday examples (see "Interpret Patterns of Index Scores" earlier i this chapter).

General Ability Index versus Auditory Memory Index

A low score (below 7) indicates that the information that clients have recalled based on having heard it is a weakness compared with their overall ability. High scores (above 13) indicate that their memory for information they have heard is a relative strength.

General Ability Index versus Visual Memory Index

A low score (below 7) indicates that the information that clients have recalled based on having seen it is a weakness compared with their overall ability. This index has been found to be one of the most sensitive measures of impairmen (Pearson, 2009a). High scores (above 13) indicate that their memory for information they have seen is a relative strength.

General Ability Index versus Visual Working Memory Index

A low score (below 7) indicates that clients' ability to concentrate on, hold, organize, and manipulate complex visual information is a relative weakness compared with their overall ability. They are likely to have difficulty working with both where the information was located ("visual space") and the details of what was seen ("visual details"). High scores (above 13) indicate that holding and manipulating visual information (both spatially and for details) is a relative strength.

General Ability Index versus Immediate Memory Index

A low score (below 7) indicates that clients' short-term ("immediate") memory for information they have seen or heard is a relative weakness compared with their overall ability. High scores (above 13) indicate that their short-term ("immediate") memory is a relative strength.

General Ability Index versus Delayed Memory Index

A low score (below 7) indicates that clients' long-term ("delayed") memory for information they have seen or heard is a relative weakness compared with their overall ability. This measure is one of the more clinically sensitive measures (Pearson, 2009a). High scores (above 13) indicate that their long-term ("delayed") memory for information they have seen or heard is a relative strength. "Long"-term assessment on this index was based on a 20- to 30-minute delay. Since the Delayed Memory Index is comprised of many memory components, it can be conceptualized as a measure of global memory (similar to the General Memory Index on the 1997 WMS-III). Given its sensitivity combined and the fact that it is a global measure of memory, it should be one of the most important comparisons.

ADDITIONAL CONSIDERATIONS: MALINGERING AND EVALUATING CHANGE

Secondary gain is frequently an issue for assessments related to personal injury litigation, workers' compensation, long-term disability, or defendants in criminal injury proceedings. Due to the potential for gain, malingering is a distinct possibility. Surveys of neuropsychologists found that estimates for feigning deficits were as high as 30% among personal injury and workers' compensation cases (Mittenberg et al., 2002) and up to 40% for litigants involved with traumatic brain injury (Larrabee, 2005). Memory problems are particularly likely to be exaggerated since they are often the most frequently reported problems among these populations. As a result, clinicians need to be particularly careful to evaluate the validity of a client's complaints. Terms that are similar to malingering but somewhat more neutral include "suboptimal performance," "inconsistent effort," or "feigning."

A number of specialty instruments are available to detect suboptimal cognitive effort and are recommended to help make a more definitive assessment (see Boone, 2007; Larrabee, 2005; Strauss et al. 2006). Best practice even requires multiple measures to be used. Possible strategies to detect malingering on the WMS-IV might be to focus on the Logical Memory Delayed Recognition task that requests clients to state whether (yes or no) an item was included in one of the previously read stories. Because random guessing would produce a score of 50%, scores of less than this suggest that the client is malingering (see Killgore & Dellapietra, 2000). Malingering may also be suggested if recognition does not improve in comparison to recall because recognition tasks are easier than free recall tasks. A final quite general indicator is dramatic differences between a person's day-to-day functioning (based on corroborating sources) and performance on WMS-III measures. The *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009c) will provide additional strategies to detect malingering, including analyses of guessing for Logical Memory Recognition, Verbal Paired Associates, Designs Spatial, and Designs Content (James Holdnack, personal communication, January 7, 2008). If an examinee performs lower than guessing on the listed subtests, malingering is a possibility.

Sometimes Wechsler memory scale scores are used to document deterioration or to monitor improvement. It is tempting to peruse pretest and posttest scores and quickly infer that some sort of actual change has occurred in the patient's level of functioning. For example, a client might have had a WMS-IV Delayed Memory Index score of 80 directly after a head injury and, 3 months later, achieved a score of 85. It might be inferred that the patient's memory has improved. However, this finding does not take into consideration factors such as practice effects, regression to the mean, or the relative reliability of the measure. The improvement between the pretest of 80 and the posttest of 85 might simply be the result of the patient's practicing the tasks 3 months previously, or the difference might simply be measurement error (reflected in its test-retest reliability). The *WAIS-IV/WMS-IV Advanced Clinical Solutions* (Pearson, 2009c) will provide strategies for calculating whether actual change has occurred ("reliable change index"). However, these calculations will account for the unreliability of the instrument, which does not necessarily mean that the personal or social significance of the change in scores has been demonstrated (see Beutler & Moleiro, 2001). Determining the personal and clinical meaning of changed scores requires clinicians to integrate information from a wider variety of sources to support any inferences related to actual change in the client's functioning.

RECOMMENDED READING

Heaton, R. K., Taylor, M. J., & Manly, J. (2003). *Demographic effects and use of demographically corrected norms with the WAIS-III and WMS-III.* In D. S. Tulsky, D. H. Saklofske,G. J. Chelune, R. K. Heaton, & R. J. Ivnik (Eds.), *Clinical interpretation of the WAIS-III and WMS-III* (pp. 181–210). San Diego: Academic Press.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *Memory*. In O. Spreen & E. Strauss (Eds.), *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed.) (pp. 679–686). New York: Oxford University Press.