11 | Memory I: Tests

Memory is the capacity to retain information and utilize it for adaptive purposes (Fuster, 1995). Efficient memory requires the intact functioning of many brain regions, including some that are especially susceptible to injury or disease. Moreover, many common neurological and psychiatric conditions produce a decline in that efficiency. One in three individuals aged 75 and above without dementia complains about memory deficits (Riedel-Heller et al., 1999). Moreover, memory complaints in outpatient settings may be the most frequent reason for neuropsychological referral. Thus memory assessment is often the central issue in a neuropsychological examination.

The use of the same word—memory—to identify some very different mental activities can create confusion. Patients as well as some clinicians lump many kinds of cognitive dysfunction under the umbrella of "memory impairment." In contrast, some patients whose learning ability is impaired claim a good memory because early recollections seem so vivid and easy to retrieve. Many older adults report memory problems when referring to an inability to retrieve common words or proper names consistently. This word finding difficulty—*dysnomia*—can occur along with efficient retrieval of episodic memories and, conversely, patients who have problems recalling episodic memories are not necessarily dysnomic. Deficits in processes outside the memory system can affect memory performance: these include attention and concentration, information processing speed, organization, strategy, effort, and self-monitoring (Ganor-Stern et al., 1998; Howieson and Lezak, 2002). Maintaining terminological distinctions between the different aspects of memory and the other functions necessary for efficient memory will help the clinician keep their differences in mind when evaluating patients and conceptualizing findings and theory.

Because memory impairments can take a variety of forms, no one assessment technique demonstrates the problem for all patients. Knowledge about presenting complaints, the nature of the brain injury or the neuropsychological syndrome, and the differing etiologies of memory disorders should guide the selection of memory tests. In every examination the examiner's choice of memory tests should depend upon clinical judgment about which tests are most suitable for answering the question under study for *this patient*. Therefore this chapter presents the tests in most common use plus a few of particular interest because of their potential research or clinical value, or because the format merits further exploration. Most tests of short-term and working memory are discussed in Chapter 9, pp. 351–358 because of their kinship to attentional processes (Cowey and Green, 1996; Howieson and Lezak, 2002). At least as many more memory tests show up in the literature as are described here.

EXAMINING MEMORY

For most adults it is useful to begin the examination of attention before proceeding with memory tests because of its fundamental role in memory performance. If someone performs poorly on simple attentional tasks such as span of immediate verbal retention (e.g., Digit Span Forward) or simple mental tracking (e.g., counting backwards by 3s or 7s), it may not be possible to get a valid measure of retention. For some patients, it may be necessary to delay the examination until a different time or under different circumstances in order to get a valid memory assessment.

A comprehensive memory evaluation should include (1) orientation to time and place; (2) prose recall to examine learning and retention of meaningful information which resembles what one hears in conversation, such as Wechsler's Logical Memory stories or other stories developed to test verbal recall; (3) rote learning ability which gives a learning curve and is tested with both free and recognition trials, such as the Auditory Verbal Learning Test or the California Verbal Learning Test; (4) visuospatial memory such as the Complex Figure, followed by a recognition trial when available; (5) remote memory, such as fund of information; and (6) personal—autobiographical—memory. All tests designed to measure learning should include one or more trials following a delay period filled with other tasks to prevent rehearsal, and both free recall and recognition or cued recall should be examined following the delay. When tests calling upon a motor response are not appropriate or produce equivocal findings, visual recognition tests can be substituted. A unilateral lesion may affect recall of verbal and nonverbal material differentially with left hemisphere lesions more likely to compromise verbal memory and right hemisphere lesions particularly disrupting visuospatial recall (Abrahams et al., 1997; Loring, Lee, Martin, and Meador, 1988; Milner, 1958; Sass et al., 1995). Thus, inclusion of both verbal and visuospatial tests is necessary for the assessment of memory problems specific to the type of material being learned.

When assessing memory the examiner should also compare aspects of cognition that are not heavily dependent on memory with the memory performance. The examiner can usually integrate the memory tests into the rest of the examination so as to create a varied testing format, to avoid stressing those memory-impaired patients who may be concerned about their deficits, and to use nonmemory tests as interference activities when testing delayed recall. Much mental status information can be obtained quite naturalistically during the introductory interview. For example, rather than simply noting the patient's report of years of schooling and letting it go at that, the examiner can ask for dates of school attendance and associated information such as dates of first employment or entry into service and how long after finishing school these events took place. Although the examiner will frequently be unable to verify this information, internal inconsistencies or vagueness are usually evidence of confusion about remote personal memory or difficulty retrieving it.

Three memory testing procedures must be part of every aspect of memory assessment if a full understanding of the patient's strengths and weaknesses is to be gained. (1) Immediate recall trials are insufficient tests of learning, retention, or the efficiency of the memory system. To examine learning (i.e., whether material has been stored in more than temporary form), a delay trial is necessary. In addition, a few patients who process information slowly will recall more on a delay trial than initially, thus demonstrating very concretely their slowed ability to digest and integrate new information. Freed, Corkin, and their coworkers (1989) call this late improvement *rebound* when it follows diminished performance on an early delay trial. (2) Interference during the delay period will prevent continuous rehearsal. Absence of some intervening activity between exposure to the stimulus and the subject's response leaves in question whether recall following delay was of learned material or simply of material held in continually rehearsed temporary storage. (3) When the subjects' recall is *below normal limits,* it is not possible to know whether reduced retrieval is due to a learning impairment or a retrieval problem. In these situations, some means of assessing learning that bypasses simple recall must be undertaken to decide this critical issue. The most direct of these, and often the simplest, is to test learning by recognition. Other techniques include use of cues, comparing recall of meaningful material with recall of meaningless material (as meaning can serve as an internal cue), or the method of *savings* (in which the patient is given the same test at a later time to see whether the material is learned more quickly the second time, i.e., as a measure of forgetting; see p. 478).

The examiner needs to take special care to recognize when a poor performance on memory tests occurs because of impairment from other possible sources of reduced functioning. Elderly persons frequently have vision or hearing problems that adversely affect proper registration of the stimulus. Patients with frontal lobe or certain kinds of subcortical damage may lack the spontaneity or drive to reproduce all that they remember. When the patient exhibits diminished initiation or persistence, the examiner should press for additional responses. With story material, for example, it may be possible to encourage a complete recall by asking such questions as, "How did it begin?" or "What was the story about?" or "What happened next?" and so on. When the task involves reproduction of configural material, the patient can be encouraged with, "That's fine; keep going," or by being asked, "What more do you remember?" Depressed patients also may lack the drive to give their best performance on memory tests and may benefit from supportive prompting.

Memory tests, perhaps more than most cognitive tests, are influenced by practice effects (see pp. 116–117; McCaffrey, Duff, and Westervelt, 2000b). Many patients are examined repeatedly to measure their course over time or to examine the validity of data in forensic cases. In these cases it is desirable to have alternate test forms of equivalent difficulty for reassessment purposes. Using different but equivalent forms of verbal memory tests can reduce if not eliminate significant practice effects. A small practice gain is more likely to occur on visuospatial memory tests even when different forms are used due to "learning to learn" the even less familiar visuospatial procedures (Benedict and Zgaljardic, 1998). There has been a paucity of memory tests with multiple equivalent forms although more and more they are being developed.

VERBAL MEMORY

While many verbal memory tests are available, only a few have reliable norms based on careful standardization. Even with many tests available, the examiner may occasionally find that none quite suits the needs of a particular patient or research question, and will devise a new one. Verbal memory tests are presented here by content in order of increasing complexity. Not every kind of test is represented under every content heading, but taken altogether, the major techniques for examining verbal memory functions are reviewed.

Verbal Automatisms

Material learned by rote in early childhood and frequently used throughout life is normally recalled so unthinkingly, effortlessly, and accurately that the response is known as an *automatism*. Examples of automatisms are the alphabet, number series from 1 to 20 or 100 by 10's, days of the week and months of the year, a patriotic slogan or a long-practiced prayer. Automatisms are among the least perishable of the learned verbal habits. Loss or deterioration of these well-ingrained responses in nonaphasic patients may reflect attentional disturbances or fluctuations of consciousness in acute conditions. It occurs in nonacute conditions only when there is severe, usually diffuse, cerebral damage, such as in advanced dementia. To test for automatisms, the examiner simply asks the subject to repeat the alphabet, the days of the week, etc. More than one error usually indicates brain dysfunction.

Letters and Digits

Brown-Peterson Technique (L.R. Peterson and Peterson, 1959; L.R. Peterson, 1966; Baddeley, 1986)

To study short-term retention, a popular method has been this distractor technique which is also called the *Peterson task* (e.g., Baddeley, 1986), the *Peterson and Peterson procedure* (e.g., H.S. Levin, 1986), and other variations on the Peterson name, or it may be referred to as *Auditory Consonant Trigrams* (*ACT*) (Mitrushina, Boone, and D'Elia, 1999; Spreen and Strauss, 1998, use the acronym CCC). The purpose of the distractor task is to prevent rehearsal of material being held for short-term retention testing. The Brown-Peterson technique is included in this chapter because it assesses short-term memory and, unlike many working memory tests (covered on pp. 358–364), it does not require manipulating information held in memory. Rather, the task is to recall the information as presented.

Upon hearing (or seeing) three consonants presented at the rate of one-per-second, the subject is required to count backward from a given number until signaled to stop counting, and then to report or identify the stimulus item. The examiner tells the subject the two- or three-digit number from which to begin counting immediately after saying or showing the test item. For example, the examiner says, "V J R 186" when the subject begins counting—"185, 184," etc.—until stopped at the end of a predesignated number of seconds and is expected to recall the item (see Table 11.1 for a consonant trigram stimulus format). With this technique, normal subjects have perfect recall with no distraction delay: they recall about 80% of the letters correctly with a distraction duration of 3 sec, approximately 70% to 80% correct recall with 9 sec delays (Stuss, Stethem, and Poirier, 1987). Longer durations produced a wider range of normal performances: from 50% to 80% with delays of 18 sec, and around 67% when the delay is as long as 36 sec. Giving five trials of three consonants each for a total of 15 possible correct responses at each delay interval, Stuss and his colleagues report standard deviations typically within the 1.6 to 2.8 range for the 9 sec delay, increasing to 2.1 to 3.6 for the 36 sec delay for various age groups (see also N. Butters, Sax, et al., 1978; Mitrushina, Boone, and D'Elia, 1999, report on two versions of the test).

TABLE 11.1 Example of Consonant Trigrams Format*

Stimulus	*Starting Number*	*Delay (sec)*	*Responses*
QLX	—	0	
SZB	—	0	
HJT	—	0	
GPW	—	0	
DLH	—	0	
XCP	194	18	
NDJ	75	9	
FXB	28	3	
JCN	180	9	
BGQ	167	18	
KMC	20	3	
RXT	188	18	
KFN	82	9	
MBW	47	3	
TDH	141	9	
LRP	51	3	
ZWS	117	18	
PHQ	89	9	
XGD	158	18	
CZQ	91	3	

Number correct
0″ Delay ________
3″ Delay ________
9″ Delay ________
18″ Delay ________
Total

*Courtesy of Edith Kaplan.

This technique has also been administered with visual stimuli. In the sequential format the consonants are printed individually on cards and shown one at a time in sets of three; in the simultaneous format all three consonants appear together on each stimulus card (Edith Kaplan, personal communication [mdl]).

Test characteristics. Differences in sex, age—from late teens up to 69 years—or education levels (high school completion or less vs. more than high school) were not statistically significant (Stuss, Stethem, and Poirier, 1987). Nevertheless, women showed a tendency for better recall than men, persons with more than a high school education had slightly higher scores on average, and older subject groups did a little less well than younger ones. Education effects penalized those with fewer years of schooling regardless of age (Bherer et al., 2001). Small but significant practice effects do occur (Stuss, Stethem, and Poirier, 1987). Stuss noted that,

> The Brown-Peterson is an interesting test. My concern is that it is very sensitive but not necessarily very specific. I think it is a wonderful working memory test, and one has to be very careful about interpreting [failure] as "poor test motivation." However, the test is very multifactorial and people can fail for various reaons. I am therefore cautious in its interpretation. (personal communication, Dec., 1996 [mdl]).

Neuropsychological findings. The Brown-Peterson technique is useful for documenting very short-term memory deficits (i.e., rapid decay of memory trace) in a variety of conditions. One of the early uses of the Brown-Peterson task in a patient population was Baddeley and Warrington's (1970) study of amnesic patients in which they reported no difference between Korsakoff patients and controls. Subsequent investigators, however, found severe impairments in Korsakoff patients (N. Butters and Cermak, 1980). Leng and Parkin (1989) noted that the performance deficits of Korsakoff patients were associated with their frontal lobe dysfunction rather than the severity of their memory problems, and also that patients with temporal pathology did better than those with Korsakoff's syndrome. Further implicating the sensitivity of this technique to frontal lobe dysfunction is Kapur's (1988b) finding that patients with bifrontal tumor, but not those with a tumor in the region of the third ventricle, recalled significantly fewer items than control subjects. B. Milner (1970, 1972) found that patients with right temporal lobectomies performed as well as normal controls on this test, but the amount recalled by those with left temporal excisions diminished as the amount of hippocampus loss increased. Again temporal lobe epilepsy patients with a left hemisphere focus performed less well than patients with a right hemisphere focus on a task recalling a single word after interference, but in this study both patient groups scored lower than controls (Giovagnoli and Avanzini, 1996). However, a visual presentation of word triads resulted in equally impaired recall by right and left temporal lobe seizure patients (Delaney, Prevey, and Mattson, 1982).

Data on multiple sclerosis patients are mixed: in one set of studies, they tended to differ very little from control subjects (Rao, Leo, Bernardin, and Unverzagt, 1991; Rao, Leo, and St. Aubin-Faubert, 1989); in others, MS patients exhibited deficits (I. Grant, McDonald, Trimble, et al., 1984; Grigsby, Ayarbe, et al., 1994). The distraction effect is much greater for Alzheimer patients than for normal subjects in their age range (E.V. Sullivan, Corkin, and Growdon, 1986), and it distinguishes both Huntington (N. Butters, Sax et al., 1978; D.C. Myers, 1983) and Parkinson patients (Graceffa et al., 1999) from controls. Schizophrenics show a rapid decline in recall on this task and produce an unusual number of intrusion errors (Fleming et al., 1995). Stuss, Ely, and their colleagues (1985) report that this test was the most sensitive to mild head injury in a battery of commonly used tests.

Occasionally consonant trigrams offers bonus information about a patient's susceptibility to attentional disorders. When counting backwards, the patient may skip or repeat a decade, or drop numbers out of sequence without being aware of the error(s). This occurrence suggests mental tracking and/or self-monitoring problems which should be further explored.

> A 30-year-old native English speaker of Polynesian stock incurred an episode of cerebral hypoxia during a surgical procedure. She had dropped out of high school to work as a cashier in a fast food outlet. In the neuropsychological examination she obtained only *low average* to *average* scores on verbal skill and academic tests—excepting for a *high average* verbal fluency production. Yet on tests of visuoperception and construction she achieved scores in the *high average* and even *superior* (Block Design SS = 14) ability ranges and performed *within normal limits* on both the Category Test and Raven's Matrices. Chief complaints (of her family) involved executive disorders: passivity, anergia, impaired organizing ability, and disinhibited shopping. Together these problems have rendered her socially dependent.
>
> On Consonant Trigrams this cooperative patient recalled 9/15 letters after the 3 sec delay trials, 5/15 after the 9 sec delay, and 2/15 after 18 sec, demonstrating a significant working memory problem. In addition she had difficulty keeping track of what she was doing when counting backwards: of the 15 items, she made no errors on only six; on others she skipped decades ("51-40-49-48 . . . ok"), she counted forward ("82-83-84 . . . I'm going upwards") but was usually not aware of errors, and she tended to skip numbers ("81-79-78 . . . " "156-154-153 . . ."), thus also displaying a severe mental tracking disability made worse by defective self-monitoring.

Variants of the Brown-Peterson technique. This paradigm has been adapted to specific research or clinical questions in a number of ways. The mode of presentation may be written—usually the stimuli are presented on cards—as well as oral. The stimuli may be words instead of consonants, and the number of stimuli—whether words or consonants—may be as few as 1 (e.g., see Leng and Parkin, 1989; E.V. Sullivan, Corkin, and Growdon, 1986). While the distracting subtraction task is usually by 3's, some studies have used 2's. Of three different distracting conditions in one study, two called for subtraction (by 2's, by 7's), and one simply required rapid repetition of "the" during the different time intervals (Kopelman and Corn, 1988). "The" repetition produced minimum interference compared with subtraction distractors, while subtraction by 2s or by 7s was equally effective. In another study, of three conditions using 10, 20, and 30 sec intervals between presentation of three consonants and request for response, one "distractor" involved repeating the syllable "bla," one required simple addition, with no distractor in the third; recall was almost perfect with no distractor and a little less than perfect in the "bla bla" condition, but dropped significantly with addition—particularly for subjects with ≤12 years' education (Bherer et al., 2001). In yet another variant, subjects had to recall eight triads of women's given names after counting backwards for 20 sec (Kapur, 1988b). Using three stimuli at a time—whether words or consonants—and subtraction by 3s for the usual duration ranges resulted in similar findings across studies (D.C. Myers, 1983), suggesting that the paradigm is more important than the contents in eliciting the Brown-Peterson phenomenon. Using recall of three monosyllabic words, Eustache and his colleagues (1995) observed an age effect for subjects ranging from 20 to 69 years.

Increasing the Tested Span

Many elderly subjects and patients with brain disorders have an immediate memory span as long as that of younger, intact adults. Thus, simple span tests, as traditionally administered, frequently do not elicit the immediate recall deficits of many persons with reduced memory capacity. To enhance sensitivity to these problems, longer and more complex span formats have been devised.

Supraspan

A variety of techniques for examining recall of strings of eight or more random numbers have demonstrated the sensitivity of the supraspan task to age, educational level, brain impairment, and anticholinergic medication (Crook, Ferris, et al., 1980; H.S. Levin, 1986). When given strings of numbers or lists to learn that are longer than normal span (i.e., span under *stimulus overload* conditions), the excess items serve as interference stimuli so that what is immediately recalled upon hearing the list represents partly what span can grasp, and partly what is retained (learned) despite interference. Problems in identifying and scoring the supraspan (e.g., it may begin at 7, 8, 9, or 10 in the normal population; should partial spans be counted?) resulted in some complex scoring systems that are unsuited to clinical use.

In normal subjects, supraspan recall will be at or a little below the level of simple span (e.g., see S.M. Rao, Leo, and St. Aubin-Faubert, 1989; pp. 421, 426 for initial recall of word list learning tasks) but will be two or more items shorter than simple span in many brain disorders. Digit span—forward or reversed—did not discriminate multiple sclerosis patients from normal subjects in the Rao, Leo, and St. Aubin-Faubert study, yet when given just one digit more than their maximum forward span, patients averaged two and one-half recalled digits fewer than the controls (2.95 vs. 5.46, respectively). Patients with right temporal lobe resections had impaired performances on a verbal supraspan learning task despite achieving intact verbal memory scores on the Wechsler Memory Scale (WMS) (Rausch and Ary, 1990).

Telephone Test (Crook, Ferris, et al., 1980; Zappalá et al., 1989)

To make the span test practically meaningful, 7- or 10-digit strings have been presented in a visual format, as if they were telephone numbers to be recalled. It is interesting to note that the longer the string, the shorter the amount of recall (see Table 11.2).

Serial Digit Learning (or Digit Sequence Learning) (Benton, Sivan, Hamsher, et al., 1994)

Subjects with less than a twelfth grade education hear a string of eight digits to learn (form D8); subjects with 12 or more years of schooling learn a nine-digit span (form K9). The digit string is repeated either until the subject has recalled it correctly for two consecutive trials or through all 12 trials. The maximum score of 24 is based on a scoring system in which each correct trial earns two points, one omission or misplacement drops

TABLE 11.2 Telephone Test Scores for Two Age Groups

	AGE (YEARS)	
Digit String	*≤50*	*>50*
Seven	5.93 ± .20	4.79 ± .29
Ten	4.24 ± .28	2.93 ± .26

From Zappalá et al. (1989)

the score to 1 point, and 2 points are added for each trial to 12 that did not have to be given. Defective performance is defined by a score of 7 or less for high school graduates (form K9), and 6 points or less for those at lower education levels. Age becomes a relevant variable after 65 years, which makes this test more sensitive to the mental changes of aging than simple digit span (Benton, Eslinger, and Damasio, 1981). Education contributes positively to performance on this test, but sex does not affect recall efficiency (Benton, Sivan, Hamsher, et al., 1994). Factor analysis suggests that performance is more closely a function of attention and information processing than learning (Larrabee and Curtiss, 1995).

Neuropsychological findings. Although intragroup variability for right and for left temporal lobe seizure patients was so great that the difference between their respective mean scores of 12.7 ± 7.2 and 8.3 ± 8.5 did not reach significance (Loring, Lee, Martin, and Meador, 1988), a χ^2 comparison of the number of failures in each group was significant ($p < .045$; see Lezak and Gray, 1991). However, even the large intragroup variability did not obscure pre–post left temporal lobectomy changes as documented on this test, with this group's average score dropping from an initial 13 to 5 after surgery (G.P. Lee, Loring, and Thompson, 1989). Patients with right temporal lobectomies showed, on the average, only a 2-point drop from their presurgery scores. This test is sensitive to more than verbal memory deficits, as patients with bilateral damage tend to perform less well on it than those with strictly lateralized dysfunction (Benton, Eslinger, and Damasio, 1981; Benton, Sivan, Hamsher, et al., 1994). Patients with lead toxicity also perform below expectation on this test (Stewart et al., 1999).

Tombaugh and Schmidt (1992) present a similar 12-trial format but use a sequence two digits longer than the subject's longest span and require three correct trials before discontinuing early. The rationale for this procedure is that adjusting the supraspan length on the basis of each individual's forward digit span equates the level of difficulty for everyone. They include a delayed recall trial with as many as six additional learning trials should the initial delayed recall be failed. Normative data for adults 20–79 years show a significant age effect. Scores of 70- to 80-year-old persons run 25% lower than scores for normal subjects under 40 (Tombaugh, Grandmaison, and Schmidt, 1995).

Words

The use of words, whether singly in word lists or combined into phrases, sentences, or lengthier passages, introduces a number of dimensions into the memory task that can affect test performances differentially, depending upon the patient's age, nature of impairment, mental capacity, etc. These dimensions include familiar–unfamiliar, concrete–abstract, low–high imagery, low–high association level, ease of categorization, low–high emotional charge, and structural dimensions such as rhyming or other phonetically similar qualities (e.g., see Baddeley, 1976; Mayes, 1988). The amount of organization inherent in the material also affects ease of retention. This is obvious to anyone who has found it easier to learn words than nonsense syllables or sentences than word strings. When using words for testing memory—and particularly when making up alternate word lists, sentences, etc.—the examiner must be alert to the potential effects that these dimensions can have on the comparability of items, for instance, or when interpreting differences between groups on the same task.

When developing material for testing memory and learning functions, the examiner may find Toglia and Battig's *Handbook of Semantic Word Norms* (1978) a useful reference. These authors give ratings for 2,854 English words (and some "nonwords") along the seven dimensions of concreteness, imagery, categorizability, meaningfulness, familiarity, number of attributes or features, and pleasantness, thus enabling the examiner to develop equatable or deliberately biased word lists on a rational, tested basis. A "meaningfulness" list of 319 five-letter (alternating consonant with vowel, e.g., "vapor," "money," "sinew") words and word-like constructs (i.e., *paralogs*) was developed by Locascio and Ley (1972). Paivio and his colleagues (1968) graded 925 nouns for concreteness, imagery, and meaningfulness. Palermo and Jenkins's *Word Association Norms* (1964) provides a great deal of data on word frequencies and their relatedness. An exhaustive reference for frequency of 86,741 English words is available (J.B. Carroll, Davies, and Richman, 1971).

Brief word learning tests

When circumstances require that memory be assessed quickly, such as at the hospital bedside, a short word learning task provides useful information. Probably the word-learning test familier to most clinicians comes from the mental status examination used by medical practitioners, especially psychiatrists and neurologists, to evaluate their patients' mental conditions. In the course of the evaluation interview the patient is given three or four unrelated, common words (some examiners use a name or date, an address, and a flower name or florist's order, such as "two dozen yellow roses") to repeat, with instructions to remember these items for recall later. The patient must demonstrate accurate immediate repetition of all the words or phrases so that there is no question about their having been registered.

For some patients, this may require several repetitions. Once assured that the patient has registered the words, the examiner continues to question the patient about other issues—work history, family background—or may give other brief items of the examination for approximately 5 min. The patient is then asked to recall the words. The widely used Mini-Mental State Examination (MMSE) tests memory with recall of three words after a few minutes with an intervening task (M.F. Folstein et al., 1975; see pp. 706–709). Strub and Black (2000) give *Four Unrelated Words* with recall after delays of 5, 10, and 30 mins and provide norms for five decades from the 40s to 80s.

Most persons under age 60 have no difficulty recalling all three or four words or phrases after 5 or 10 mins (Strub and Black, 2000). Thus, correct recall of two out of three or even three out of four raises the question of a retention deficit in middle-aged and younger persons (Beardsall and Huppert, 1991). Most data sugggest that approximately 50% of adults, including those over 85 years, can recall all three words and another 30%–40% can recall two of the words (Bleeker, Bolla-Wilson, Kawas, and Agnew, 1988; Heeren, Lagaay, von Beek et al., 1990). In another study approximately 25% of healthy adults aged 50 and older (up to 95) recalled all three words and 40% recalled two of the three words (Cullum, Thompson, and Smernoff, 1993). All studies agree that recall of only one of three words at any age usually indicates that verbal learning is impaired. Using a cut-off of less than two words, this memory test had an 82% accuracy rate in distinguishing patients with mild dementia from controls (Derrer et al., 2001). Stuss, Binns, and their collaborators (2000) report that when TBI patients could recall "three little words" predicted return of continuous memory.

A three-item stimulus (table, red, 23 Broadway) has been used as a brief memory assessment with 2,000 presumably intact persons in the 50 to 93 age range (Jenkyn et al., 1985). With a criterion of failure of one or more errors, failed performances were made by 14% to 19% of subjects between the ages of 50 and 64, by 24% to 28% of those in the age range 65 and 79, and by 55% of persons over age 80. Unfortunately, how many persons made only one error is not reported. Yet it is evident that the one error = failure criterion may be too stringent for older persons. Again using only three words, 28 healthy subjects 75 years and older recalled an average of 2.0 ± 1.0 words (Beardsall and Huppert, 1991). The relative crudity of the three-word format has led some clinicians to favor one with four words (Petersen, 1991; Strub and Black, 2000). Godwin-Austen and Bendall (1990) use a 7-word address.

Among the many variants to the basic three- or four-word format is one in which the examiner identifies their categories when naming the words (e.g., "Detroit, a city; yellow, a color; lily, a flower; apple, a fruit."). On the first recall trial, the examiner asks for the words. Should the patient omit any, the examiner can then see whether cuing by category will aid the patient's recall. When cuing improves recall, a retrieval rather than a storage problem is implicated. Upon satisfying himself that the patient could recall several words after a short time span, Luria (1973b) used two three-word lists, giving the patient "Series 2" after the patient had learned the three common words in "Series 1." When the second series had been learned, Luria then asked for recall of Series 1 as a test of the patient's capacity to maintain the organization and time relationships of subsets of learned material.

C. Ryan and Butters (1982) used four words in a version of the Brown-Peterson technique. Following the one-per-second reading of four unrelated words (e.g., anchor, cherry, jacket, and pond), patients were given a three-digit number with instructions to count backward from that number by threes for 15 or 30 seconds, at which time they were instructed to recall the words. This technique was quite sensitive in eliciting an age gradient for normal subjects that was paralleled, at significantly lower levels, by alcoholics at the three tested age levels. Since the four (or three) words differ on each subsequent trial, this format is effective in bringing out perseverative tendencies (N. Butters, 1985).

In still another variation, two three-word sets (rose, ball, key; brown, tulip, honesty) were given for immediate recall, each word set administered at a different time during the examination (Cullum, Thompson, and Smernoff, 1993). Subjects were not told that they would be asked to recall the words later. Although the words are repeated up to three times as needed to retain for immediate recall, and the delayed recall trial comes only two to three minutes later, fewer than 30% of subjects age 50+ recalled all three of the first set of words, while one-third of them recalled one or none; only 10% recalled all three of the second set, with 60% recalling none. The expected age gradient appeared, with subjects in the 80- to 95-year age range remembering the fewest words; education was not associated with recall prowess. As the lower word frequency for "tulip" and "honesty" may account for the very great differences in rate of recall between the two word lists, the authors caution that word difficulty be considered when giving word learning tests.

Strub and Black (2000) use four words that can be cued in several ways and ask for recall at 10 and 30 min as well as 5 min. Should any words be missed on spontaneous recall, the examiner provides different cues, such as the initial phoneme of the abstract word, the category of the color, a familiar characteristic of the flower, etc. When cueing fails, they recommend a recognition technique (e.g., "Was the flower a rose, tulip, daisy, or petunia?") to help determine whether the patient's problem is one of storage or retrieval. The additional 10 and 30 sec recalls elicited a rebound effect in which recall improved with delay for each of their five age groups (e.g., recall at 5 and 10 secs for subjects in their 60s was 2.0 and 3.0 words, respectively; for the 80s it was 2.1 and 2.7 words); 30 sec recalls for all but the 40s group were even higher than 10 sec recall (e.g., 3.5 for the 60s group). Moreover, both stage I and II Alzheimer patients showed the rebound effect at 10 sec with a slight drop at 30 sec that was still higher than the 5 sec recall (e.g., stage I: 1.6, 1.9, 1.8 at 5, 10, and 30 secs).

Benson Bedside Memory Test (D. Frank Benson, personal communication [dbh])

Frank Benson used eight words in an informal examination of memory (see Table 11.3). The eight words are read to the patient and with recall after each of four trials. Free recall is obtained after a 5 to 10 min delay followed by a category-cued recall for any omissions, followed by multiple choice prompting if necessary. Although this task takes only minutes it is sensitive to delayed recall impairment. Most adults can acquire 7 or 8 of the words during the four presentations and should be able to recall approximately 6 freely and the remainder with cues.

TABLE 11.3 Benson Bedside Memory Test

Words	*Category Cue*
Cabbage	Vegetable
Table	Furniture
Dog	Animal
Baseball	Sport
Chevrolet	Automobile make
Rose	Flower
Belt	Article of clothing
Blue	Color

Word Span and Supraspan

The number of words normal subjects recall immediately remains relatively stable through the early and middle adult years. Five age groups (20s to 60s) comprising a total of 200 men, were tested with familiar one-syllable words in lists ranging in length from four to 13 words (Talland, 1965b). Beyond five-word lists, average recall scores hovered around 5.0. The five age groups did not differ on recall of lists of four to seven words. The two oldest groups showed a very slight but statistically significant tendency to do a little less well than the youngest groups on the 9- and 11-word lists, and the three oldest groups did less well on the 13-word list. The greatest difference between the oldest and youngest groups was on the 9-word list on which the 20–29 age group averaged 5.6 words and the 60–69 age group averaged 5.0 words. A significant drop with aging in number of words recalled was also documented by Delbecq-Derouesné and Beauvois (1989). Recall data from 12 lists of 15 words each for five age ranges (20–25 to > 65) indicated that subjects in the 55–65 age range and older retrieved significantly fewer words, recalling many fewer from the beginning and middle of the lists than did three younger age groups. When tested first with a two-word list and adding a word with each successful repetition while maintaining the original word order, the word span of a group of control subjects again averaged 5.0 (E. Miller, 1973). Control subjects learned word lists of one, two, and three words longer than their word span in two, four, and more than ten trials, respectively.

Word list learning tests provide a ready-made opportunity to examine supraspan. Rather than use random words, some examiners test supraspan with shopping lists to enhance the task's appearance of practical relevance (Delis, Kramer, Kaplan, and Ober, 1987, 2000; Flicker, Ferris, and Reisberg, 1991; Teng, Wimer, et al., 1989). Of those tests with unrelated words in most common use, the Selective Reminding procedure usually presents a 12-word list and the Auditory–Verbal Learning Test (AVLT) list contains 15 words. Based on tests from 301 adults, Trahan, Goethe, and Larrabee (1989) found that on first hearing a 12-word list, the average recall of younger adults (18–41) was approximately six, recall dropped to an average of five words for persons age 54–65, persons 66–77 years old recalled between four and five words, and the average for a 78+ group was four words. On the basis of these data, Trahan and his colleagues recommended that recall of fewer than four words be considered impaired up to age 54; and that for ages 54 and older, the impaired classification begin with recalls of two or less. Slightly higher spans have been reported for Trial I of the AVLT in samples of healthy, well-educated subjects (Ivnik, Malec, Smith, et al., 1992a). M. Schmidt (1996) computed metanorms for nine adult age groups divided by sex.

On supraspan learning tasks, it appears that both short-term retention and learning capacities of intact subjects are engaged (S.C. Brown and Craik, 2000; see also Vallar and Papagno, 2000, for a discussion of the many systems contributing to supraspan recall). Many brain impaired patients do as well as normal subjects on the initial trial but have less learned carry-over on subsequent trials (e.g., Lezak, 1979). Short-term retention in patients whose learning ability is defective also shows up in a far better recall of the words at the end of the list than those at the beginning (the *recency effect*), as the presentation of new words in excess of the patient's immediate memory span interferes with retention of the words first heard. Normal subjects, on the other hand, tend to show a *primacy* as well as a recency effect, consistently having better recall for the words at the beginning of the list than for most of the other words. Moreover, when the full list is repeated for each learning trial, subjects whose memory system is intact are much more likely to develop an orderly recall pattern that does not vary much from trial to trial except as new words are added. By trial IV or V, many subjects with good learning capacity repeat the list in almost the same order as it is given.

On word list tests in which unrelated words are presented in the same order on each learning trial, the subject's learning strategy can be examined for efficiency. On the initial hearing most normal individuals show a primacy and recency effect in which they recall most easily the first and last few words from the list but they tend to switch strategies after the second or later trials and begin their recall with the words they had not yet said, thereby minimizing proactive interference effects. In addition to these strategies, many subjects make semantic associations between the words and recall subgroups of words in the same order from trial to trial (e.g., on the AVLT, school-bell; on the California Verbal Learning Test [CVLT], recognition of one or more of the predefined categories). A review of the order in which patients recall words over the five trials will show whether they are following this normal pattern. Patients who fail to show this or any other pattern may have approached the task passively, may be unable to develop a strategy, or may not appreciate that a strategy is possible. Asking the patient at the conclusion whether any particular technique was used for learning the words often clarifies whether strategies were developed intentionally.

Impairment in the ability to put time tags on learned material is assessed by the subject's accuracy in distinguishing each of the two lists on the short-term and delayed recall trials and on the recognition trial of the AVLT or CVLT, or by the presence of intrusions from previously administered tests. For example, items from the Boston Naming Test when seen earlier might appear among the recalled words. For confused patients, even words from the instructions, such as "remember," may be produced. Intrusions of words not in any test shows a tendency for interference from internal associations and, sometimes, disinhibition.

A few times in the course of the series of learning trials most persons will repeat on the same trial a word already given on that trial. This kind of repetition is not "perseveration." Rather, most patients who repeat an abnormal number of words on word list learning tests have attentional problems such that they have difficulty keeping track of what they have already said while searching their memory for other words; in short, they cannot do two things at once: monitor their performances and engage in a memory search. *Perseveration* refers to mental stickiness or "stuck in set" phenomena that are more likely to occur with specific patterns of cognitive dysfunction such as those associated with significant frontal lobe damage, some aphasic disorders, etc. *Repetition* must not be confused with *perseveration*.

Auditory-Verbal Learning Test (AVLT) (A. Rey, 1964; M. Schmidt, 1996)

In 1916 Edouard Claparède developed a one-trial word list learning test composed of 15 words which were later used by André Rey to form the AVLT (Boake, 2000). This easily administered test affords an analysis of learning and retention using a five-trial presentation of a 15-word list (list A), a single presentation of an interference list (List B), two postinterference recall trials—one immediate, one delayed—and recognition of the target words presented with distractors. By this means the examiner easily obtains measures that are crucial for understanding the kind and severity of a patient's memory deficits: immediate word span under overload conditions (trial I), final acquisition level (trial V), total acquisition (ΣI–V), amount learned in five trials (trial V − trial I), proactive interference (trial I − trial B), retroactive interference (trial V − trial VI), delayed recall (trial VII), recognition, number of repetitions, and number and types of intrusions. Retention should be examined after an extended delay, from 20 to 45 minutes—most usually, around 30. In some instances the examiner may wish to determine retention after longer periods, such as one hour or the next day. The original French words and their order were translated without change to English. Other language versions include Flemish (Lannoo and Vingerhoets, 1997), German (H. Mueller et al., 1997), Hebrew (Vakil and Blachstein, 1993), and Spanish (Miranda and Valencia, 1997).

For trial I, the examiner reads a list (A^1) of 15 words (see Table 11.4) at the rate of one per second after giving the following instructions:

TABLE 11.4 Rey Auditory-Verbal Learning Test Word Lists

List A[1]	*B*[1]	*AC*[2]	*BC*[2]	*A/JG*[3]	*B/JG*[3]	C
Drum	Desk	Doll	Dish	Violin	Orange	Book
Curtain	Ranger	Mirror	Jester	Tree	Armchair	Flower
Bell	Bird	Nail	Hill	Scarf	Toad	Train
Coffee	Shoe	Sailor	Coat	Ham	Cork	Rug
School	Stove	Heart	Tool	Suitcase	Bus	Meadow
Parent	Mountain	Desert	Forest	Cousin	Chin	Harp
Moon	Glasses	Face	Water	Earth	Beach	Salt
Garden	Towel	Letter	Ladder	Knife	Soap	Finger
Hat	Cloud	Bed	Girl	Stair	Hotel	Apple
Farmer	Boat	Machine	Foot	Dog	Donkey	Chimney
Nose	Lamb	Milk	Shield	Banana	Spider	Button
Turkey	Gun	Helmet	Pie	Radio	Bathroom	Log
Color	Pencil	Music	Insect	Hunter	Casserole	Key
House	Church	Horse	Ball	Bucket	Soldier	Rattle
River	Fish	Road	Car	Field	Lock	Gold

[1]Taken from E.M. Taylor, *Psychological Appraisal of Children with Cerebral Defects.* © 1959 by Harvard University Press.
[2]Developed by Crawford, Stewart, and Moore (1989). © Swets and Zeitlinger.
[3]Developed by Jones-Gotman, Sziklas, and Majdan (personal communication, March 1993).

I am going to read a list of words. Listen carefully, for when I stop you are to tell me as many words as you can remember. It doesn't matter in what order you repeat them.

On first hearing the long list some patients may be distracted by fear of failure, so it is desirable to include in the instruction:

There are so many words that you won't remember them all the first time. Just try to remember as many as you can.

The examiner writes down the words recalled in the order in which they are recalled, thus keeping track of the pattern of recall, noting whether the patient has associated two or three words, proceeded in an orderly manner, or demonstrated hit-or-miss recall. *Examiners should not confine themselves to a structured response form* but rather take down responses on a sheet of paper large enough to allow for many repetitions and intrusions as well as for high-level—and therefore very wordy—performances. Use of record sheets in which words from the list are checked or numbered in order of recall from trial to trial delays the inexperienced examiner as some patients recall the words so fast that finding the words to check is difficult. Moreover, preformed record sheets do not allow the examiner to keep track of where intrusions or repetitions occur in the course of the subject's verbalizations on any one trial. It is usually possible to keep up with fast responders by simply recording the word's initial first two or three letters when more than one word on the list begins with the same letter (e.g., CUrtain, COFfee, COLor). Should patients ask whether they have already said a word, the examiner informs them, but does not volunteer that a word has been repeated as this tends to distract some patients and interfere with their performance. It also may alert some patients to monitor their responses—a good idea that may not have occurred to them without external advice.

When patients indicate that they can recall no more words, the examiner rereads the list, following a second set of instructions:

Now I'm going to read the same list again, and once again when I stop I want you to tell me as many words as you can remember, *including words you said the first time.* It doesn't matter in what order you say them. Just say as many words as you can remember, whether or not you said them before.

This set of instructions must emphasize inclusion of previously said words, for otherwise some patients will assume it is an elimination test.

The list is reread for trials III, IV, and V, using trial II instructions each time. The examiner may praise patients as they recall more words. Patients may be told the number of words recalled, particularly if they are able to use the information for reassurance or as a challenge. On completion of each ten-word trial of a similar word learning test, Luria (1966) asked his patients to estimate how many words they would recall on the next trial. In this way, along with measuring verbal learning, one can also obtain information about the accuracy of patients' self-perceptions, appropriateness of their goal setting, and their ability to apply data about themselves. This added procedure requires very little time or effort for the amount of information it may afford, and it does not seem to interfere with the learning or recall process. On completion of trial V, the examiner tells the patient:

Now I'm going to read a second list of words. This time, again, you are to tell me as many words of this second list as you can remember. Again, the order in which you say the words does not matter. Just try to tell me as many words as you can.

The examiner then reads the second word list (B^1), and writes down the words in the order in which the patient says them. Following the B-list trial, the examiner asks the patient to recall as many words from the first list as possible (trial VI). Also without forewarning, the 20- to 45-minute delayed recall trial (VII) measures how well the patient recalls what was once learned. Normally few if any, words recalled on trial VI are lost after this short a delay (e.g., Mitrushina, Boone, and D'Elia, 2000; M. Schmidt, 1996).

The score for each trial is the number of words correctly recalled. A total score, the sum of trials I through V, can also be calculated. Words that are repeated can be marked *R*; *RC* when patients repeat themselves and then self-correct; or *RQ* if they question whether they have repeated themselves but remain unsure. Subjects who want to make sure they did not omit saying a word they remembered may repeat a few words after recalling a suitable number for that trial. However, lengthy repetitions, particularly when the subject can recall relatively few words, most likely reflect a problem in self-monitoring and tracking, along with a learning defect.

Words offered that are not on the list are errors and marked *E*. Frequently an error made early in the test will reappear on subsequent trials, often in the same position relative to one or several other words. Intrusions from list A into the recall of list B or from B into recall trial VI are errors that can be marked *A* or *B*.

See Table 11.5 for an example of scored errors. The 28-year-old ranch hand and packer who gave this set of responses had sustained a right frontotemporal contusion requiring surgical reduction of swelling just two years before the examination. Since the accident he had been unable to work because of poor judgment, disorientation, and personality deterioration.

This method of marking errors enables the examiner to evaluate the quality of the performance at a glance. Patients who make intrusion errors tend to have difficulty in maintaining the distinction between information coming from the outside and their own associations; those who give a List A response on Trial B, or a List B response on later trials tend to confuse data obtained at different times. Some, such as the patient whose performance is given in Table 11.5, have difficulty maintaining both kinds of distinctions, which suggests a serious breakdown in self-monitoring functions.

TABLE 11.5 Sample AVLT Record Illustrating Error Scoring

I	*II*	*III*	*IV*	*V*	*B*	*VI*
Hat 1	Drum 1	River 1	Drum 1	River 1	Desk 1	Drum 1
Garden 2	Curtain 2	House 2	Curtain 2	House 2	Ranger 2	Curtain 2
Moon 3	Bell 3	Turkey 3	Hat 3	School 3	Glasses 3	Bell 3
Turkey 4	House 4	Farmer 4	School 4	Bell 4	Bell A	Parent 4
Hose EA	River 5	Water EC	Parent 5	Farmer 5	Pet EA	School 5
	Hose EA	Color 5	Farmer 6	Drum 6	Fish 4	Moon 6
	Drum R	Drum 6	Color 7	Curtain 7	Glasses R	Teacher EA
	Bell R	Curtain 7	Nose 8	Bell R	Mountain 5	Turkey 7
	Curtain R	Garden 8	Turkey 9	School R	Cloud 6	Coffee 8
	Drum R	Hat 9	Color R	Parent 8	Bell AR	Color 9
		Hose EA	School R	Coffee 9		
		Garden R	Nose R	School R		
		Turkey R	Drum R	Parent R		
		Farmer R	Turkey R	Color 10		
		School 10	Farmer R	Moon 11		
		Parent 11				
4	5	11	9	11	6	9
1 EA	1 EA	1 EA	6 R	4 R	2 A (1 R)	1 EA
	4 R	1 EC			1 EC	
		3R			2 R	

A recognition trial should be given to all patients except those who recall 14 or more words on trial VII and have made no errors (confabulations, list confusions, associations, or other intrusions), for the likelihood of recognition errors by these latter subjects is slim. In testing recognition, the examiner asks the patient to identify as many words as possible from the first list when shown (or read if the patient has a vision or literacy problem) a list of 50 words containing all the items from both the A and B lists as well as words that are semantically associated (S) or phonemically similar (P) to words on lists A or B; or the alternate word sets (see Table 11.6). The instruction is given as the patient is handed the recognition sheet and a pencil:

I am going to show you a page with words on it. Circle the words from the first list I read to you. Some of the words you see here are from the first list and some are from the second list I read to you only once. Some of these words were not on either list. Just circle the ones from the first list, the list I read five times.

Some subjects circle relatively few words and need encouraging. It is possible to keep two scores by giving them a different colored pencil after they said they were finished, telling them:

TABLE 11.6 Word Lists for Testing AVLT Recognition, Lists A–B

Bell (A)*	Home (SA)	Towel (B)	Boat (B)	Glasses (B)
Window (SA)	Fish (B)	Curtain (A)	Hot (PA)	Stocking (SB)
Hat (A)	Moon (A)	Flower (SA)	Parent (A)	Shoe (B)
Barn (SA)	Tree (PA)	Color (A)	Water (SA)	Teacher (SA)
Ranger (B)	Balloon (PA)	Desk (B)	Farmer (A)	Stove (B)
Nose (A)	Bird (B)	Gun (B)	Rose (SPA)	Nest (SPB)
Weather (SB)	Mountain (B)	Crayon (SA)	Cloud (B)	Children (SA)
School (A)	Coffee (A)	Church (B)	House (A)	Drum (A)
Hand (PA)	Mouse (PA)	Turkey (A)	Stranger (PB)	Toffee (PA)
Pencil (B)	River (A)	Fountain (PB)	Garden(A)	Lamb (B)
		Recognition Lists AC-BC[1]		
Nail (A)	Envelope (SA)	Ladder (B)	Foot (B)	Water (B)
Sand (SA)	Car (B)	Mirror (A)	Bread (PA)	Joker (SB)
Bed (A)	Face (A)	Screw (SA)	Desert (A)	Coat (B)
Pony (SA)	Toad (PA)	Music (A)	Street (SA)	Captain (SA)
Jester (B)	Silk (PA)	Dish (B)	Machine (A)	Tool (B)
Milk (A)	Hill (B)	Pie (B)	Head (SPA)	Fly (SPB)
Plate (SB)	Forest (B)	Wood (SB)	Girl (B)	Song (SA)
Heart (A)	Sailor (A)	Ball (B)	Horse (A)	Doll (A)
Jail (PA)	Dart (PA)	Helmet (A)	Soot (PB)	Stall (PA)
Insect (B)	Road (A)	Stool (PB)	Letter (A)	Shield (B)
		Recognition Lists A/JB-B/JB		
Rock (PB)	Star (SA)	Soap (B)	Television (SA)	Violin (A)
Corn (PB)	Peel (SA)	Frog (SB)	Hotel (B)	Beach (B)
Pear (SA)	Lock (B)	Dog (A)	Piano (SA)	Radio (A)
Tree (A)	Banana (A)	Orange (B)	Spider (B)	Bus (B)
Cork (B)	Toad (B)	Cousin (A)	Bucket (A)	Doctor
Bread	Uncle (SA)	Bathroom (B)	Soldier (B)	Chest
Sofa (SB)	Earth (A)	Gloves (SA)	Scarf (A)	Knife (A)
Stair (A)	Hospital (SB)	Field (A)	Wife (SA)	Donkey (B)
Ham (A)	Grass (SA)	Armchair (B)	Train (SB)	Hunter (A)
Casserole (B)	Lunchbox (SA)	Blanket (PA)	Suitcase (A)	Chin (B)

*(A) Words from List A; (B) words from list B; (S) word with a semantic association to a word on list A or B as indicated; (P) word phonemically similar to a word on list A or B, (SP) words both semantically and phonemically similar to a word on the indicated list.

[1]Reprinted with permission (Crawford, Steward, and Moore, 1989).

There were 15 words on that list. See if you can find the rest of them even if you have to guess.

This technique allows the examiner to distinguish between those patients who do not recognize the additional words and make many errors from those who are overly cautious and use a high confidence threshold in their responding.

Others—often patients whose judgment appears to be compromised in other ways as well—check 20 or even 25 of the words, indicating that they neither appreciated the list's length nor maintained discrimination between list A, list B, and various kinds of associations to the target words. These patients can be instructed that the list contained only 15 words and asked to review the recognition sheet, marking with an X only those they are sure were on the list. Without this procedure the accuracy of their recall and ability to sort out what comes to mind cannot be ascertained.

The recognition procedure measures how much was learned, regardless of the efficiency of spontaneous retrieval. Comparison of the recognition and delayed recall scores provides a measure of the efficiency of spontaneous retrieval. Recognition scores below 13 are relatively rare among intact persons under age 59 (Mitrushina, Boone, and D'Elia, 2000; M. Schmidt, 1996), and scores under 12 are infrequent among 55- to 69-year-olds (Ivnik, Malec, Smith, et al., 1992a; Mitrushina, Boone, and D'Elia, 1999). Further, the recognition score examines the patient's capacity to discriminate when or with what other information a datum was learned. This technique may elicit evidence of disordered recall like that which troubles patients with impaired frontal lobe functions who can learn readily enough but cannot keep track of what they have learned or make order out of it. If the patient's problem is simply difficulty in retaining new information, then recognition will be little better than recall on trial VII.

The third word list (C) is available should either the A- or B-list presentations be spoiled by interruptions, improper administration, or confusion or premature response on the patient's part. List C is really an emergency list as words from it are not represented on the AB recognition sheet, thus reducing the recognition format's sensitivity to intrusion and confusion tendencies. Evidence that list C is easier than list B suggests that scores one point higher might be expected for list C trial B and for list A trials VI and VII when using list C as a distractor (Fuller et al., 1997). However, list C was found to be comparable to list A, with individual measures correlating in the .60 to .77 range, and all but three mean differences (favoring list A and appearing on trials IV, V, and VI) no greater than one word (J.J. Ryan, Geisser, et al., 1986). When list C was compared with list A as an alternate learning list in a large study of young gay and bisexual men who were HIV seronegative, it was mostly equivalent although it was slightly more difficult to learn (Uchiyama et al., 1995). Another study found essentially no difference between lists A and C for trials I, III, V, VI, VII, and the recognition trial (Delaney, Prevey, et al., 1988).

As is typical of memory tests practice effects can be pronounced (see pp. 116–117; McCaffrey, Duff, and Westervelt, 2000b). For example, significant improvement on almost all measures appeared on retesting after almost one month, with many increases exceeding one word and an almost three-word difference appearing on trial I (Crawford, Stewart, and Moore, 1989). Thus, the same lists should not be given twice in succession. Ideally, the examiner will have alternate lists with the recognition trial sheet available. Alternate forms are *parallel forms* if they produce results equivalent to the original versions. Crawford, Stewart, and Moore (1989) and Majdan and her colleagues (1996) have developed parallel lists (see Table 11.4) with appropriate sets of words for recognition testing (see Table 11.6). M. Schmidt (1996) provides other parallel forms in English and in German plus three of the four lists Rey (1964) said he had "borrowed [*empruteés*]" from Claparède.

However, it is not always possible to know in advance that the patient had been given the AVLT in a recent examination by someone else. When the parallel list material is not at hand for a second examination, the examiner can reverse the A and B lists, giving the B list five times and using the A list as interference. This manipulation reduces practice effects for all trials except the interference trial, as some patients will show remarkably good recall of the A list even after a year or more.

Normative data. Most young adults (ages 20–39) recall six or seven words on trial I and achieve 12 or 13 words by the fifth trial. The change in number of words recalled from trials I to V shows the rate of learning—the learning curve—or reflects little or no learning if the number of words recalled on later trials is not much more than given on trial I. In general, approximately 1.5 words are lost from trial V to trial VI, i.e., following the interference trial list (B); although after age 64 the spread between trials V and VI gradually increases from almost 2.0 (ages 65–69) to 3 (ages 75–79, 80+) (Sinnett and Holen, 1999). Little if any loss occurs between trials VI and VII, the delayed recall trial. Usually no more than one error shows up on the recognition trial (Mitrushina, Boone, and D'Elia, 1999; M. Schmidt, 1996). Marked variations from this general pattern will likely reflect some dysfunction of the memory system.

Michael Schmidt's (1996) metanorms are reliable for most purposes as several normative studies with relatively large samples have contributed to them. M.E. Harris and his colleagues (2002) have updated their recognition trial accuracy norms. A study of 1,818 homosexual men provides norms for this demographic group as part of a multicenter study of HIV (Uchiyama et al., 1995). M. Schmidt (1996) summarized the AVLT literature through 1995. Extensive reviews of normative data also appear in other sources (Mitrushina, Boone, and D'Elia, 1999; Spreen and Strauss, 1998).

Test characteristics. Word list learning is among the most sensitive verbal memory test formats because of the relative freedom from associative context compared with, for example, prose material. In offering an explanation for the effectiveness of every AVLT learning measure (each trial, Σ trials I–V, *learning* [high score − trial I], in distinguishing normal control subjects from a group of patients with "medically confirmed neuropathologies," J.B. Powell and his colleagues (1991) suggested that these scores "reflect the combined functioning of a wider cross section of neurobehavioral mechanisms, including arousal, motivation, attention/concentration, auditory perception, verbal comprehension, immediate verbal memory span, short-term verbal memory storage and retrieval, and progressive learning abilities" (p. 248). This study found that each of these AVLT scores discriminated between these groups better than each of the Halstead-Reitan measures, the Stroop (Dodrill format), and either Logical Memory or Visual Reproduction (WMS).

It is not surprising that age effects show up on word list tests. Using a Hebrew version of the AVLT, Vakil and Blachstein (1997) found modest changes below the age of 60 compared to increasingly reduced recall after 60. In this study the measures most affected by age were trial V and total acquisition score (ΣI–V), list B, and the first delayed recall (trial VI). Minimal age effects were found for the forgetting rate. This task becomes challenging for persons over 70 years. They typically recall five words on trial I, achieve 10 words by trial V, lose two or three words between trials V and VI, and make two or three errors on the recognition task (M. Schmidt, 1996). Healthy elderly subjects, in comparison with younger ones, show greater forgetting of words at the end of the list during delayed recall (Carlesimo et al., 1997). This suggests that older subjects rely much more on short-lived memory processes—i.e., the immediate recall of the last words on the list—than do younger people.

Sex too plays a role, as women's means on many of the AVLT measures tend to run higher than men's means, from as little as .1 (on a recognition trial) to more than 2.0 words (on recall items) (Bleecker, Bolla-Wilson, Agnew, and Meyers, 1988; Geffen, Moar, et al., 1990). Instances in which men's mean scores are the same or better than women's scores are relatively rare (e.g., R.M. Savage and Gouvier, 1992). Education, verbal facility as measured by vocabulary (WAIS-R), and general mental ability also contribute significantly to performances on this test (Bolla-Wilson and Bleecker, 1986; Ripich et al., 1997; Selnes, Jacobson, et al., 1991; Uchiyama et al., 1995; Wiens et al., 1988).

This test has high test-retest reliability. Using alternate forms with a retest interval of one month, correlations ranged from .61 to .86 for trials I–V and from .51 to .72 for delayed recall and recognition (Delaney, Prevey, Cramer, et al., 1992). Test–retest reliability correlation coefficients after one year ranged from .38 (for trial B) to .70 (for trial V) (W.G. Snow, Tierney, Zorzitto, et al., 1988). Factor analytic studies show that the learning measures of the AVLT (V, VI, recognition) correlate significantly—mostly in the .50 to .65 range—with other learning measures (Macartney-Filgate and Vriezen, 1988; J.J. Ryan, Rosenberg, and Mittenberg, 1984). The supraspan measure, trial I, probably reflects its large attentional component in negligible (.17 to −.13) correlations with the learning measures (Macartney-Filgate and Vriezen, 1988). An evaluation of the comparability of the AVLT with the CVLT produced correlations of .32 for trial I, .33 for trial V, .47 for total words recalled, and .37 for short delay recall (Crossen and Wiens, 1994). A factor analysis of scores made by 146 normal volunteers for Trials I, V, B, VI, VII, Recognition, and a temporal order measure produced three basic factors: retrieval, storage, and acquisition (short-term memory) (Vakil and Blachstein, 1993). The first factor included performance on temporal order and trials VII, B, and V; the second factor included only the Recognition score; and trials I and B entered into the third factor.

Neuropsychological findings. Ordinarily, the immediate memory span for digits and the number of words recalled on trial I will be within one or two points of each other, providing supporting evidence regarding the length of span. Larger differences usually favor the digit span and seem to occur in patients with intact immediate memory and concentration who become confused by too much stimulation (stimulus overload). These patients tend to have difficulty with complexity of any kind, doing better with simplified, highly structured tasks. When the difference favors the more difficult word list retention task, the lower digit span score is usually due to inattentiveness, lack of motivation, or anxiety at the time Digit Span was given.

Slowness in shifting from one task to another can show up in a low score on trial I. When this occurs in a person whose immediate verbal memory span is *within normal limits,* recall B will be two or three words longer than that of trial I, usually *within normal limits.* In these cases trial II recall will show a much greater rate of acquisition than what ordinarily characterizes the performance of persons whose initial recall is abnormally low; occasionally a large jump in score will not take place until trial III. When this phenomenon is suspected, the examiner should review the pattern of the patient's performance on other tests in which slowness in establishing a response set might show up, such as Block Design (e.g., a patient who gets one point at most on each of the first two standard administration designs [i.e., 5 and 6], and does designs 7, 8, and 9 accurately and, often, faster than the first two; or a verbal fluency performance in which the patient's productivity increases with each trial, even though the difficulty of the naming task may also have increased). In those cases in which recall of list B is much lower (by two or three words) than immediate recall on trial I, what was just learned has probably interfered with the acquisition of new material; i.e., there is a *proactive interference* effect. When proactive interference is very pronounced, intrusion words from list A may show up in the list B recall too.

Most patients with brain disorders show a learning curve over the five trials. The appearance of a curve, even at a low level—e.g., from three or four words on trial I to eight or nine on V—demonstrates some ability to learn if some of the gain is maintained on the delayed recall trial, VII. Such patients may be capable of benefiting from psychotherapy or personal counseling and may profit from rehabilitation training and even formal schooling since they can learn, although at a slower rate than normal. Occasionally a once-bright but now severely memory impaired patient will have a large immediate memory span, recalling eight or nine words on trial I, but no more than nine or ten on V and very few on VI. Such a performance demonstrates the necessity of evaluating the scores for each trial in the context of other trials.

This test has proven useful in delineating memory system deficits in a variety of disorders. Some TBI patients will have a reduced recall for each measure but demonstrate a learning curve and some loss on delayed recall but a near normal performance on the recognition trial, indicating a significant verbal retrieval problem (Bigler, Rosa, et al., 1989; Peck and Mitchell, 1990). These patients tend to make a few intrusion errors. The AVLT has been effective in predicting psychosocial outcome after TBI (S.R. Ross et al., 1997).

With localized lesions, the AVLT elicits the expected memory system defects: Frontal lobe patients perform consistently less well than control subjects on recall trials but, given a recognition format for each trial, they show a normal learning curve (Janowsky, Shimamura, Kritchevsky, and Squire, 1989). Patients with left anterior temporal lobectomies show impaired delayed recall of words from the list (Majdan et al., 1996). Degree of left hippocampal atrophy measured by MRI in patients with temporal lobe epilepsy has been associated with severity of total recall and delayed recall deficits on this test (Kilpatrick et al., 1997). Before anterior temporal lobectomy, patients with left temporal lesions differed from those with lesions on the right only in lower scores on recall trials (VI and VII) and recognition; but after surgery they differed greatly on all AVLT measures (Ivnik, Sharbrough, and Laws, 1988). Miceli and his colleagues (1981) also found that patients with right hemisphere lesions did significantly better than those with lesions involving the left hemisphere, even though these latter patients were not aphasic. Korsakoff patients showed minimal improvement on the five learning trials, but when provided a recognition format for each trial they demonstrated learning that progressed much slower than normal and never quite reached the trial V normal level of virtually perfect recognition (Janowsky, Shimamura, Kritchersky, and Squire, 1989; Squire and Shimamura, 1986). These latter authors note that the usual recall format of the AVLT discriminates effectively between different kinds of amnesic patients.

Degenerative diseases have differing AVLT patterns. Low recall on almost all measures except for rate of forgetting, has been reported for multiple sclerosis patients compared to controls (Bravin et al., 2000). Patients with advanced Huntington's disease have, on the average, a greatly reduced immediate recall (fewer than four words), show a small learning increment, and drop down to trial I levels on delayed recall; a recognition format demonstrates somewhat more learning, and they are very susceptible to identifying a word as having been on the learning list when it was not (a false positive error) (N. Butters, Wolfe, et al., 1985, 1986). Patients with early Alzheimer type dementia have a very low recall for trial I and get to about six words by trial V (Bigler, Rosa et al., 1989; Mitrushina, Satz, and Van Gorp, 1989). They have particular difficulty recalling words after a delay with distraction (Woodard, Dunlosky, and Salthouse, 1999). While they recognize about two more words than they can recall, their performances are characterized by many more intrusions than any other diagnostic group (Bigler, Rosa, et al., 1989). Patients with mild cognitive deficits who are at risk for subsequent development of dementia have impaired performance (Petersen, Smith, Waring, et al., 1999; Tierney et al., 1996).

AVLT variants. Patients obviously incapable of learning even 10 of the 15 words experience the standard administration as embarrassing, drudgery, or both. Others may be easily overwhelmed by a lot of stimuli, or too prone to fatigue or restlessness to maintain performance efficiency with a 15-word format. Yet these patients often need a full-scale memory assessment. They can be given only the first 10 words, using the standard procedures. Although a 10-word ceiling is too low for most persons—controls and patients alike—it elicits discriminable performances from patients who, if given 15 words, would simply be unable to perform at their best. Minden, Moes, and their colleagues (1990) used this method to examine multiple sclerosis patients who, by virtue of impaired learning and retrieval functions, easy fatigability, and susceptibility to being overwhelmed and confused due to a reduced processing capacity, may perform better on a 10-word list. The 35 normal control subjects recalled 6 ± 1.4 words on trial I, 9.1 ± 1.2 on trial V, 5.1 ± 1.2 on list B, 7.6 ± 2.3 on trial VI, 7.1 ± 2.9 on trial VII, and then recognized 9.4 ± 1.0 of the words. MS patients were impaired on all measures relative to the controls.

In order to minimize cultural bias in the original AVLT word list for World Health Organization (WHO) research on HIV-1 infection (e.g., there are no turkeys and few curtains in Zaire), two new word lists were constructed from five common categories: body parts, animals, tools, household objects, and vehicles—all presumed "to have universal familiarity" (WHO/UCLA-AVLT) (Maj et al., 1993). List lengths and administration format remain the same. A comparison between subjects in Zaire and Germany indicated low intercultural variability with this new form. When given along with the original word list to persons in a Western country, correlations were in the .47 to .55 range.

Another administration variation ensures that the patient has attended to the words on the list. Using a list of ten words taken from AVLT lists B and C, Knopman and Ryberg (1989) requireed patients to read each word aloud, shown individually on index cards, and follow each word with a sentence they make up using that word. Dementia patients were able to accomplish this task. This was repeated for a second learning trial. Recall followed an interposed task five minutes after the second learning trial. This technique discriminated 55 normal subjects (M recall = 6.0 ± 1.8) from 28 Alzheimer patients (M recall = 0.8 ± 1.0), with no overlaps between the two groups. Correlations with a retest of the normal subjects six months later gave a coefficient of .75.

Vakil, Blachstein, and Hoofien (1991) also use this task to examine incidental recall of temporal order by giving subjects the A list, on which the order differs from the administration sequence, and asking them to rewrite the list in its original form. By giving two sets of administration instructions—one for intentional recall in which subjects are told that they should remember the word order, the other for incidental recall in which the need to remember the word order is not mentioned—Vakil and his colleagues demonstrated that much of temporal order judgment comes automatically. Correlations with other AVLT scores indicate a relationship between the incidental recall of temporal order and retention but not acquisition (Vakil and Blachstein, 1993).

California Verbal Learning Test (CVLT) (Delis, Kramer, Kaplan, and Ober, 1987); *California Verbal Learning Test–Second Edition (CVLT-II)* (Delis, Kramer, Kaplan, and Ober, 2000)

This word list task is designed to assess the use of semantic associations as a strategy for learning words. Each of the 16 words in each CVLT list belongs to one of four categories of "shopping list" items: for example, List A—"Monday's" list—contains four names of fruits, of herbs and spices, of articles of clothing, and of tools; List B—the "Tuesday" interference list—also contains names of fruits and of herbs and spices plus four kinds of fish and four kitchen utensils. The CVLT-II categories—no more shopping lists—for List A are vegetables, animals, ways of traveling, and furniture, with vegetables and animals in List B along with musical instruments and parts of buildings. The CVLT-II includes an alternate form and a short form (CVLT-IISF). The words are read at a rate slightly slower than one per second.

Category items are presented in a randomized order with instructions to recall the words in any order, thereby assessing the subject's spontaneous use of semantic associations. While examination of the use of strategies offers an advantage, it creates disadvantages as well. CVLT performance is a measure of the interaction between verbal memory and conceptual ability, so scores cannot be evaluated as exemplars of the patient's learning ability *per se* because of the possible confounding effects of concept apprehension and conceptual organization (Delis, 1989). However, when it is important to assess whether and how well a patient uses learning strategies based on concept formation, this test offers an advantage.

The procedure is identical for the two CVLT editions and similar to the AVLT. Following five trials with List A, the interference List B is read to the subject. Two "short delay" recalls of List A are obtained. The first of the two recall trials is "free" recall in which the request for the subject to "tell me all" remembered items from List A is identical to the AVLT free recall procedure. Immediately following the free recall trial the examiner asks the subject to recall items in each of the categories (fruits, then herbs and spices, etc.). For subjects who used semantic clustering during the learning phase, cueing at delayed recall offers little additional benefit. However, subjects who failed to make the semantic associations during the learning trials often benefit from this cueing. The enhanced recall due to cueing at the short delay also should carry over to the free recall requested 20 minutes later. This "long delay" trial measures recall of List A under the same two conditions, "free" and "cued." In one study, category cueing following short delay free recall failed to facilitate long delay recall in college students. The students receiving standard cueing instructions recalled the same number of words as students receiving no cueing during short delay recall even though the cued students used more semantic clustering in their delayed free recall than did those who received no cueing (P.K. Shear et al., 2000). The failure to find performance enhancement associated with cueing might have been due to a near ceiling effect as these students recalled an average of over 14 of the 16 words on trial 5.

The recognition trial format also differs from that of the AVLT in its oral presentation of just 44 items (48 on CVLT-II): of course all items of List A are included, but only eight from List B, two from each of the two overlapping categories and two from each of the two new categories. On the CVLT, four nonlist items each come from one of the four List A categories (e.g., tools: hammer; spices and herbs: pepper), eight items bear a phonetic resemblance to List A items (e.g., grapes: tapes, parsley, pastry); and the remaining eight are simply items one might find in a very large supermarket (e.g., film, clock). In my experience [mdl], these latter items have never elicited false positive errors.

Practice effects, which tend to be particularly prominent on memory tests (McCaffrey, Cousins, Westervelt, et al., 1995), led to the development by Delis, Kramer, and their coworkers (1983, 1987) of an alternate form of the CVLT that is parallel to the original in every respect (Delis, McKee, et al., 1991). Sixteen of the 19 scores (Principal CVLT Variables, see below) generated by this test were significantly correlated, half of these equal to or greater than $r = .67$.

Responding to the problems inherent in the original CVLT, Delis and his coworkers developed the CVLT-II. It is intended to replace the first test and not simply serve as an alternate form. Differences in the normative samples between tests preclude the interchangeability of standard score equivalents of raw scores between versions. The major changes are that the CVLT-II uses different categories of items that have higher familiarity than on the original form and the normative sample is much larger. An optional forced-choice recognition measure is obtained approximately 10 to 15 minutes after the yes/no recognition. Because forced-choice with completely unrelated items is easier than yes/no recognition, this measure was added to detect motivation lapses. Revisions have been made in calculating certain scores; repetitions are now called "repetitions," not "perseverations" and a new clustering score has been added.

A comparison of the CVLT and CVLT-II in a sample of 62 healthy adults tested approximately seven days apart on the two tests showed good equivalency in Total Trials 1–5 recall and Long-Delay Free Recall (Delis, Kaplan, Kramer, and Ober, 2000). Data for Short-Delay Recall and the cued recalls were not given. The CVLT and AVLT were compared in 60 normal adults and were found to provide similar information although scores were slightly higher on the CVLT (Crossen and Wiens, 1994). Raw scores for learning and short delay retention were nearly identical across tests for a TBI group given both in spite of the design differences in the tests (Stallings et al., 1995). German (Hildebrandt et al., 1998) and Korean (Kim and Kang, 1999) versions of the CVLT have been validated.

CVLT scores. In addition to the acquisition scores for trials 1, 5, and B, and retention of List A following free and cued trials for short and long delays, and *Recognition Hits,* other scores include: *List A Total Recall,* which is the sum of trials 1 through 5; *Learning Slope,* which quantifies the rate of learning; *Recall Consistency* on the learning trials; plus scores that reflect learning strategies (i.e., *Semantic Cluster Ratio, Serial-order Cluster Ratio,* and scores for primacy, recency tendencies); plus still more scores for comparing Trials I and B, free and cued recall, and recall and recognition and scores for such other response characteristics as repetitions (given the misnomer, *Perseverations,* see p. 422), *False Positives,* and *Intrusions* (for free and for cued recall trials). Together, not counting Number Correct scores for trials 2, 3, and 4, these total at least 19 different CVLT scores (27 on the research edition of the alternate CVLT form).

The CVLT-II changes the ways in which semantic and serial clustering scores are calculated, and a new subjective clustering score has been added. *Intrusions* and *repetitions* are scored and subtyped according to noncategory or category characteristics, and further into synonym/subordinate intrusions and across-list intrusions. Both *proactive* (intrusions from List A into List B recall) and *retroactive* (from List B into delayed recall and recognition trials) interference can be documented. Also included are scores for evaluating signal detection efficiency and response biases.

Many of the 27 CVLT scores (42 CVLT-II scores) with normative data are highly intercorrelated such that the most useful data are obtained from the List A trial 5, List A $\Sigma 1$–5, semantic clustering, free and cued recall (both short and long delays), intrusions (both free and cued recall), and recognition hits and false positives (Elwood, 1995). All but semantic clustering is easily scored by hand, and in most cases the examiner knows whether or not the subject has used semantic clustering based on the obtained recall pattern without using a complex computational formula. A guide for calculating all of the supplementary scores by hand is included in the test manual. The publisher markets a computer scoring system which is helpful for complex calculations of such scores as Recognition Discriminabilty (CVLT-II) and Learning Slope, although the neuropsychological import of these scores remains in question.

Normative data. The CVLT manual provides normative data for 273 males and females in seven age groups covering 17 to 80 years. These data have been criticized for poor standardization and inflated norms (Elwood, 1995). Norms demographically corrected for age, education, ethnicity (Caucasians and African Americans), and sex are available (M.A. Norman et al., 2000). Norms for elderly males and females have been compiled by Paolo, Tröster, and Ryan (1997a). The CVLT-II has an improved normative sample of 1,087 adults in seven age groups ranging from 16 to 89 years and stratified according to the U.S. census by age, sex, ethnicity, educational level, and region of the country. Because age and sex account for significant differences between individuals, norms are provided for males and females within each age group.

Test characteristics. Age-associated CVLT performance decrements are seen, particularly at age 65 and beyond (Paolo, Tröster, and Ryan, 1997a; Pope, 1987). Significant age × Total Trials 1–5 correlations of $-.61$ for the original test and $-.65$ for the alternate form reflect the important contribution of age to recall (Delis, McKee, et al., 1991). For example, the learning curve is a little flatter for older persons. However, at least into the 60s, immediate span (trial 1) and recognition remain essentially unchanged, with relatively few false positive errors (J.H. Kramer, Blusewicz, and Preston, 1989). A trend toward fewer clusters with advancing age has been reported (Pope, 1987). Interestingly, while the number of intrusions increased with advancing age, false positive errors on the recognition trial continue to be small (around 1.1 ± 2.2) into the 75 to 91 age range (Pope, 1987). Education correlations are positive and significant, but lower than those for age (.36 and .39 for original and alternate test forms, respectively) (Delis, McKee, et al., 1991).

In the sample used to produce demographically corrected norms, Caucasians outperformed African Americans and women were superior to men (M.A. Norman et al., 2000). It is unknown if the ethnicity effect would have been significant if the African Americans had been as well-educated as the Caucasians. Many studies show that women tend to outperform men on learning and recall measures of this test (Paolo, Tröster, and Ryan, 1997a; Ragland et al., 2000; Reite et al., 1993; Wiens, Tindall, and Crossen, 1994) although no sex differences showed up for the recognition trial or for error types (J.H. Kramer, Delis, and Daniel, 1988). Women use a semantic clustering strategy more often than men (Berenbaum et al., 1997; J.H. Kramer, Delis, and Daniel, 1988). Personal experience [dbh] suggests that men are less likely than women to make a distinction between the categories of fruit and herbs/spices, instead treating them all as "edibles." Some men are unfamiliar with spices such as paprika, further contributing to a breakdown in use of categories on the original CVLT. Based on a study of elderly men twin pairs, CVLT measures of learning and recall have a large genetic component (56% of total variance), whereas the genetic component is not significant on learning strategy or recognition memory (Swan et al., 1999).

Reliability studies undertaken by the test's authors give split-half (odd–even item, odd–even learning trials, 2 × 2 categories) reliability correlation coefficients of .77 to .86 (Delis, Kramer, Fridlund, and Kaplan, 1990). In older adults tested slightly over a year later, the stability coefficient for acquisition and long-delay free recall was .76 (Paolo, Tröster, and Ryan, 1997b). CVLT-II reliability correlations are also high (Delis, Kaplan, Kramer, and Ober, 2000). Split-half reliability correlations of scores from Total Trials 1–5 range from .87 to .89, and alternate form reliability ranges from .72 to .79 for various measures. Test–retest (21 days later) reliability was .82 for Total Trials, although it was much lower for some of the many variables, most notably Total Learning Slope (.27) and Total Repetitions (.30).

Factor analyses have consistently shown a large general verbal learning factor with small effects of response discrimination, learning strategy, proactive interference, and serial position (Delis, Freeland, et al., 1988; Millis, 1995; Schear and Craft, 1989; Vanderploeg, Schinka, and Retzlaff, 1994). An alternative four-factor model (attention span, learning efficiency, delayed recall, and inaccurate recall) has also been described (Wiegner and Donders, 1999). Together these studies demonstrate how different are various aspects of verbal memory, thus supporting the conceptual framework of this test. Attention plays a significant role in performance (Rapport, Axelrod, et al., 1997; Vanderploeg, Schinka, and Retzlaff, 1994). WAIS-R Vocabulary alone accounted for up to 13% of the variance (P.A. Keenan, Ricker, Lindamer, et al., 1996). Correlations with other memory tests are reported to be "modest" (Schear and Craft, 1989).

Neuropsychological findings. Clinical studies using the CVLT-II are few in number at the time of this writing. The CVLT-II performance of a small group of patients with circumscribed frontal lobe lesions was compared to that of control subjects (Baldo, Delis, et al., 2002). The frontal patients had a depressed learning curve, an increased tendency to make intrusions, reduced semantic cluster, and impaired yes/no recognition performance because of a tendency to endorse semantically related distractors and words from the interference list. Both groups slightly benefited from cueing and both recalled slightly more words in the Long-Delay Free Recall than in the Short-Delay Free Recall. The interpretation of these findings supported the theory that the frontal lobes play an important role in strategic memory processes and source memory. Using a 12-item version of the CVLT (three categories, four items each), Ricker, Müller, and their coworkers (2001) looked for cerebral blood flow changes by PET scanning during recall and recognition trials, and found differences between small samples of normal control subjects and of patients with severe traumatic injuries. Changes occurred predominantly in frontoparietal regions for both groups. On free recall, controls showed significant blood flow increases in the left middle frontal gyrus, left medial cerebellum, and left angular/supramarginal gyrus whereas the patients showed only increase in the left angular supramarginal gyrus. On recognition testing, both groups had increased blood flow in the middle frontal gyri bilaterally, with the TBI patients experiencing a greater increase than control subjects.

Performance patterns generated by the CVLT effectively discriminate many—but not all—patient groups. The CVLT has been used to examine the interaction between memory and the frontal executive impairments associated with TBI (Deshpande et al., 1996; Haut and Shutty, 1992; Kibby et al., 1998; Novack et al., 1995). Crosson, Novack, and their colleagues (1989) describe three different memory problem patterns among TBI patients—consolidation deficits, impaired encoding, and retrieval deficiencies—which show up primarily in their *number correct* and *error* scores on the recognition trial. Deshpande and her colleagues (1996) elicited five learning subtypes using cluster analysis: "Active, Disorganized, and Passive" and two similar to a "Deficient subgroup." For example, the Active group were least impaired, the Disorganized group the most impaired, unable to profit from repetitions.

The interpretation of recognition–recall discrepancies as retrieval deficits has been challenged (Cicchetti, 1997; M.C. Wilde et al., 1995, 1997) and defended (Veiel, 1997a,b). Briefly, the argument is that recognition is easier than free recall for patients with retrieval deficits as it puts less demand on retrieval. Wilde and his colleagues (1995) tested this assumption by comparing TBI patients who had recognition–recall discrepancies with those who did not. Two predictions about patients with these discrepancies were not supported: that they would have better cued recall than free recall, and that they would have less consistent recall as measured by the Consistency Index. One prediction, that patients with recognition–recall discrepancies would show fewer intrusions during both free and cued recall, was supported. Wilde and his colleagues concluded that this study gave little support for the retrieval interpretation while Veiel accepted the partial support.

The CVLT performances of patients with left temporal lobe seizures differ from those of patients with seizures associated with right temporal lobe dysfunction and from normal control subjects on a number of variables including fewer words learned and fewer total number of words recalled, less semantic clustering, and poorer retrieval, although their recognition level was similar to that of the other two groups (B.P. Hermann, Wyler, Richey, and Rea, 1987). Mean scores also showed that patients with left temporal lobe involvement had a greater loss from trial V to short-term delayed recall and more false positive errors, but intragroup variability was so large that these differences did not reach significance. Most of the CVLT variables did not discriminate between patients with right temporal lobe involvement and normal subjects. Left anterior temporal lobectomy seizure patients have greater deficits after surgery than patients with comparable surgeries in the right hemisphere (K.G. Davies et al., 1998a,b). Using a German language version of the CVLT, Hildebrandt and his colleagues (1998) found that patients with strokes involving both the left prefrontal cortex and left temporal lobe were impaired in encoding and recall compared to a group with right posterior lesions. Performance of the left temporal lobe groups was significantly worse than that of the left prefrontal group for final learning level, short-term cued recall, and learning slope. Reliance on a serial clustering strategy was a distinguishing characteristic of patients with left mediotemporal lesions. Recognition performance was relatively intact in all the stroke groups. B.J. Diamond and his colleagues (1997) observed that patients with anterior communicating artery aneurysm were defective in both acquisition and retention of information, in scanty use of semantic cluster, and in making an abnormally large number of intrusion and false positive errors.

The CVLT readily elicits the memory problems of Alzheimer patients, even those with very mild disease (Bayley et al., 2000; Bondi, Monsch, Galasko, et al., 1994). Alzheimer patients recall fewer words from the beginning of the list than controls (Bayley et al., 2000); their greatest difficulty is with free recall on long delay (Massman, Delis, and Butters, 1993; Zakzanis, 1998). Learning and retention deficits in Alzheimer patients are associable to atrophy of the mesial temporal lobe and thalamus (Libon, Bogdanoff, et al., 1998; Stout et al., 1999). Many Parkinson patients do not spontaneously use semantic strategies as well as do controls (Knoke et al., 1998; van Spaendonck et al., 1996b). Because of the many variants of this disease, not all Parkinson patients will be impaired on the CVLT (Filoteo et al., 1997). Multiple sclerosis patients show a diminished acquisition rate, tend to use serial rather than semantic clustering, and recall fewer words than controls after delay (B.J. Diamond et al., 1997). However, their percentage savings score and recognition discrimination are unimpaired. Patients with HIV associated dementia perform poorly on acquisition but show relatively good retention over time (D.A. White et al., 1997). Using 14 CVLT scores, Murji and colleagues (2003) identified four subtypes in their 154 HIV patient subgroups, based on different loadings of four factors: Attention Span, Learning Efficiency, Delayed Recall, and Inaccurate Recall. The subtypes—Normal, Atypical, Subsyndromal, and Frontal–striatal—varied with respect to scores on a depression inventory and on tests of cognitive functioning. Interestingly, the poorest performers (Frontal–striatal) had less education than those in the other groups: the effects of a cognitive reserve were offered as one possibility to account for this; certainly others can be conjectured.

Each of the major degenerative diseases affects the CVLT variables somewhat differently, permitting a number of group-by-group discriminations and potentially aiding differential diagnosis in the individual case (Delis, Massman, et al., 1991; J.H. Kramer, Levin, et al., 1989). For example, more than 35% of the responses made by a group of Alzheimer patients were intrusions, compared with less than 10% for high functioning Huntington and Parkinson patients. Alzheimer patients forgot almost 80% of trial 5 recall following delay, high functioning Parkinson patients' loss was almost 30%, and delay allowed high functioning Huntington patients to improve their recall a little. "Perseveration" (i.e., repetitions, which may or may not have been true perseverations) also helped to differentiate these three groups, as high-level Huntington patients repeated themselves eight times on the average during the learning trials while Alzheimer patients made an average of about three repetitions and high level Parkinson patients averaged only one. Both Alzheimer and Huntington patients displayed very poor recall of the words at the beginning of the list, but Huntington patients recognized these words while Alzheimer patients gave no evidence of having learned them (Massman, Delis, and Butters, 1993). Compared to Alzheimer patients, patients with ischemic vascular dementia show superior recognition discriminability as their performance pattern is more similar to patients with subcortical dementia (i.e., Parkinson's disease, Huntington's disease) (Libon, Bogdanoff, et al., 1998). Patterns of deficits in Huntington patients also have been reported in asymptomatic Huntington gene carriers (Hahn-Barma et al., 1998).

Alcoholic patients differ from normal controls on most variables of this test, but their level of semantic clustering is similar and their learning curves have similar slopes although consistently lower than those of normal subjects (J.H. Kramer, Blusewicz, and Preston, 1989). Alcoholics are also relatively susceptible to interference effects. Korsakoff patients produced a profile similar to that of Alzheimer patients, making this test a poor discriminator of these two conditions, particularly in the early stages of Alzheimer's disease (Delis, Massman, et al., 1991). The CVLT may aid discrimination between early Alzheimer's disease and depression, as depressed patients show about the same score increase on the recognition trial relative to recall as normal subjects, but Alzheimer patients' recognition score is only a little better than their recall if at all (Massman, Delis, Butters, et al., 1992). Schizophrenic patients do not organize their recall by semantic categories (Kareken et al., 1996).

CVLT short forms. Shorter versions have been developed for severely impaired patients; being briefer they will be less stressful for both failing patients and concerned examiners. On a 9-word list composed of three 3-word categories, dementia patients of mild to moderate severity perform worse than controls (Libon, Mattson, et al., 1996). Vascular dementia patients were superior to Alzheimer patients on delayed recall measures and made fewer intrusions and false positive recognition responses. Woodard, Goldstein and their coworkers (1999) found moderate correlations with the WMS-R Logical Memory and Visual Memory as well as with nonmemory tests, suggesting that dementia patients' low memory test scores may indicate general cognitive deterioration. A six-word version containing three exemplars of two categories is useful for demonstrating the learning and recall impairment of patients with moderate Alzheimer's disease (Kaltreider et al., 1999).

The CVLT-II manual also gives a short version of the test. The CVLT-IISF has nine words in three categories, uses only one list instead of two, and gives only four learning trials. It calls for delayed recall at two intervals—30 secs (filled with counting backward as a distraction) and 10 mins, followed by a yes/no recognition trial. As with the standard version, a forced-choice recognition trial is optional.

CERAD Word List Memory (W.G. Rosen, Mohs, and Davis, 1984)

The Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (J.C. Morris, Heyman, et al., 1989; J.C. Morris, Edland, et al., 1993) includes a list of ten unrelated words for examining memory, a procedure incorporated in the Alzheimer's Disease Assessment Scale (ADAS) (W.G. Rosen, Mohs, and Davis, 1984). The short list is a suitable length for the very elderly and for Alzheimer patients who are likely to become distressed by longer lists. Its brevity is also useful for patients who are difficult to manage (Lamberty, Kennedy, and Flashman, 1995) and also would be appropriate for severely amnesic patients for whom longer word lists would be too taxing. The procedure has the advantage that the patient reads the words printed in large letters on cards, bypassing the hearing problems common to this age group and assuring registration of each word. The words are shown at a rate of one every 2 secs and presented in a different order on each of the three learning trials. Recall follows each trial. After a 3 to 5 min delay retention is tested by free recall and a recognition trial in which ten unrelated distractor words are intermixed with the target words. An alternate list of words of equal difficulty is available for repeat testing.

Age and education norms have been developed for Caucasian Americans (Ganguli et al., 1991; Welsh, Butters, Mohs, et al., 1994; Welsh-Bohmer et al., 2000) and African Americans (Unverzagt, Hall, Torke, et al., 1996). A significant age effect is reflected in the norms (Howieson, Holm, Kaye, et al., 1993; Unverzagt, Hall, Torke, et al., 1996; Welsh, Butters, Mohs, et al., 1994). In a sample of Caucasian Americans, women outperformed men and education affected final acquisition level but not delayed free recall (Welsh, Butters, Mohs, et al., 1994); yet no sex differences were reported in a more recent sample (Welsh-Bohmer et al., 2000). For African Americans, education contributed to acquisition, recall, and recognition scores; the only score on which women were superior was acquisition (Unverzagt, Hall, Torke, et al., 1996).

Correlations of CERAD and CVLT performances of probable Alzheimer patients on comparable variables were mostly at the $p < .001$ level (e.g., Total Words, $r = .65$; Last Trial Recall, $r = .48$; Recognition Hits, $r = .40$). However, lower correlations (e.g., False Positives, $r = .22$; Forgetting Rate Percent, $r = .15$) indicate that while, for the most part, these two tests measure similar aspects of verbal learning in this patient group, differences in CERAD and CVLT scores are sufficiently great that conclusions drawn from one of these tests cannot be automatically assumed for the other (Kaltreider et al., 2000).

The acquisition and free recall measures are sensitive to memory loss associated with early stage dementia (Greene et al., 1996; Howieson, Dame, Camicioli, et al., 1997; Welsh, Butters, Hughes, et al., 1992). Performance declines progressively with increasing severity of dementia (Welsh, Butters, Hughes, et al., 1991). K. Welsh and her colleagues (1991) observed that the delayed recall measure was the most useful in detecting Alzheimer's disease while others have suggested that the total acquisition score identifies Alzheimer patients best (Derrer et al., 2001). Impaired word recognition becomes evident with progression of dementia. Test sensitivity and scores were approximately the same for African American and Caucasian Alzheimer patients when differences between groups in age, education, and disease severity were statistically corrected (Welsh, Fillenbaum, Wilkinson et al., 1995).

Hopkins Verbal Learning Test (HVLT) (Brandt, 1991), *Hopkins Verbal Learning Test-Revised (HVLT-R)* (R.H.B. Benedict, Schretlen, et al., 1998)

In its original form this is a word list learning task in which 12 words, four in each of three semantic categories, are presented for three learning trials. This is followed by a 24-word recognition list containing all 12 target words plus six semantically related foils and six unrelated ones (Brandt, 1991). The words on each of six 12-word lists differ for each list. The six lists and the recognition format for each are given in Brandt (1991). The 1998 revision includes a 20 to 25 minute delayed recall trial that is forewarned plus the subsequent yes/no 24-word recognition trial. Scores include one for each learning trial, a total acquisition score, a learning measure, delayed free recall, percent retention, and delayed recognition. Recognition scores are calculated for true positives, false positives, a discrimination index (true positives − false positives), and a measure of the recognition trial response bias, *Br* (the sum of "yes" responses).

Test characteristics. Stability coefficients over nine months using different forms were moderate for total recall ($r = .50$) in healthy older adults (Rasmusson et al., 1995). The six alternate forms are equivalent for the recall trials but recognition scores differ slightly (R.H.B. Benedict, Schretlen, et al., 1998).

In many ways this is a short version of a CVLT-type task, as indicated by a relatively high correlation ($r = .74$) for total learning for the two tests (Lacritz and Cullum, 1998). Validity studies demonstrated the comparability of HVLT-R recall and recognition measures to memory measures from other tests, particularly verbal memory tests (Lacritz, Cullum, al., 2001; A.M. Shapiro et al., 1999). Unimpaired adults achieve ceiling scores easily (Lacritz and Cullum, 1998). Normative data from unimpaired young adults on the HVLT-R produced a mean recall of 11 out of 12 words on the last learning trial, and the delayed recall mean was 10.6. Healthy, well-educated older adults (mean age, 70.7 ± 9.3) approach ceiling on the last learning trial. By contrast, mean performance of the normative group ages 70–88 did not approach ceiling. In a larger study of older adults, women performed better than men and there was a significant effect of age but not education (Vanderploeg, Schinka, Jones, et al., 2000). These authors provide normative data for Form A for ages 60 to 84 with adjustments for sex.

Neuropsychological findings. Patients with Alzheimer's disease and those with vascular dementia show a learning deficit on the HVLT (A. Barr et al., 1992; Frank and Bryne, 2000; Hogervorst et al., 2002; A.M. Shapiro et al., 1999). Comparing Huntington and Alzheimer patients on recognition trial scores, Brandt, Corwin, and Krafft (1992) found Alzheimer patients more likely to say "yes" to semantically related foils than the Huntington patients, and unlike control subjects who had made no false positive errors on unrelated foils, both kinds of dementia patients said "yes" to some of them. The HVLT has been useful in predicting which males will have a postconcussive syndrome after minor head injury (Bazarian et al., 1999). Of those with scores of 25 or greater on the summed learning trials, 92% did not have a postconcussive syndrome one month after injury. This relationship did not hold for females with minor head injury. In young adults the delayed recall and recognition measures proved sensitive to the effects of a relatively low blood alcohol level (0.07 mg/dl) compared to the drug-free state (Acheson et al., 1998).

Interference Learning Test (M.M. Schmidt and Coolidge, 1999)

This verbal learning task is uniquely designed to evaluate the effect of interference on learning a list of words. The ideal patient for this test would seem to be a high functioning adult who has performed adequately on a more standard word list learning task despite complaints of everyday memory problems in settings with more distractions than occur in a quiet testing room. The initial list, given in four trials, consists of 20 target words intermixed with 24 distractors, each exposed for three secs. The subject is instructed to read each word from a card but remember only the words printed on white cards and ignore those on blue cards. The target words consist of four nouns from each of three categories and eight abstract nouns. The to-be-ignored distractor words include 20 concrete nouns, some from target categories, and four abstract nouns, making them suitable for taxing source memory and inviting intrusion errors. Sixteen target words and no distractors make up a second list which is presented twice. Four words come from each of two categories, one the same as the original list and one different, and eight abstract nouns.

Recall of List 1 is obtained after two trials with List 2 (short delay) and after 30 mins (long delay). The delayed free recall is followed by a category cued recall trial and a yes/no recognition trial. The recognition words include the 20 target words from List 1, five to-be-ignored words, three words from List 2, and 14 words either from List 1 categories or unrelated to any lists. Finally, recall of the to-be-ignored words is tested with yes/no recognition. This test yields measures of learning under conditions of interference, source memory, intrusion errors, and organizational strategies. Computer scoring is available.

The normative sample consisted of 393 mostly Caucasian participants in the 16 to 93 age range ($M = 35.7$) with an average education of 14.2 years. Although the age group sizes are not given, older individuals appear to be underrepresented. Normalized *T*-scores were derived from a regression equation taking into account age and "premorbid IQ." A subsample of 20- to 29-year-olds gave no more intrusions on this test than on the AVLT or the CVLT despite efforts to produce interference with the to-be-ignored words.

Supporting the hypothesis that multiple sclerosis patients are particularly sensitive to the disruptive effects of interference during learning, a group of 30 MS patients had poorer overall recall when the distractor words were included (List 1 recall compared to List 2 recall) and made more intrusion errors than controls (Coolidge, Middleton, and Griego, 1996). In samples of patients with a variety of brain disorders, only the MS group made significantly more source errors than the control group (M.M. Schmidt and Coolidge, 1999). Clinical usefulness will determine whether the extra time needed to administer this complex task compared to the AVLT or CVLT is time well spent.

Selective Reminding (SR) (Buschke and Fuld, 1974)

The differentiation of retention, storage, and retrieval may also be accomplished with the selective reminding procedure. As this is a procedure, not a specific test, it has been given in many different ways. Subjects usually hear (or may be shown one by one on cards [Masur, Fuld et al., 1989]) a list of words for immediate recall. On all subsequent trials, subjects are only told those words they omitted on the previous trial. The procedure typically continues until the subject recalls all words on two successive trials or to the twelfth trial, although the original procedure called for repeated trials until all the words were recalled at one time. Those subjects who recall all words before the twelfth trial must reach a standard recall criterion (complete recall on three consecutive trials) before given a delayed recall trial. Not all subjects reach this criterion in 12 trials. Considerable variation can exist between subjects on the number of times each word must be presented (i.e., was not recalled in the previous trial). Some examiners give both a cued and a four-choice recognition trial after the last or twelfth trial (H.S. Levin, 1986; Spreen and Strauss, 1998) (see Table 11.7, p. 437). Most examiners ask for a free recall after 30 minutes (e.g., Hannay and Levin, 1985; Spreen and Strauss, 1998) or one hour (Ruff, Light, and Quayhagen, 1989).

The original version called for ten items, all of the same category (animals, clothing) to be read at a 2-per-sec rate (Buschke and Fuld, 1974); 12 items is the most common list length today. Erickson and Scott (1977) pointed out that such category-restricted lists lent themselves to successful guessing. Most examiners now use lists of unrelated words. A set of four comparable 12-word lists developed by Hannay and Levin (1985) is in most common use. Loring and Papanicolaou (1987) noted that different examiners have reported findings on different lists of different composition and length, making it difficult to draw generalizations from the literature. For example, McLean, Temkin, and their colleagues (1983) used a 10-item list giving a maximum of ten trials; Gentilini and his coworkers (1989) also gave ten trials but with a 15-item list; and Masur and his colleagues (1989, 1990), using the usual 12-item list, gave a maximum of six trials. See p. 437 for lists.

SR scores. Unique to selective reminding procedures is a measure of those words consistently recalled from trial to trial without further reminding: *Consistent long-term retrieval* (*CLTR*). (Masur and his colleagues [1990] further restricted the definition of this score as "the number of items the subject is able to recall on at least the last three trials without reminding.") Ten other scores can be obtained (Hannay and Levin, 1985; Spreen and Strauss, 1998) although some workers compute fewer (e.g., Ruff, Light, and Quayhagen, 1989). The full score roster for the learning trials includes, along with CLTR: *Sum recall* (ΣR); *Long-term retrieval* (*LTR*) or *Long-term storage* (*LTS*), the number of words recalled on two or more consecutive trials (i.e., without intervening reminding); in *Short-term recall* (*STR*) are words recalled only after reminding; *Random long-term retrieval* (*RLTR*) refers to words in LTS that do not reappear consistently but require further reminding; *Reminders* is the sum of reminders given in the course of the procedure; *Intrusions* are words not on the list. Three additional scores are given for the number of words recalled on cueing, by means of the multiple choice procedure, and on the delayed free recall trial. Additionally, Spreen and Strauss (1998) recommend noting the number of words recalled on the first trial (i.e., the supraspan).

Test characteristics. The test format provides numerous intercorrelated scores (Burkart and Heun, 2000; Loring and Papanicolaou, 1987; Spreen and Strauss, 1998) that are assumed to represent short-term recall, long-term recall and storage, and retrieval. Only words recalled on two consecutive trials are assumed to be in long-term storage. In support of this assumption, Beatty, Krull, and colleagues (1996) found that words retrieved from CLTR on the last acquisition trial were more likely to be recalled after delay than were words not consistently retrieved. An inadequate recall of words from long-term storage is assumed to represent a retrieval failure. However, other interpretations are possible. Loring and Papanicolaou (1987) pointed out that RLTR may represent weak encoding of words rather than a retrieval failure. Low CLTR scores of multiple sclerosis patients was interpreted as representing difficulties in the acquisition/encoding of information (J. DeLuca, Gaudino, Diamond, et al., 1998). Given the popularity of this test, the lack of comparisons

TABLE 11.7 Multiple-Choice and Cued-Recall Items for Forms 1–4 of SRT

Form 1			
1. bowl	dish	bell	view
2. love	poison	conform	passion
3. dawn	sunrise	bet	down
4. pasteboard	verdict	judgement	fudge
5. grand	grant	give	jazz
6. see	sting	fold	bee
7. pain	plane	pulled	jet
8. county	state	tasted	counter
9. voice	select	choice	cheese
10. flower	seed	herd	seek
11. date	sheep	wool	would
12. mill	queen	food	meal
Form 2			
1. shine	glow	chime	cast
2. dispute	disappear	contour	disagree
3. fat	oil	trail	fit
4. stopwatch	affluent	wealthy	worthy
5. trunk	drunk	stoned	blunt
6. fin	peg	wake	pin
7. glass	grass	plan	lawn
8. moon	beam	spark	noon
9. propose	ready	prepare	husband
10. award	prize	pot	size
11. bark	bird	duck	luck
12. leap	ranch	blade	leaf
Form 3			
1. throw	toss	through	plate
2. flower	lilt	intent	lily
3. film	movie	slave	kiln
4. waver	cautious	discreet	distinct
5. soft	loft	attic	tack
6. beet	meat	clue	beef
7. stream	street	speed	road
8. helmet	armor	bacon	velvet
9. smoke	serpent	snake	pool
10. hoed	dug	hay	dog
11. blank	bundle	pack	puck
12. ton	shirt	foil	tin
Form 4			
1. egg	shell	beg	source
2. airline	runner	darling	runway
3. fort	castle	sink	fork
4. boldness	dentist	toothache	headache
5. blown	drown	float	rib
6. body	infant	middle	baby
7. larva	lava	echo	rock
8. damp	moist	hook	stamp
9. purse	clean	pure	bare
10. ballot	vote	dish	note
11. chain	peal	strip	slip
12. trust	rise	fact	truth

Cued Recall Words

Form 1		*Form 2*		*Form 3*		*Form 4*	
BO	PL	SH	GR	TH	ST	—	LA
PA	COU	DI	MO	LI	HE	RU	DA
DA	CH	FA	PRE	FI	SN	FO	PU
JUD	SE	WEA	PR	DI	DU	TO	VO
GR	WO	DR	DU	LO	PA	DR	ST
—	ME	—	LE	BE	—	BA	TR

From Spreen and Strauss (1998).

of the selective reminding procedure with other word learning tasks with full reminding procedures in the same individuals is disappointing. Such comparisons would indicate whether the unique measures of this procedure, including CLTR, identify memory problems better than scores from the less complicated rote learning procedures.

Up to age 70, women tend to outperform men, with age of lesser importance and education contributing only little and that mostly below the college level (Ruff, Light, and Quayhagen, 1989). The Ruff group present normative data for LTS and CLTR for men and women separately, each data set stratified by age (four ranges from 16–24 to 55–70) and education (three levels, ≤12, 13–15, ≥16). They attributed at least some of the women's advantage to their greater use of a clustering strategy (e.g., by their temporal relationships—primacy or recency effects, or conceptually, e.g., *plane* and *bee* both fly). When the age range includes subjects over 70, age becomes an important variable, with sex effects of smaller but still significant consequence (Larrabee, Trahan, et al., 1988). Larrabee and his colleagues published norms for seven age groups from 18–29 to 80–91 which include all 11 of the usual scores. They provided correction values to bring men's scores up to women's levels (reproduced in full in Spreen and Strauss, 1998). See Table 11.8 for these age × sex norms for the three most used scores, CLTR, ΣR, and LTS.

Versions of this test using both six and 12 trials have been compared for sensitivity. Because 12 trials can be tedious, the shorter version would be preferable if it were shown to be as sensitive as the longer version. Norms have been reported for all scores generated on a six-trial version for cognitively intact 56 adults, ages 18 to 58 (Ehrenreich, 1995). Revised normative data for a six-trial administration have also been extracted from Larrabee and colleagues' 1988 data (Larrabee, Trahan, and Levin, 2000). As with other comparisons between six- and 12-trial versions (Drane et al., 1998; R.L. Smith et al., 1995), the correlations for various measures were high, ranging from .81 to .95 with the notable exception of a lower correlation (.51) for RLTR. The authors point out that the correlations are likely to be inflated because scores on the 12-trial version are based, in part, on cumulative scores at trial 6.

Reliability has been examined by test–retest procedures using the different forms. Test–retest reliability has correlations in the .41 to .62 range using seven of the learning measures for all four forms (Hannay and Levin, 1985), and higher—.73 and .66—for ΣR and CLTR using only Forms I and II (Ruff, Light, and Quayhagen, 1989). Larrabee, Trahan, and their colleagues (1988) reported that Form I is more difficult than Forms II, III, and IV, which were comparable. However, equivalency among the four forms has also been documented (Westerveld et al., 1994). A substantial practice effect for most of the scores appeared with four administrations using different forms of the test regardless of the order of the forms (Hannay and Levin, 1985). Correlational studies with other memory tests consistently bring out this procedure's significant verbal memory component (Larrabee and Levin, 1986; Macartney-Filgate and Vriesen, 1988).

Neuropsychological findings. Typically, studies report only one or a few scores, mostly CLTR, often with ΣR or LTS. SR measures of storage and retrieval have not only distinguished severely head injured patients from normal control subjects, as expected (H.S. Levin, Mattis, et al., 1987; Paniak et al., 1989), but have effectively documented impairment in mildly injured patients (McLean, Temkin, et al., 1983). Differences in learning efficiency show up between patients whose head injuries differ in their severity (H.S. Levin, Grossman, Rose, and Teasdale, 1979): on long-term storage, only the seriously damaged group did not continue to show improvement across all 12 trials, but leveled off (with an average recall of approximately six words) at the sixth trial. The mildly impaired group achieved near-perfect scores on the last two trials, and the moderately impaired group maintained about a one-word-per-trial lag behind them throughout, showing a much

TABLE 11.8 Norms for the Most Used SR Scores for Age Groups with 30 or More Subjects

	AGE (YEARS)			
	18–29	*40–49*	*60–69*	*70–79*
Education (years)	12.88 ± 1.7	14.71 ± 2.7	13.40 ± 3.6	13.46 ± 3.8
M/F	23/28	19/12	33/17	38/21
Scores				
CLTR	115.12 ± 19.7	107.10 ± 26.6	88.92 ± 35.8	69.68 ± 36.0
ΣR	128.18 ± 9.2	125.03 ± 12.0	114.82 ± 15.8	105.27 ± 16.7
LTS	125.00 ± 10.5	122.45 ± 15.6	107.00 ± 21.8	95.54 ± 24.9

From Larrabee, Trahan, Curtiss, and Levin (1988). Reprinted with permission.

less consistent retrieval pattern than the mildly impaired group. ΣR and CLTR also were sensitive to continuing improvements in moderately to severely injured patients over a two-year span (Dikmen, Machamer, et al., 1990). Paniak and his colleagues (1989) observed that CLTR in itself did not adequately account for a tendency of severely head injured patients to have an abnormally high rate of random recall which these authors attribute to inefficient learning but, rather, may reflect erratic retrieval mechanisms.

Lateralized temporal lobe dysfunction, whether identified on the basis of seizure site or due to anterior lobectomy, is readily discriminated by significantly depressed CLTR and LTS scores when the damage is on the left (Drane et al., 1998; Giovagnoli and Avanzini, 1996; G.P. Lee, Loring, and Thompson, 1989). However, neither CLRT nor LTS differentiated those patients whose left temporal lobectomies did not include the hippocampus from those with larger resections that did (Loring, Lee, Meador, et al., 1991). More impairment is associated with left than right frontal lesions for total words recalled, although impairment is evident with both lesion sites (Vilkki, Servo, and Surmaaho, 1998).

The selective reminding format has been used successfully to elicit memory impairment in patients with very mild Alzheimer's disease or with mild cognitive impairment that does not yet meet criteria for a dementia diagnosis (Devanand et al., 1997; Petersen, Smith, Waring, et al., 1999). Masur and his colleagues (1989) found that LTR and CLTR were the scores that best distinguished patients with early Alzheimer's disease from normal controls. They also report (1990) that SR scores—ΣR and the delayed recall score—were particularly sensitive predictors of which apparently normal elderly persons might develop Alzheimer's disease within two years of the initial examination, predicting well above baseline rates (37% and 40%, respectively) for these two scores. Prediction rates of most other SR scores were comparable, except STR (i.e., supraspan), which is generally relatively insensitive to very early dementia. In a more recent study the Buschke research group caution that using age- and education-corrected scores reduces the sensitivity for detecting dementia by as much as 28% compared to uncorrected scores (Sliwinski et al., 1997). They recommend using memory scores without age corrections for detecting mild dementia. Patients with dementia alone or combined with Parkinson's disease or stroke achieved scores significantly below those of nondemented Parkinson or stroke patients (Stern, Andrews et al., 1992). Multiple sclerosis patients performed significantly below normal control subjects on CLTR but not on delayed recognition (DeLuca et al., 1998; S.M. Rao, Leo, and St. Aubin-Faubert, 1987). However, considerable variability exists among MS patients. Beatty and his colleagues found that 25% of these patients performed normally while the remainder showed varying degrees of impairment with the SR procedure (Beatty, Wilbanks, et al., 1996).

SR variants. One variant of the SR procedure is *Free and Cued Selective Reminding (FCSR)* (Buschke, 1984; Grober, Merling, et al., 1997). The FCSR uses category cues at both acquisition and retrieval in an attempt to ensure semantic encoding and enhance recall. The subject is asked to search a card containing line drawings of four objects and to identify the one that belongs to a category named by the examiner. Each of the 16 items to be learned appears on one of four of these cards. After each item on the card is correctly identified, the card is removed and immediate recall of the four items is tested by cueing with the category prompt. The subject is corrected for any errors. Additional items are presented four at a time in the same manner. After the study phase, four recall trials are obtained in which free recall is followed by cued recall for items not spontaneously reported. Missed items are presented again with their cues. Elderly subjects recall twice as many words from long-term memory in FCSR than in SR (Grober, Merling, et al., 1997). Normative data for the elderly have been reported from the MOANS project (Ivnik, Smith, Lucas, et al., 1997) and the Einstein Aging Project (Grober, Lipton, Katz, and Sliwinski, 1998). The latter group found that age, education, and sex influenced performance but race did not. Free recall impairment on the FCSR predicted the development of dementia as much as five years in advance of the diagnosis (Grober, Lipton, Hall, and Crystal, 2000). However, the usefulness of this test is limited by ceiling effects because category cueing makes recall much easier for most adults, including well-functioning elderly.

The *Double Memory Test (DMT)* was designed to be a more difficult version of the FCSR by increasing the items per category cue from one to four and increasing the list length to 64 items (Buschke, Sliwinski, et al., 1995). Two lists are used, one in which category cues are presented during acquisition as well as retrieval, the other in which category cues are presented only during retrieval. Both patients with dementia and controls benefit when category cues are presented during acquisition compared to cueing only during retrieval. The benefit in the dementia patients is inversely related to the severity of the dementia (Buschke, Sliwinski, et al., 1997).

Tuokko and her colleagues (with Crockett, 1989; with Gallie and Crockett, 1990) offered a pictorial form of the test that documented memory deterioration in elderly patients. Two Spanish versions of the SR have been developed (Campo et al., 2000).

Word List (Wechsler, 1997)

An optional verbal memory test that comes with the Wechsler Memory Scale-III models the AVLT procedure but is shorter. The examiner reads the list of 12 unrelated words in the same order for each of four trials at the rate of one word per $1^1/_2$ secs and instructs the patient to recall them in any order. A short delayed recall of the original list follows a one-trial interference list. The patient is told to expect another recall 30 min later. Recognition is obtained with a yes/no format in which the subject identifies the target words from an equal number of foils, also semantically unrelated. The manual provides normative data for eight measures: List A trial 1 recall, total recall, learning slope, the difference between trial 4 recall and short delay recall, long delay recall, recognition, and the percent of trial 4 recall retained in the long delay recall condition. Also, norms for a "contrast" measure of recall of the first trials of both lists are available.

A floor effect with a pronounced distribution skew for long delay recall begins with a relatively young age group. It is unlikely that this score distribution justifies conversion to scaled scores (see p. 145). For example, the mean number of words recalled in the long delay condition by the 55–65 year-old group is 3.5 and a recall of only one word is the lower limit of the *average* range. No obvious reason explains the low performance by the normative group. Although the word frequency of the items is slightly lower than for some word lists, it appears comparable on such important variables as concreteness, imagery, meaningfulness, and pleasantness. Other word lists do not use exactly the same procedure, making direct comparisons impossible. However, a similar age group (55–59 years) given five trials in which to learn the 15 AVLT words, an interference trial, and tested 30 mins later with no warning gave an average recall of 10.4 words with the lower limit of the *average* range at 8.3 words (Ivnik, Malec, Tangalos, et al., 1990). Given ten words (CERAD), three trials, and no interference list, a 50–69 year-old group with low education (9.1 years) recalled an average of 7.0 words with 5.8 words at the lower limit of the *average* range (Welsh, Butters, Mohs, et al., 1994). The relatively lower education of the older groups in the WMS-III normative population may contribute to the seemingly large age effect. For the elderly, the acquisition and recognition scores appear to provide better measures of memory impairment than does recall following long delay.

Paired associate word learning tests

The format of paired associate tests consists of word pairs that are read to the subject with one or more recall trials in which the first of the pair is presented for the subject to give the associated word. Thus it is a word-learning test with built-in cueing.

Associate Learning (PAL) (Wechsler, 1945); *Verbal Paired Associates (VePA)* (Wechsler, 1987, 1997)

This is perhaps the most familiar of the paired word learning tests. The original Wechsler format consists of ten word pairs, six forming "easy" associations (e.g., baby-cries) and the other four "hard" word pairs that are not readily associated (e.g., cabbage-pen) (Wechsler, 1945). The PAL list is read three times, with a memory trial following each reading. Total score is one-half the sum of all correct associations to the easy pairs plus the sum of all correct associations to the hard pairs, made within five secs after the stimulus word is read. Thus, the highest possible score is 21. The word pairs are randomized in each of the three learning trials to prevent positional learning. In its original format this was a test of cued new learning with no procedure in place for measuring retention. Some workers have added a 30-min delayed recall (e.g., Spreen and Strauss, 1998; Stuss, Ely, et al., 1985). A fourth-trial variation on the standard administration of the Associate Learning task reverses the order of word pair presentation by telling the patient, "I'm going to give you the second word and you give me the first" (Milberg, Hebben, and Kaplan, 1996). In this way, the examiner can determine whether the word associations were truly learned, or whether the patient's correct responses represent strings of passively learned phonetic associations. WMS-II provides an alternate form of the PAL (C.P. Stone et al., 1946) which tends to be significantly more difficult than Form I. Suspected dementia patients scored one-and-one-half points higher on Form I with a correlation between Forms I and II of .73 (Margolis et al., 1985). These studies found correlations between PAL Forms I and II of .61 and .73, respectively.

Verbal Paired Associates (*VePA-R*) in the 1987 revised edition of the Wechsler Memory Scale (WMS-R) contains just eight pairs, four of the original easy pairs and four hard pairs. Scoring is based on the first three trials, although subjects who have not learned all the pairs by the third trial get up to three more trials. Approximately one-half hour later a single recall trial is given. Scoring differs from the original format in that, while correct recall of easy and hard pairs is still counted separately, easy and hard pairs alike receive the same one-point value for each correct response.

On both forms—WMS and WMS-R—the score is evaluated only for the total sum. The WMS doubled value for the hard pairs gives more weight to the presumed new learning they require making them more vulnerable to many kinds of brain damage than easy pairs, which depend to a large extent on old learned associations (R.S. Wilson, Bacon, et al., 1982). The relative weight contributed by the easy associates increases with age (DesRosiers and Ivison, 1986).

The weight differential was demonstrated in a comparison of mildly injured TBI patients, patients whose injuries were severe, and normal subjects on the WMS version of paired associates, for the difference between the average recall of severely injured patients and control subjects for easy pairs was only 3.6 words, but for the hard pairs it was 4.9 (Uzzell, Langfitt, and Dolinskas, 1987). It also showed up when performances of multiple sclerosis patients were compared with those of control subjects as here easy words did not distinguish the groups on any of the three trials, but the hard word score differed significantly for the two groups on all trials (Minden, Moes, et al., 1990). However, these studies, in which easy and hard pairs were evaluated separately, are thoughtful exceptions to the WMS manuals' norms based on the combined scores. Of course the juxtaposition of hard and easy pairs in one test has the practical result of testing two different activities (i.e., recall of well-learned verbal associations and retention of new, unfamiliar verbal material), but the combined score obscures the status of each.

In *Verbal Paired Associations-III* (1997), all items are "hard," thus doing away with the relative insensitivity of the easy items. Administration and scoring are based on four trials with eight pairs. Following a delayed recall approximately 30 mins later, a recognition trial includes the previously presented word pairs intermixed with pairs of new words. This latter task, for which no norms are provided, appears to be too easy for most patients. Rather, the score from this recognition test is combined with the Logical Memory II recognition score in the normative tables which precludes comparison of this score with the normative sample!

In including only hard pairs, the test publishers eliminated one of the virtues of the earlier pair sets, as easy pairs provide opportunities for memory impaired patients to have some experience of success on otherwise consistently frustrating and even humiliating failures in a memory examination. When examining a psychologically frail elder, I [mdl] still use the word pair set from the WMS-R. Moreover, a subject's failure to make these easy associations can no longer alert the clinician to possible poor motivation.

Clinical experience [dwl, mdl] suggests that the delayed recall format lacks useful sensitivity. Although a paired associate recognition trial is given immediately after the delayed free recall, there are no item pairs in which the first word is incorrectly paired with the second word of a different word pair or a new, nonlist word. Thus the patient need only recognize either of the two words since they always appear as the correct target pair, defeating the purpose of testing acquired *word associations.*

Test characteristics. By inspection, small but consistent age decrements show up for this test (see Mitrushina, Boone, and D'Elia, 1999; Spreen and Strauss, 1998; Wechsler, 1987, p. 52). Within age ranges above 60, some studies found that age contributes little to score differences on the WMS version of this test (J.S. Bak and Greene, 1981; Ivnik, Smith, et al., 1991), while pronounced score declines for elderly subjects have been documented too (Kaszniak, Garron, and Fox, 1979; Margolis and Scialfa, 1984; Zagar et al., 1984). It may be suspected that when scores from the oldest age groups are similar to those in young–old groups, differences in the demographic composition of these age groups contribute to their performance similarities as education and income tend to be positively correlated with health and longevity. McCarty, Siegler, and Logue (1982), avoiding the demographic factors involved in longevity, retested their subjects over a decade or more and found that only the hard-pairs score decreased significantly. The more difficult VePA-III version shows a steady decline with age; subjects over 75 years generally recall no more than three or four of the pairs.

Despite the preponderant evidence for sex effects on word learning tests, evaluations for sex biases with appropriate norms or correction scores are not given in the WMS-R or WMS-III manuals. Commingling of scores from the story recall task, Logical Memory, on which men perform equally well if not better than women and Verbal Paired Associates for most test evaluation purposes possibly masks sex differences. One study considering sex differences on this test in the WMS version found that women performed better only on the hard associates, which makes sense in light of the sex differences that appear on other tests of rote verbal learning. In two other WMS studies, women tended to average about 1 point higher than the men at each age level until the 60s, when they had an approximately two-point lead (Ivison, 1977). However, sex differences have not always shown up on the WMS version (Trahan, 1985). Young women outperformed young men on the VePA-III first trial recall, total recall, and percent retention but not delayed recall (M.R. Basso et al., 2000). Education effects have been reported for both easy and hard pairs on the WMS version (Ivnik, Smith, et al., 1991).

Normative data from a number of studies have been compiled for the WMS forms of Verbal Paired Associates (Mitrushina, Boone, and D'Elia, 1999; see also Spreen and Strauss, 1998). The manuals give normative data for ages 16–74 for the VePA-R and 16–89 for the VePA-III. Unfortunately, for the WMS-R, only extrapolated data are provided for age groups 18–19, 25–34, and 45–54 as these groups were not included in the standardization (see p. 483).

A short-term (7 to 10 days) test–retest reliability correlation of .53 and a significant 1.33 point gain in mean score were documented for hypertensive patients (McCaffrey, Ortega, et al., 1992). This correlation is in line with a test–retest (after one year) reliability correlation of .63 obtained by W.G. Snow, Tierney, and their colleagues (1988). Three-week retesting of subjects from a broad age range (17–82) produced a relatively high reliability coefficient (.72) with an average 1.31 score gain ($p < .05$) (Youngjohn, Larrabee, and Crook, 1992). On retesting with the WMS-R, only small gains accrued on the learning trials for all three standardization groups (ages 20–24, 55–64, 70–74), with virtually no gains on delayed recall (Wechsler, 1987; see also McCaffrey, Duff, and Westervelt, 2000b for more test–retest data). Stability coefficients were highest on the VePA compared to all other WMS-III tests; retesting 2 to 12 weeks later resulted in a 1 to $1^1/_2$ point gain (The Psychological Corporation, 1997).

Examination of the construct validity of this test generally demonstrates a significant verbal learning component (Bornstein and Chelune, 1989; Larrabee and Levin, 1986) or a general learning factor that is relatively independent of verbal skills (Chelune, Ferguson, and Moehle, 1986; Larrabee, Kane, et al., 1985), depending on the mix of tests being studied. When easy and hard pairs are analyzed separately, the association of hard pairs with other verbal learning measures becomes evident (Macartney-Filgate and Vriezen, 1988).

Neuropsychological findings. Jones-Gotman (1991b) pointed out that this test falls short of the ideal for a verbal memory test as the words lend themselves readily to visual imagery, thus allowing the enterprising subject to use a dual encoding strategy. Yet, despite this potential drawback, both earlier versions of the test are sensitive to the effects of lateralized lesions: with patient groups of mixed etiologies, both learning and delayed recall means of patients with left-sided lesions were significantly below those made by patients whose dysfunction was on the right side (WMS-R: Chelune and Bornstein, 1988; WMS: Saling et al., 1993; Vakil, Hoofien, and Blachstein, 1992). These findings are supported by studies of patients who had temporal lobectomies for seizure control, as those with left-sided excisions not only performed at significantly lower levels than those whose surgery involved the right temporal lobe (WMS-R: L.H. Goldstein et al., 1988), but they also performed below their already depressed presurgery scores on this test (WMS: Ivnik, Sharbrough, and Laws, 1988). In a study of patients with temporal lobe epilepsy, poor VePA delayed recall of the left temporal group was the only WMS-III test that statistically distinguished patients with left and right foci (N. Wilde et al., 2001). Slightly but significantly and fairly consistently lower scores on delayed recall of the WMS version distinguished patients who had apparently "recovered" from mild head injuries from normal control subjects (Stuss, Ely, et al., 1985).

Paired-associate learning has proven useful not only in eliciting the learning deficits of Alzheimer type dementia (Bondi, Salmon, and Kaszniak, 1986) but also in documenting the progress of deterioration, even in the early stages (Kaszniak, Poon, and Riege, 1986; Storandt, Botwinick, and Danziger, 1986). However, of a group of seven patients with memory complaints associated with neurological disorders, of whom four had Alzheimer diagnoses, only one performed WMS paired associates at a level below that of the 12-person control group (Lussier et al., 1989). N. Butters, Salmon, Cullum, and their coworkers (1988) found that by calculating a *savings score* (delayed recall ÷ last immediate recall [whether trial 3 or higher] × 100) they could demonstrate retention levels by both young and old control subjects that were significantly better than those of patients with Huntington's disease or amnesia due to head injury or Korsakoff's psychosis. This differential did not show up for Alzheimer patients who, it may be presumed, had such low immediate recall scores that they could not lose much and still make any response at all. Moreover, the savings differential was smaller than for other WMS-R tests (Logical Memory and Visual Reproduction). Unfortunately these workers did not evaluate patients' learning and retention performances per se on this test as they used the WMS-R manuals' recommended combined VePA and Logical Memory score.

The paired associate format is useful in showing memory impairment in patients with basal ganglia disease. Although newly diagnosed Parkinson patients were impaired on the PAL, they showed a good savings score at a 1-hour delay (J.A. Cooper, Sagar, Jordan, et al., 1991). Memory for the hard pairs of the PAL distinguished presymptomatic gene carriers for Huntington's disease from noncarriers (Hahn-Barma et al., 1998). The VePA-R differentiated relatively young patients with Parkinson's disease from matched controls even though the groups were indistinguishable on the Logical Memory WMS-R test (Camicioli, Grossmann, et al., 2001). Squire and Shimamura (1986) found that the PAL discriminated very well between a group of amnesics of mixed etiology and persons with mildly depressed memory functioning due to either depression or chronic alcoholism. It also proved to be sensitive in documenting the more subtle differences between depressed patients and normal control subjects. TBI patients can be taught to improve their performance on this test using mental imagery instructions (Twum and Parente, 1994).

Paired associate learning variants. The paired associate learning format lends itself to a seemingly unlimited number of modifications—in length, difficulty level, number of trials, scoring methods, etc. (e.g., Delbecq-Dérouesné and Beauvois, 1989; Morrow, Robin, et al., 1992; see also H.S. Levin, 1986). Some examples are given here.

In an early study, Inglis (1959) randomized the order of administering just three word pairs (cabbage–pen, knife–chimney, sponge–trumpet), giving them until the subject reached a criterion of three consecutive correct responses for all three pairs or until 30 trials, dropping out word pairs once the criterion was reached. The score was the number of times a word pair had to be repeated. Not surprisingly, a control group's mean score of 13 ± 6.6 was significantly lower than that for elderly psychiatric patients (59 ± 25).

The first ten word pairs of Wechsler's Similarities test given to TBI patients were used to test incidental learning by asking for a free recall of the pairs, and then giving a cued recall trial using the first word of each pair as the cue (Vilkki, Holst, Öhman, et al., 1990). Both free and cued recall correlated significantly with duration of coma (−.48 free recall, −.43 cued recall) and ventricular enlargement (−.33, −.32 for free and cued recall, respectively). This technique was also highly sensitive to the presence of diffuse damage and to the left-lateralized damage after surgical repair of subarachnoid hemorrhage due to a ruptured aneurysm (Vilkki, Holst, Öhman, et al., 1989). A year after surgery, free recall, but not cued recall, discriminated between patients functioning normally and those with obvious neuropsychological deficits.

Rich and her colleagues (1997) examined the interference effect of using a different set of second words during a second learning task, following an AB–AC paradigm. Huntington patients were slow to learn the initial pairs relative to controls and were disproprotionately impaired in learning the new associations. However, they did not recall more B words during the AC portion of the test and, therefore, did not show increased suceptibility to proactive interference as was hypothesized.

Choosing among word-learning tests [mdl]

Many word list tasks are available today. The examiner's selection should depend on what test characteristics are most relevant to the examination questions, the patient's condition and demographic status, and the ease of administration and scoring.

For verbal learning per se, my preference for the AVLT rests on a number of test variables: Unlike the SR procedure, all subjects are exposed to the same number of stimuli, and since they are given in the same order, position effects (primacy, recency) become evident as well as other strategies the subject might use. The addition of both immediate and delayed recall trials and a recognition trial allows the examiner to see both the effects of interference and those of delay on recall; the recognition trial, of course, is the best measure of how much the subject has actually learned and the extent of recall efficiency. Both administration and scoring are much simpler than those of the SR, requiring no arithmetic operations, and the data are immediately available. In fact, I score as I give the test. Moreover, little seems to be gained (but much time lost) by the elaborate SR scoring procedures. Moreover, as Loring and Papanicolaou (1987) note, a number of SR measures "have typically . . . high correlations in both clinical and control samples (i.e., total recall, LTS, LTR, CLTR), suggesting that these measures are assessing similar constructs." These authors further note that the seeming parcellation into "long-term storage" and "retrieval" makes an arbitrary distinction between these terms, basing LTS on Buschke's definition requiring two consecutive trials and overlooking the possibility that erratic recall of a word may reflect tenuous storage rather than a retrieval problem. In fact, this test does not measure retrieval as understood in the usual sense of the efficiency of delayed recall compared with recognition tested immediately following delayed recall (e.g., see Delis, 1989; Loring and Papanicolaou, 1987).

In comparing the SR procedure with the AVLT and CVLT list learning administration (standard procedure), using 20 words for ten trials and college undergraduates as subjects, MacLeod (1985) discovered that with the SR procedure these subjects required only 30 to 40 item exposures to reach the 20-word criterion, compared to more than 100 single-item exposures for the standard procedure. However, criterion was reached one full trial sooner for an animal list by means of the standard procedure compared to the SR procedure, and the standard procedure also led to criterion an average of one-half trial sooner than the SR procedure on a random word list. In recommending the SR procedure as "faster to administer . . . because fewer items need be presented on each study trial," MacLeod did not reckon with the learning problems of neurologically impaired patients who may require many more word repetitions during 12 (often discouraging, certainly boring) trials in which half or more of the words must be repeated each time. MacLeod's work indicates that, overall, the SR and the standard procedure are about equally effective in measuring learning competency, at least in bright young people.

When examining the incidental use of concept formation (compared with the structured format of Similarities, e.g.), the subject's use of strategy in learning, and/or whether cueing helps (e.g., when focusing on a patient's potential to benefit from remediation training), the CVLT provides valuable information as it documents the benefits of prepackaged concepts for learning. However, because of the CVLT's built-in conceptual confounds, the AVLT is a better test for verbal rote memory in itself. The CVLT may also be used for a second examination to avoid practice effects on the AVLT, although CVLT produces slightly higher scores (Crossen and Weins, 1994).

Verbal Paired Associates (WMS-R) are particularly useful when the patient appears incapable of learning more than a very few words on a list test (administration of story recall early in the examination gives a general idea of the patient's level of verbal learning). With VePA, verbal learning can be examined by means of the hard pairs while the easy ones give the patient some success opportunities so that the test is not experienced as too defeating. Moreover, the built-in cues also help to determine whether the patient can benefit from cueing strategies for remediation. Although the WMS version with its greater weighting on hard pairs appears to work a little better as a measure of verbal learning, the availability of normative data, particularly for the delayed recall trial, facilitates WMS-R interpretation.

Story Recall

In many ways story recall tests most resemble everyday memory demands for the meaningful discourse found in conversation, radio and television, and written material. They provide a measure of both the amount of information that is retained when the material exceeds immediate memory span, and the contribution of meaning to retention and recall. Prior experience may prompt the story listener to attend to the setting, characters, actions, and outcome of a story. The comparison of a patient's memory span on a story recall test with a word list task will tell how much the inherent organization and meaningfulness of the prose material can facilitate memory or, conversely, how much syntactic processing or overload of data can compromise functioning.

Story recall administration presents a number of problems because not all examiners present the test material in exactly the same way. Ideally, the stories are enunciated carefully in a natural speech pattern with a slight pause between sentences for clarity. Presentation rates that are too fast hinder recall in intact persons (Shum, Murray, and Eadie, 1997), an effect likely to be greatest in the elderly and patients whose brain disorder has slowed their processing of information. Also, asking patients "Anything else?" at the end of recall allows them an opportunity to provide information out of order that might have come to mind during or after the recall process. Some patients will spontaneously provide this additional recall while others will not [dbh].

Scoring issues. Scoring story recall presents a number of problems since few people repeat the test material exactly. This leaves the examiner with the problem of deciding how much alterations differ from the text to require loss of score points. Common alterations include a variety of substitutions (of synonyms, of similar concepts, of less precise language, of different numbers or proper names); omissions (large and small; irrelevant to the story, relevant, or crucial); additions and elaborations (ranging from inconsequential ones to those that distort or alter the story or are frankly bizarre); and shifts in the story's sequence (that may or may not alter its meaning).

Rapaport and his colleagues (1968) addressed questions of how to judge these alterations by scoring as correct all segments of the story in which "the change does not alter the general meaning of the story or its details." Without a more elaborate scoring scheme, this rule is probably the most reasonable one that can be followed in a clinical setting. Talland and Ekdahl (1959) made a welcome distinction between verbatim and content (semantic) recall of paragraphs. They divided meaningful verbal material into separate scoring units for verbatim recall and for content ideas, which are credited as correctly recalled if the subject substitutes synonyms or suitable phrases for the exact wording (see p. 449). Several scoring methods have been devised for the Logical Memory (LM) test that take minor alterations and/or gist into account (see p. 445). These may be generally applicable to tests of story recall. Unless scoring rules for alterations are specified or a method for scoring slight alterations is used, the examiner will inevitably have to make scoring decisions without concise, objective standards. In most cases, the likelihood that a score for a story recall test may vary a few points (depending on who does the scoring and how the scorer feels that day) is not of great consequence. The sophisticated psychological examiner knows that there is a margin of error for any given score. However, alterations in some patients' responses may make large segments unscorable as verbatim recall, although the patient demonstrated a quite richly detailed recall of the story. Other patients may reproduce much material verbatim, but in such a disconnected manner, or so linked or elaborated with bizarre, confabulated, or perseverated introjections that a fairly high verbatim recall score belies their inability to reproduce newly heard verbal material accurately.

Logical Memory (LM-O, LM-R) (Wechsler, 1945, 1987, 1997)

Free recall immediately following auditory presentation characterizes most story memory tests. The original Wechsler Memory Scale version, Logical Memory (LM-O) employs this format. The examiner reads two stories, stopping after each reading for an immediate free recall. The LM manuals do not specify the speed of presentation of the stories, which may vary considerably across examiners (Shum, Murray, and Eadie, 1997). The format for the WMS-R Logical Memory test (LM-R) differs in a number of ways from the WMS-O. Most important for the usefulness of the test is the addition of a 30-minute delayed recall of the stories. The Anna Thompson story has remained the first story in all versions with only minor variations in each subsequent edition. The second story was changed for the 1987 revision. In WMS-III, not only is another new paragraph paired with the venerable Anna Thompson (the WMS-R Robert Miller story was considered too likely to evoke an emotional reaction in some people and thus bias recall), but it is given in two learning trials which increases the likelihood of retention over a 30-minute delay. The second reading may aid patients who are so overwhelmed by the amount of information contained in the story that they lose track of what they are hearing. Repeating the first story rather than the second may have better addressed the problem of anxious patients "freezing" at the beginning of the test (Cannon, 1999). Delayed recall may be prompted with a set of cues provided for each story. Two alternate paragraphs of equivalent difficulty to LM-R have been developed for use when repeat testing is required (J. Morris et al., 1997).

Scoring. Scoring of the stories requires the examiner's judgment. Score differences due to variability both in subjective criteria and in scoring methods produced a variegated set of LM-O average performances by normal young and middle-aged subjects (Loring and Papanicolaou, 1987; see also Mitrushina, Boone, and D'Elia, 1999). Scoring guidelines introduced with the LM-R improved interrater reliability with coefficients above .95 (K. Sullivan, 1996; Wechsler, 1987). The manuals for both LM-R and LM-III provide a general rule—based on "item(s) correctly repeated," for scoring each of the 25 items of a story and examples of both satisfactory and failed responses. However, the size, complexity, and scoring criteria of individual items differ considerably: several items consist of just one name with no variations credited, other one-name items allow several variations; some words have to be precisely included (e.g., cafeteria), while others may be indicated by similar expressions (e.g., "cops" is an acceptable substitute for "police"); some words can be scored as correct even if they occur in an incorrect context (e.g., South). These scoring anomalies suggest that two persons with similar recall abilities may earn quite different scores if one hit on the items calling for a single word response and the other recalled the same amount of material or even more but did not give many of the specified person and place names. For these reasons—and to capture distortions or confabulations—recall should be recorded verbatim. Unfortunately the WMS record forms do not leave space for recording subjects' responses on them.

A detailed scoring system proposed for LM-O is designed to bring out qualitative response differences that classify response segments (single ideas) according to whether they are essential propositions (e.g., that a robbery took place), detail propositions (e.g., the protagonist's name was Anna Thompson), or self-generated propositions (i.e., intrusions) (Webster et al., 1992). These authors also developed a cued recall format for the Anna Thompson story which contains 12 questions, each open-ended and followed by a choice of three answers. A recognition task consisting of multiple-choice questions developed for LM-R showed a ceiling effect for well-educated subjects (Fastenau, 1996b). A thematic scoring option has also been added for both LM-III stories, designed to indicate the number of main ideas recalled.

The score for the additional learning provided by the second presentation of Story B is referred to as a "learning slope;" the manual provides comparison data from the normative sample. In some cases this score may be critical in the interpretation of overall performance.

An 85-year-old man without memory complaints in a study of healthy aging recalled seven elements each from LM-III A and the first administration of Story B. Following the second presentation of Story B his recall doubled, showing the advantage of giving a second trial, perhaps because of age-related slow information processing. His benefit from the second trial held through the delay interval. He retained ten elements of Story B while recalling only four from Story A. However, following WMS-III scoring rules, combining his delayed recall score of Story A with that of Story B ($\Sigma = 21$) placed him only in the *average* range for his age thus failing to show his *above average* ability to retain well-learned information over time.

TABLE 11.9 WMS-III Logical Memory Recognition Scores as a Function of Age or LM-II Scores in an Elderly Sample

			RECOGNITION SCORES		
Age (years)	*Sample* (n)	*LM-II (mean)*	*Recognition (mean)*	*Range*	*25th Percentile*
70–79	26	31.6 ± 8.3	27.1 ± 2.3	21–30	25
80–89	70	27.3 ± 7.0	25.7 ± 2.7	16–30	24
90–99	30	21.2 ± 8.8	23.1 ± 3.2	16–29	20
LM-II Range					
All	7	0–9	19.1 ± 2.6	16–23	16
All	16	10–19	23.1 ± 2.4	18–26	21
All	56	20–29	24.9 ± 2.5	18–30	23
All	47	30–39	27.4 ± 2.5	24–39	26
All	6	40–45	29.9 ± 0.9	28–30	28

From the Oregon Brain Aging Study.

A formula is provided for calculating percent retention of the LM-III stories over the delay interval. The manual gives no normative story content data for the LM-III yes/no recognition test which follows the 30-minute delayed recall of both stories. Rather, the recognition score from this test is added to the Verbal Paired Associates recognition score to produce a composite recognition score, again leaving interpretation of the data to the examiner's imagination. To correct this deficiency, at least for a group of healthy, well-educated (M = 14.7 years) subjects in a longitudinal study of aging (Hickman et al., 2000), recognition scores are given in Table 11.9. This sample (61 men, 71 women) had a mean age of 84.8. No sex effect appeared on the delayed recall (LM-II) or the Recognition trial. Recognition scores ranged from chance (16) to perfect in a negatively skewed distribution. They correlated with age ($r = -.46$), but the correlation with LM-II scores ($r = .73$) was stronger.

Test characteristics. Immediate recall of both earlier LM versions remains fairly stable through middle age and then progressively declines (Mitrushina, Boone, and D'Elia, 1999; Sinnett and Holen, 1999; Wechsler, 1987). The LM-III immediate recall shows a slow, steady decline between the ages of 55 and 89 years with the oldest age group (85–89 years) recalling about half of the recall of the youngest normative group (Wechsler, 1997b). Delayed recall data vary for LM, perhaps in part because of administration and test differences. Delayed recall on LM-R may begin its decline as early as the 20s, level off until the 50s, and then continue to shrink (Wechsler, 1987). However, the omission of data for years 18–19, 25–34, and 45–54 makes this a tenuous generalization. Norms have been developed for LM-R for older age groups (Ivnik, Malec, Smith, et al., 1992c; Lichtenberg and Christensen, 1992; Marcopulos, McLain, and Giuliano, 1997; E.D. Richardson and Marottoli, 1996). Delayed recall on LM-III begins to decline fairly steadily from about age 45. Age decline on LM-III delayed recall is largely explained by poorer immediate recall (Haaland, Price, and LaRue, 2003). A steady decline in recall of thematic units also occurs with age. The relatively lower education of the older groups in the WMS-III normative population makes these norms questionable when evaluating the performances of better educated older persons.

Sex effects are not prominent. Overall, women have the advantage. They outperformed men on immediate recall of LM-R (Ragland et al., 2000). Ivison (1986) found slightly higher scores by women on "Anna Thompson," slightly higher scores by men on the second LM-O story, perhaps reflecting the stories' different content. Women with greater temporal lobe cerebral blood flow performed better on immediate and delayed recall of the LM-R than those with lower blood flow, but this correlation was not found in males (Ragland et al., 2000). Education, often used as the most convenient measure of intellectual ability, makes a significant contribution to performance on LM (Abikoff et al., 1987; Compton et al., 1997; E.D. Richardson and Marottoli, 1996; Ylikoski et al., 1998), as does socioeconomic status (Sinnett and Holen, 1999).

Correlations between LM-R stories A and B for different age groups were in the .68 to .80 range for immediate recall and .68 to .85 for the delay trial (Wechsler, 1987). The WMS manual does not give any information about the relationship between the two LM stories. For LM-R the test manual reports that subjects in the 20–24 year age group made the greatest gains when retested within four to six weeks: +7.4 on immediate recall of the two stories, +9.4 on delayed recall (out of a possible 50 points) (Wechsler, 1987; see also McCaffrey, Duff, and Westervelt, 2000b, for other control group gains). Two raters scoring LM-R blindly achieved almost perfect agreement (Woloszyn et al., 1993). The retest gain on LM-III over 2 to 12 week intervals was reported to be about 2 points for immediate and delayed recall when the age groups were combined (The Psychological Corporation, 1997). Practice effects can be observed with lengthy retest intervals, even up to a year (Hickman et al., 2000; Theisen et al., 1998).

Correlational studies consistently demonstrate a relationship between the immediate recall trial of this test and other learning tests (Kear-Colwell, 1973; Macartney-Filgate and Vriezen, 1988), and an even stronger association of delayed recall with other learning tests (Bornstein and Chelune, 1989; Woodard, Goldstein, et al., 1999). This latter group described LM-R as the "purest" measure of episodic memory compared to a word list learning task and a visuospatial memory task because of its relatively low association with nonmemory measures. Both immediate and delayed trials have larger associations with verbal tests (e.g., WIS Information, Vocabulary) than does the associate learning format, probably reflecting the verbal organization and syntax required both for repeating the stories and giving answers to these two WIS tests (Larrabee, Kane, et al., 1985).

Neuropsychological findings. Because of its age and popularity, a wealth of clinical studies have used LM. Thus LM data exist for almost all known brain disorders. This review covers LM patterns for the most commonly seen or neuropsychologically relevant conditions.

A delayed LM-O recall distinguished a group of mild TBI patients with apparent "good recovery" from control subjects whose average recall score was $2^1/_2$ units greater than that of the patients (Stuss, Ely, et al., 1985). In another study, TBI patients recalled less of the Anna Thompson story than controls, particularly losing details in the middle portion of the story while showing relatively well-preserved primacy and recency effects (S. Hall and Bornstein, 1991). Soccer (European football) concussions in long-term adult players are associated with impaired LM-R performance (Matser, Kessels, Lezak, et al., 1999). LM-R was more accurate than a word list learning task and a paired associate learning task in differentiating patients with mild head injuries from matched controls (Guilmette and Rasile, 1995). However, not all studies have found LM to be sensitive to mild TBI. Brooker's (1997) review identifies other WMS-R tests as more sensitive to the effects of mild TBI and mild dementia in group comparisons, apparently because of LM's large within-group variability. For example, the LM-O version of this test did not distinguish moderately injured from severely injured head trauma patients prior to entrance into a rehabilitation program, although the moderately injured group's average score (5.85) was barely *marginal to normal limits,* while that of the severely damaged group was *defective* (4.25) (Trexler and Zappala, 1988). Yet, significant improvement in the first year after head injury was registered by LM-O, which also distinguished the head injured patients from their controls even after showing improvement at two years posttrauma (Dikmen, Machamer, et al., 1990); moreover, the LM-R score contributed to the prediction of improvement and level of social integration of TBI patients six months after discharge from acute rehabilitation (Hanks, Rapport, et al., 1999).

A "percent forgetting" score reflecting the difference between immediate and delayed recall showed that right temporal lobectomy patients were more likely than those with temporal excisions on the left to have improved verbal memory, and less likely to perform worse than preoperatively (Ivnik, Sharbrough, and Laws, 1988). Delaney, Rosen, and their colleagues (1980) found that only delayed recall differentiated right from left temporal lobectomy patients. Similar findings showed the expected right–left differential in recall score levels for patients with seizure foci who subsequently had temporal lobectomies, but a "percent retained" score was the only one that correlated significantly with neuronal loss in the excised tissue (Sass et al., 1992). The volume of the left hippocampus significantly predicted LM-R immediate, delayed, and percent retention scores in seizure patients who had not undergone surgery (R.C. Martin, Hugg, Roth, et al., 1999). Groups of patients with lateralized lesions of mixed etiologies also performed differently on LM-R, of course with the patients whose damage was on the right outperforming the left lesioned group (Chelune and Bornstein, 1988; P.M. Moore and Baker, 1996). Patients with carotid artery disease performed significantly better than Alzheimer patients but significantly worse than control subjects on LM; no differences showed up between the two groups with lateralized carotid involvement (Kelly, Kaszniak, and Garron, 1986). A scoring system that distinguishes between "Essential," "Detail," and "Self-generated" propositions brought out response differences between patients with lateralized lesions and normal control subjects (Webster et al., 1992). For example, normal control subjects gave more essential and detail propositions than did the patients, patients with left-sided lesions tended to make fewer responses in all categories, and patients with lesions in the right hemisphere gave more intrusion responses.

Like other learning tests, LM has been useful as an aid both in identifying dementia and in tracking its progression (Storandt, Botwinick, and Danziger, 1986; R.S. Wilson and Kaszniak, 1986). Scores below those of controls have been found in patients before the appearance of clinical evidence of Alzheimer's disease (Howieson, Dame, et al., 1997; Rubin, Storandt, Miller, et al., 1998) or in asymptomatic Huntington's disease gene carriers (Hahn-Barma et al., 1998). Characteristically, Alzheimer patients have poor recall after the delay interval. The Savings Score developed by Nelson Butters, Salmon, Cullum, and their coworkers (1988; Tröster, Butters, Salmon et al., 1993) (see p. 442) shows that Alzheimer patients have more pronounced forgetting over the delay interval than Huntington patients). This test is also sensitive to the memory and learning deficits of multiple sclerosis (Minden, Moes, et al., 1990). MS patients show the usual pattern of recalling main elements compared to nonessential details but recall less than controls (Lokken et al., 1999).

Babcock Story Recall Format (Babcock, 1930; Babcock and Levy, 1940)

Like Logical Memory in the Wechsler tests, the modified format uses a pair of stories of similar length and difficulty.[1] After initial reading and recall of the first story, the first story is reread and one or two tests are interpolated for approximately 10 min when a recall is requested. Immediate thereafter, a second story is read and its administration follows the Babcock format of immediate recall upon first hearing, then rereading, and another approximately 10 min interference period, and then delayed recall of the second story. One set of data on normal subjects found an approximately 4-point gain on second recall of 21-item stories (Rapaport et al., 1968). Freides, Engen, and their colleagues (1996) used the double reading format on two 29-unit stories and the original Babcock-Levy[1] story divided into 22 units; recall after the second reading immediately followed a drawing recall trial so that delay was only from 80 to 120 sec. College students' average unit gain was from 5 (story 1 [Babcock-Levy], first reading $M = 12.8 \pm 3.5$) to 6.6 (story 2, first reading $M = 16.5 \pm 5$). Given only the Babcock-Levy story, elderly subjects (in 60s, 70s, 80s) on average made 5 unit gains from first reading: $M = 10.5 \pm 3.4$ (60s), $M = 9.0 \pm 3.3$ (70s), $M = 8.0 \pm 3.6$ (80s) (see p. 449 for second story and Table 11.10).

It is noteworthy that LM-III now includes the Babcock format for the second of its two stories.

When using story pairs, the decision about which story recall format to use, one without rereading after the first recall or Babcock's, depends on whether the examiner is more interested in testing for proactive interference or learning. The stories in each of these tests can be adapted to either format. The Babcock format may be more likely to elicit interference effects because it was read twice and the second story is introduced immediately after the delayed recall of the first. The two readings in the Babcock format seem to make more neuropsychological sense than a single reading of a story, as patients with a limited auditory span, or whose grasp of information as it goes by them is restricted by slow processing, will register only a small portion of the story on first hearing it. Immediate recall provides an appropriate opportunity for documenting these problems which then can be distinguished from defective learning by rereading the story. Delayed recall will then give a clearer picture of learning capacity. By the same token, patients whose delayed recall drops significantly even with a second reading leave little doubt about the fragility of their recall capacity. Of special interest are intrusions of content or ideas from the first to the second paragraph and wide disparities in amount of recall.

Story Sets

Story recall elicits the most information about a subject's ability to handle meaningful verbal information when two stories are given in tandem. Since neuropsychological examinations are often repeated, sometimes within weeks or even days, the best way to deal with practice effects is to have multiple story sets available.

Each of the four forms of the Rivermead Behavioural Memory Test (B. [A.] Wilson, Cockburn, and Baddeley, 1985) contains a 21-unit (from 54 to 65 words in each) story suitable for tandem presentations. The authors acknowledge the local nature of some place names and colloquialisms in the stories, advising examiners to substitute more familiar ones as needed (e.g., I substitute "Beaverton," a Portland suburb, for "Brighton" [mdl]).

P. Green and Kramar (1983)[2] developed the *CogniSyst Story Recall Test,* a series of stories at five levels of difficulty (from 22 words, 10 items to 56 words, 25 items) with six stories in each set.

Two/ semi-trailer trucks/ lay on their sides/ after a tornado/ blew/ a dozen trucks/ off the highway/ in West Springfield./ One person/ was killed/ and 418 others/ were injured/ in the

[1]December 6./ Last week/ a river overflowed/ in a small town/ ten miles/ from Albany./ Water/ covered the streets/ and entered the houses./ Fourteen persons/ were drowned/ and 600 persons/ caught cold/ because of the dampness/ and cold weather./ In saving/ a boy/ who was caught under a bridge,/ a man/ cut his hands.

[2]The complete test sets and norms can be ordered from Wm. Paul Green, Ph.D., 17107-107 Ave., #201, Edmonton, Alberta, T5S 1G3, Canada. (e-mail: paulgreen©shaw.ca)

TABLE 11.10 Expected Scores for Immediate and Delayed Recall Trials of the Babcock Story Recall Test

*Mental Ability Level**	*Sample* (n)	IMMEDIATE RECALL			DELAYED RECALL		
		Q_1	*Median*	Q_3	Q_1	*Median*	Q_3
Average	27	12	13	14	13	15	16
High average	41	12	14.5	17	16	17	19
Superior	45	13	15	18	15	17	19

*For statistical definitions of these levels, see Chapter 6.
Adapted from Rapaport et al. (1968).

Wednesday storm/ which hit an airport/ and a nearby residential area./ The governor/ will ask/ the President/ to declare/ the town/ a major disaster area.

Cowboy Story

Because it has been included in many mental status examinations since it first appeared in 1919, this is the paragraph best known to medical practitioners. Talland (1965a; with Ekdahl, 1959) used it to make a welcome distinction between verbatim and content recall of paragraph. He divided it into 27 memory units for quantitative verbatim recall and identified 24 content ideas (italicized words or phrases), which are credited as correctly recalled if the subject substitutes synonyms or suitable phrases for the exact wording.

A cowboy/ from Arizona/ went to San Francisco/ with his dog,/ which he left/ at a friend's/ while he purchased/ a new suit of clothes./ Dressed finely,/ he went back/ to the dog,/ whistled to him,/ called him by name/ and patted him./ But the dog would have nothing to do with him,/ in his new hat/ and coat,/ but gave a mournful/ howl./ Coaxing was of no effect/; so the cowboy went away/ and donned his old garments,/ whereupon the dog/ immediately/ showed his wild joy/ on seeing his master/ as he thought he ought to be./ (Talland, 1965a).

On immediate recall testing, a 22-subject control group gave an average of 8.32 of the 27 verbatim memory units; their average content recall score was 9.56. Healthy subjects in three age groups ($M_{ages} = 43.5$, 55.9, and 67.5) read this story aloud (Fastenau, Denburg, and Abeles, 1996). The youngest group recalled 15.1 units immediately and lost an average of less than one unit after a 20-minute delay. The oldest group recalled less, 13.5 units, but retained this level of recall over delay. Age accounted for about 10% of the variance in these data.

Story Memory Test (Heaton, Grant, and Matthews, 1991)

This story recall test is unique in its multiple presentations and normative data that includes a 4-hour delay. The 29-item story authored by Ralph Reitan is presented for up to five trials or until the subject has obtained at least 15 points, whichever comes first. The procedure is advantageous for patients with slow information processing or attentional deficits who may not have sufficient exposure when material is presented only once. A tape recording of the story presents items at the rate of one scorable unit per second. Patients with attentional or hearing problems might benefit from a "live" presentation. Recall units are scored so that partially correct information receives partial credit. The *Learning* score is the number of points recalled on the last learning trial divided by the number of trials taken to reach criterion. The *Memory* score is a percent of loss over time: percentage of the difference between the amounts recalled on the last learning trial and on the 4-hour recall. Age- and education-corrected norms are presented in the manual.

African Americans do not perform as well as Caucasians on this test, which has been attributed in part to differences in dialect (Manly, Miller, Heaton, et al., 1998). In this study the use of Black English affected the Learning score because different word usage by African Americans resulted in loss of points. Memory was not affected by the use of Black English because it is scored as a percent loss. A factor analysis of immediate memory (trial 1) showed loading with CVLT trial 1 while the Learning score loaded with CVLT learning (trials 1–5); verbal fluency contributed to both of these scores (DiPino, Kabat, and Kane, 2000). Delayed recall loaded positively with CVLT delayed recall and negatively with digits backward and Judgment of Line Orientation.

Stories in memory batteries

Memory batteries frequently include story recall tests. The *Memory Assessment Scales* (J.M. Williams, 1991; see pp. 488–490) contains a short story that is read once to the patient. Immediate and delayed recalls are obtained, although only responses to nine questions about the stories are scored. This story would be useful for severely impaired patients who might fail other memory tests as the cueing provided by questions about the stories facilitates recall. The *Denman Neuropsychology Memory Scale* (Denman, 1984, 1987; see pp. 490–491), on the other hand, offers a 42-item story, which is longer than most tests in order to avoid the ceiling effects that can occur with shorter stories. The *Learning and Memory Battery* (*LAMB*) (Tombaugh and Schmidt, 1992; see pp. 493–494) contains a 31-item paragraph of information about a person which is read twice with free and cued recall trials following each reading. Delayed recall takes place after 20 minutes and includes free and cued recall as well as multiple-choice questions regarding missed material. The *Randt Memory Test* (Randt and Brown, 1986; Randt, Brown, and Osborne, 1980; see pp. 487–488) contains five 25-word, 20-item stories, which could be used in pairs. All five stories follow an identical formula in identical sequence: date (3 items), place (2 items), catastrophe (3 items), locale (4 items), consequence that includes three numbers (8 items). Erickson and Howieson (1986) suggested that they read more like a list than a story. Since the strict similarity of the items could lead to confusion for even the most efficient learner, these stories should be used together—or even a day apart—only with caution.

VISUAL MEMORY

Tests of visual memory often call for a visuomotor response, typically drawing. This, of course, can complicate the interpretation of defective performance since failure may arise from constructional disability, impaired visual or spatial memory, or an interaction between these or other factors. Even on recognition tasks without a constructional response, perceptual impairments such as hemispatial inattention are potential performance confounds. Therefore, the quality of a patient's responses when compared to other neuropsychological measures should enable the examiner to estimate the relative contributions of perception, constructional or visuomotor skill, and memory to the final product.

To minimize verbal mediation, visual memory test stimuli often use abstract designs or nonsense figures, although some visual memory tests (e.g., Continuous Recognition Memory Test; WMS-III Family Pictures) contain both visual and verbal elements and do not strive to assess material-specific memory function. Even attempts to create a hypothetically "pure" nonverbal visual memory test by using complex or unfamiliar stimuli cannot fully eliminate verbal associations—which are thought to contribute to the poorer lateralizing ability of most visual memory tests compared to their verbal counterparts (Barr et al., 1997; Feher and Martin, 1992).

The measurement of learning (rate, efficiency, retention) requires material of sufficient difficulty that only very exceptional persons would be able to grasp and retain it with one or two exposures, and there must be enough learning trials to permit emergence of a learning curve. A number of visual learning tests meet these requirements—some do not. Several more or less follow André Rey's AVLT paradigm.

Facial recognition is another form of visual memory. The Recognition Memory for Faces portion of the Recognition Memory Test (Warrington, 1984) and the Wechsler Memory Scale-III Faces test (Wechsler, 1997b) use this format. The Warrington test is discussed in Chapter 12 and the WMS-III Faces is discussed below.

Visual Recognition Memory

Recognition testing is important for evaluating visual memory when free recall is impaired. It also overcomes the output limitations of patients who cannot adequately draw due to hemiparesis or some other physical limitation. Most newer visual memory test formats include a recognition component. In this section, only visual memory tests that solely rely on recognition testing will be discussed.

Recurring Figures Test (D. Kimura, 1963)[1]

This test is important, in part, because it established the recurring stimulus paradigm for visual memory assessment. The material consists of 160 cards containing either geometric or irregular nonsense figures which are each shown for 3 sec in eight trial blocks. The trial sets are not apparent to the subject as the stimuli are continuously presented. Eight of the first 20 designs repeat in each subsequent trial block. The subject indicates which designs were seen previously. Because the targets are "new" during the first trial block, correct recognition of all targets yields a score of 56. False positive responses are subtracted from the total to correct for guessing. Normative information is based upon corrected recognition performance.

Test characteristics. Performance appears unaffected by age (Rixecker and Hartje, 1980): Kimura's control subjects, all in their 20s, had an average corrected score

[1]The test material may be ordered from D.K. Consultants, Department of Psychology, University of Western Ontario, London, Ontario, Canada, N6A 3K7.

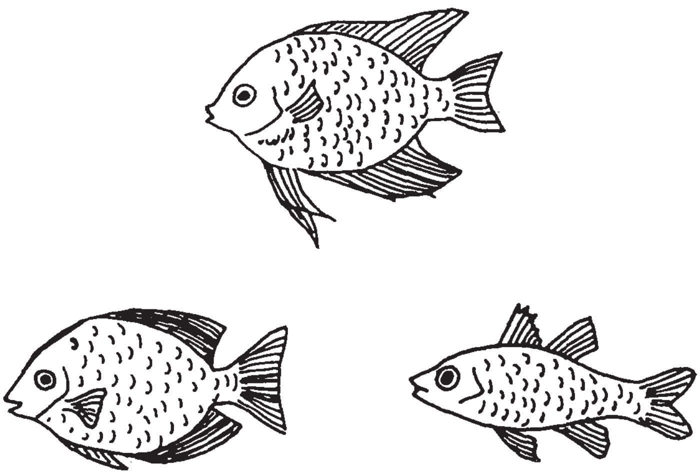

FIGURE 11.1 One of the eight target figures (top) of the Continuous Recognition Memory Test and two foils. (Courtesy of H. Julia Hannay)

of 38.9 which is similar to the performance of an independent sample of healthy young subjects ($M = 37.5$) (Aguirre et al., 1985) and other, older, groups in their 30s ($M = 40.9$) and 40s ($M = 36.6$) (Nielsen, Knudsen, and Daugbjerg, 1989). One exception ($M = 28.5$) was reported for control subjects of whom most were in their forties (Newcombe, 1969). For subjects in the 20 to 65 year age range, performance was unaffected by age but education differences were found (Rixecker and Hartje, 1980). A reliability coefficient of .94 was reported.

Neuropsychological findings. Kimura (1963) found no difference between right and left temporal lobectomy patients. Right temporal lobe patients made more false positive errors, however, producing a significant left vs. right group difference for education corrected scores. Both groups remembered geometric better than nonsense figures but left temporal patients recognized a greater proportion of nonsense figures.

Corrected scores differentiated TBI patients from healthy control subjects, although no difference in number of false positive errors was observed (D.N. Brooks, 1974). Patients with histories of alcoholism obtained an average score of 32.7, in sharp contrast to the *defective* scores ($M = 11.3$) made by six Korsakoff patients (Squire and Shimamura, 1986). Noting some extreme instances of intragroup variability (range = 5 to 24 for Korsakoff patients, which overlaps the control subjects' range of 20 to 40), Squire and Shimamura suggested that, "this test may fail to detect amnesia in some cases because it does not incorporate a long delay, and because the critical items are repeated several times" (p. 873). This criticism can be applied to all recurring recognition tests that do not incorporate a delayed recall component.

Continuous Recognition Memory Test (CRMT) (Hannay and Levin, no date; Hannay, Levin, and Grossman, 1979)[1]

This test consists of 120 line drawings of various flora (e.g., mushrooms, flowers, a cauliflower, a potato) and fauna (e.g., fish, seashells, dogs) organized into six blocks of 20 drawings each (see Fig. 11.1). The first set of blocks introduces the eight target drawings plus 12 foils. Each of the subsequent blocks contains all eight target figures plus eight similar ones (in each of the eight categories from which target figures are drawn), plus four drawings from other-than-target categories (e.g., vegetables). The subject sees each drawing for 3 sec and, similar to Kimura's Recurring Figures, indicates whether the drawing is "old," i.e., previously seen, or "new." The original format of the test includes a set of drawings with each target figure on top of a page and repeated again with the similar foils randomized below to test perceptual accuracy. This format can be readily turned into a recognition trial simply by covering the target figures on top of each page with a 3 × 5 card and asking the subject to point out the drawing which is exactly like the one seen before, "six different times."

This test extends Kimura's scoring system to include signal detection techniques. Scores are obtained for *Hits*

[1]This test may be ordered from H.J. Hannay, 4046 Grenoch Lane, Houston, TX 77025.

(correct recognition of recurring targets), *False Alarms* (incorrectly calling a new, nonrecurring figure "old"), and *Misses* (failure to recognize a target stimulus). The *Correct Responses* score is calculated from the formula Hits + (60 − False Alarms). A d' score can also be calculated as a measure of perceptual discrimination. Impairment levels have been determined for each of these scores: Correct Responses < 87; Hits < 36; False Alarms > 16; Misses > 4 (Hannay, Levin, and Grossman, 1979). The test format is sufficiently similar to the Continuous Visual Memory Test described below that either can follow the other for retesting.

The application of signal detection methods to recognition memory testing is not universally accepted (J.T.E. Richardson, 1994). Drake and Hannay (1992) explained that signal detection analysis is justifiable for evaluating scores when it meets the necessary assumptions of signal detection theory (i.e., a constant sensory input [signal and noise], random presence of signal that occurs with a fixed signal strength, with signal and noise at times being indistinguishable).

Test characteristics. For 66 TBI patients, neither age nor education was associated with scores on this test (Hannay, Levin, and Grossman, 1979). However, both age and education effects, but not sex differences, were observed in a large ($n = 299$) sample of healthy subjects in the 10 to 89 year age range (Trahan, Larrabee, and Levin, 1986). Hits and False Alarms tap into different aspects of memory performance, with Hits associated with learning and memory, and False Alarms reflecting attention to visual detail (Fuchs et al., 1999).

Larrabee and Curtiss (1995) modified this test by including a 30 min delayed recognition trial. Unlike the Visual Reproduction test of the Wechsler Memory Scale, spatial ability did not contribute to CRMT performance. Somewhat surprisingly, the relationship to the Wechsler Memory Scale "general memory factor" was similar for both immediate acquisition and delayed recognition.

Neuropsychological findings. While this test did not discriminate between healthy subjects and patients with mild TBI, Correct Response did identify from 67% to 85% of moderately and severely injured patients from a broad age range (Hannay, Levin, and Grossman, 1979). Relatively fewer brain injured adolescents with moderate (29.4%) to severe (41.5%) injuries performed at *defective* levels (Hannay and Levin, 1989). Correct Response scores of patients with brain injury, mostly due to trauma, tended to be negatively associated with ventricular enlargement (H.S. Levin, Meyers, et al., 1981). Adolescents with left hemisphere contusions or hematoma or with diffuse damage performed at a somewhat lower level than those with identifiable right-sided or bilateral lesions (Hannay and Levin, 1989).

Continuous Visual Memory Test (CVMT) (Trahan and Larrabee, 1988)

With the Hannay and Levin CRMT format as a model, this test differs only in the abstract nature of the stimulus designs and in such details as number of items (112), number of target figures (7), number of times each target figure appears (7), and exposure time (2 sec). Scoring follows the CRMT procedures. Besides a trial for perceptual accuracy, the CVMT includes a recognition trial after a 30-minute delay. Normative data are available for ages 18 to 70+ (Trahan and Larrabee, 1988; Trahan, Larrabee, and Quintana, 1990). Cut-off scores for *Total* (score), a d' score (the perceptual discrimination measure calculated from z-scores for *Hits* and *False Alarms* listed in the record form), and *Delay* have been calculated for four age groups, 18–29, 30–49, 50–60, and 70+, and are presented with the normative data in the test booklet. However, several studies in independent samples suggest that the recommended cut-off scores tend to misclassify some healthy elderly subjects as impaired (S. Hall, Pinkston, et al., 1996; Paolo, Tröster, and Ryan, 1998a).

Test characteristics. Performance levels go down slowly but steadily from age 30 on, mostly due to an increase in false alarms (Trahan, Larrabee, and Quintana, 1990). Because of these age declines, the suggested cut-off scores for younger subjects are inappropriate for elderly application (S. Hall, Pinkston, et al., 1996). A comparison between subjects with 12 or fewer years of education and those with 16 or more years found no differences between groups (Trahan and Larrabee, 1988). Interitem reliability correlations go from .80 to .98 (for both recurring and nonrecurring items) (Trahan and Larrabee, 1988). Trahan and Larrabee (1988) reported a strong association between Total score and the WMS-R Visual Reproduction test's delay trial while finding no association between Delay and Block Design. These and other congruent data indicate that Delay is a measure of visual memory "relatively independent of visual–spatial ability" (Trahan and Larrabee, 1988). In their factor analytic studies, d' was associated with "a general cognitive factor" but no memory factors. The delayed recognition score has been reported to be the best measure of visual memory in some factor analytic studies (Larrabee, Trahan, and Curtiss, 1992) but not in others (Larrabee and Curtiss, 1995). However, like the Continuous Recognition Memory Test, spatial ability contributes little when compared to Wechsler's Visual Reproduction test (Larrabee and Curtis, 1995).

Test–retest stability coefficients in 12 healthy subjects reported in the manual are .85 for Total score, .80 for *d'*, and .76 for Delayed Recognition, although somewhat lower scores have been reported for larger samples (.53 to .66) (Trahan, Larrabee, Fritzsche, and Curtiss, 1996). As with other memory tests, stability coefficients after a one year retest delay were substantially lower (.44 to .49) in one healthy elderly sample (Paolo, Tröster, and Ryan, 1998b). Based upon these test–retest reliability data, an 11-point difference in Total score between assessments reflects a significant change ($p = .10$) that exceeds the standard error of the difference score and is not likely due to chance (Paolo, Tröster, and Ryan, et al., 1998b). An alternative form of the CVMT is available for repeat testing (Trahan, Larrabee, Fritzsche, and Curtiss, 1996).

Neuropsychological findings. The average scores for both right- and left-lateralized stroke patient groups were significantly lower than those for control subjects on all measured variables (Trahan, Larrabee, and Quintana, 1990). However, while 50% of patients with right-sided lesions failed on Total and 63% performed in the impaired range on Delay, of the patients with left-sided strokes only 20% and 23% failed on these measures, respectively. In a sample of patients with lateralized temporal lobe epilepsy, the CVMT did not discriminate seizure onset laterality, although overall cognitive functional and visuoperceptual processing were related to CVMT scores (Snitz et al., 1996). Almost all (92%) of a small group of Alzheimer patients had difficulty discriminating targets from false alarms. However, only about half had Total or Delay scores below the acceptable level (Trahan and Larrabee, 1988).

WMS-III Faces (Wechsler, 1997b)

Memory for faces has a rich tradition in memory assessment (Warrington, 1984), and in particular, for assessing memory functions associated with the nondominant hemisphere (B. Milner, 1968). This test of facial recognition memory is similar to Warrington's Recognition Memory subtest. A series of 24 faces is shown at the rate of one every 2 sec. Memory is assessed with a recognition format in which the target face pictures are shown one-by-one interspersed among 24 foils. The subject's task is to indicate which faces had been previously seen. Delayed recognition is tested with the 24 target faces mixed in with 24 new foils. Three scores can be obtained: *Recognition Total* ("yes" for targets, "no" for foils) on the immediate and on the delay trials, and *percent recalled.* Scores are converted to standard scores ($M = 10 \pm 3$) for each age group (see p. 142). As with many other WMS-III tests, the percent recalled score can be compared to normative tables as a supplemental score.

Test characteristics. The following data come from the Psychological Corporation's (1997) statistical analyses of this test. Performance for both the immediate and delayed components is fairly stable through young adulthood. It begins to decline in middle age and decreases more rapidly in the 70s and beyond. Percent recall shows little age effect as the average recall (i.e., with scaled score = 10) in the oldest age group (85–89 years) corresponds to a retention of 92 to 94%. The average reliability coefficient was .74 for both the immediate and delayed conditions. The test–retest stability coefficient over a short period of 2 to 12 weeks is .67 for immediate and .62 for delayed recognition. The stability for percent retention ranged from 81% to 89% for the two groups.

Although men may outperform women on visuospatial tests such as Judgment of Line Orientation, this difference is not seen on the WMS-III Faces test (M.R. Basso, Harrington, et al., 2000). Faces scores do not correlate as well with the WMS-III "general visual memory factor" or with other visual memory tests as do other measures of visual memory (Millis, Malina, et al., 1999). This low correlation may simply reflect a different aspect of visual memory assessed by Faces than the other tests.

A yes/no recognition format with no built-in control for guessing or response bias lends itself to guessing. This creates a problem when the accuracy of "normal" recognition does not differ much from 50%, i.e., chance. For example, for the oldest age group (85–89 years), a score of 24 (half right) on either the immediate or delayed condition results in a scaled score of 7 (16th–24th percentile range)! Moreover, a score of 34 might be obtained either by correctly recognizing all 24 target faces and correctly saying "no" to 10 foils (making 14 false positive errors), or 10 correct recognitions (14 false negative errors) while correctly saying "no" to all 24 foils. Yet the inferences from these two patterns should be very different despite identical scores.

Clinical findings. Schizophrenia patients tend to perform poorly on tests of memory for faces, including the WMS-III Faces test (Conklin et al., 2002). In addition, first-degree relatives of these patients perform poorly on WMS-III Faces.

Family Pictures (Wechsler, 1997b)

This new test in the Wechsler Memory Scale-III is designed to measure "complex, meaningful, visually presented information," which the test developers consider a "visual analogue to the Logical Memory subtest" (Wechsler, 1997b, p. 15). Like the Continuous Visual Recognition Memory test, many aspects of Family Pictures are easily verbalized.

The test is introduced by showing a "family portrait"—colored drawings of six family members (mother, father, grandfather, grandmother, son, and daughter) and their dog. Subjects then see individual pictures for 10 sec that contain four family members appearing in different situations performing everyday tasks. Subjects are simply instructed to remember as much about each scene as possible for recall after the picture is removed. Recall is scored for reports of which of the characters were in the scene, where they were located in the picture using a 2 × 2 grid, and what each character was doing. A 30-minute delayed recall is also obtained.

Test characteristics. This test can present some scoring problems. Scoring is "character based," meaning that unless the correct person is identified, points for what was going on in the picture and in which quadrant cannot be earned. This can create a problem for some older patients who, based upon their age, may refer to the "father" as "son" and the "son" as "grandson." Further, in a least one picture (i.e., yard scene), the distinction between "father" and "son" is not as clear as in others; even when the subject correctly identifies the activity and place, misclassification of the son results in a score of 0 (it could be argued that in these cases a perfect score of 4 should be awarded since the person confusion is not due to memory failure). The examiner can usually avoid this problem by taking a little extra time during the initial identification of each family member.

Other concerns about the test have surfaced. For example, patients often confuse the picnic with the yard scene on delayed recall, making the scores difficult to interpret (e.g., describing the activities for the yard scene when prompted with "tell me everything about the picnic scene" receives no memory credit). Also, very indistinct actions are portrayed for the characters in the meal scene, and their intended actions are often missed: patients may simply report that everyone is eating, which gets credit for only one of the four persons at the table. This scene has the potential for an additional scoring problem as "looking" is listed as an example of both a 0-point and 1-point action response for two separate characters. Moreover, a verbal memory impairment (i.e., forgetting the picture titles) may also be a source of low scores.

The three components of Family Pictures (who, what, and where) are summed into a single score. Given the potential neuropsychological differences associated with a ventral "what" stream that processes object information and a dorsal "where" stream involved with spatial processing (Mishkin, Ungerleider, and Macko, 1983), it is unfortunate that normative data for each discrete component were not provided. Thus, one cannot easily tell what this test is measuring in the individual case.

Visual Recall: Design Reproduction

A number of abbreviated tests of memory for designs call for a 5- or 10-sec exposure followed immediately, or after a brief delay, by a drawing trial in which subjects attempt to depict what they remember. Probably the most popular designs are the two Memory for Designs I tasks at age levels IX and XI from the Stanford-Binet (Terman and Merrill, 1973; see Fig. 11.2, p. 455). These appear in slightly modified form in the Wechsler Memory Scale-III and in other test sets as well (e.g., Gainotti and Tiacci, 1970). Both the Stanford-Binet and the Wechsler Memory Scale administrations call for a 10-sec exposure followed by an immediate response. A third Binet design, composed of embedded diamonds, appears at age level XII (see Fig. 11.2).

Memory tests requiring reproduction of a design have been used to assess right hemisphere damage. McFie (1960) found a significant number of impaired Binet design reproductions associated with right hemisphere lesions regardless of their specific site within the hemisphere, although this disability was not associated with left hemisphere patients. More recently, extrahippocampal volumes in the right medial temporal lobe, but not the hippocampus, have been associated with Visual Reproduction performance (Kohler et al., 1998; R. Martin et al., 1999).

Visual Reproduction (Wechsler, 1945, 1987, 1997b)

This was originally developed as an immediate recall test, but many examiners added a delayed trial to the original version (VR-O) (e.g., E.W. Russell, 1988). Each of the three VR-O cards with printed designs is shown for five seconds (the third card of each form of the test has a double design; form I contains the IX- and XI-year level designs of the Binet pictures shown in Fig. 11.2). The other two form I designs are from the Babcock-Levy test battery (1940). Following each exposure, subjects draw what they remember of the design. The maximum score is 14. Scoring discrepancies can be quite large and mostly arise from differences of opinion about the degree of accuracy required, with

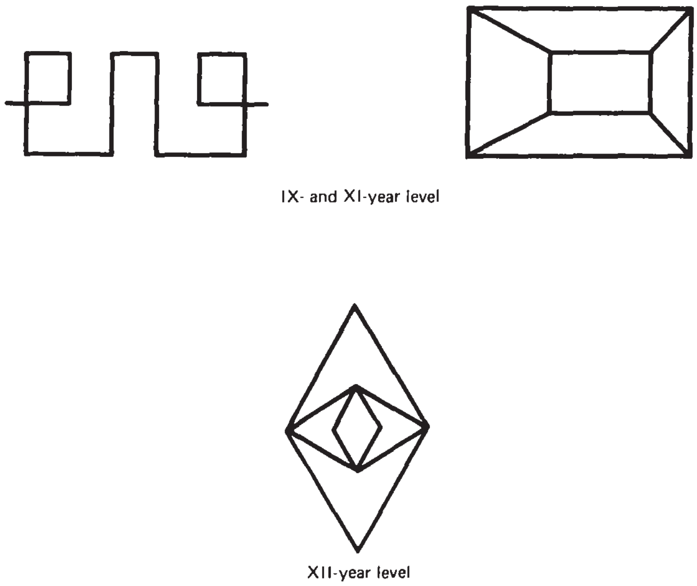

FIGURE 11.2 Memory for Designs models (Terman and Merrill, 1973. Courtesy of Houghton Mifflin Co.)

questions of proportions concerning card B drawings in particular (M. Mitchell, 1987).

Delayed recall trials have been given anywhere from 20 min to an hour later; most examiners ask for a 30-min recall. The procedures described by Trahan, Quintana, and their colleagues (1988) are typical in requesting recall after 30 min of testing during which other drawing tasks are not given ("to limit interference"). Using the same scoring criteria as those for the first trial, these workers offer normative data for four age groups from 18–19 to 70+ (for more normative data for both immediate and delay trials, see also Mitrushina, Boone, and D'Elia 1999; Spreen and Strauss, 1998).

The first formal revision of this test (VR-R) included both immediate and 30-min delay trials (Wechsler, 1987). It consists of four items, of which three contain a single figure (design A is the same on both forms of VR, B is new involving circles within circles, C is VR-O card B) and on the fourth item card, D, are two designs, one containing three and the other two geometric elements. Examples accompany very detailed scoring guidelines with new criteria for the designs retained from the original version. The maximum score is 41 points with the same scoring criteria applied to both the immediate and delayed trials.

Further modifications were introduced in the second revision (VR-III) (Wechsler, 1997b). A simple design was added to lower the floor of the test, one of the cards from VR-R was deleted, and one of the cards from the original scale that was not included in VR-R resurfaced in slightly modified form. As with VR-R, a 30-minute delayed recall is obtained. The newest version contains a 48-item recognition test and a seven-item discrimination test to identify differences in recall and recognition capacities. In addition, a copy task can be given to examine the potential role of motor difficulty. Scoring for the VR-III has been modified so that partial credit may be earned; the manual provides good examples of 0-, 1-, and 2-point responses. The maximum score is 104 points with scaled score conversions for the different age groups (see p. 484).

A scheme for scoring intrusion errors for the VR-O and VR-R was devised to document these often very interesting abnormal response distortions (D. Jacobs, Salmon, et al., 1990; D. Jacobs, Tröster, et al., 1990). Fewer intrusion errors appeared on VR-R performances compared to VR-O, which these authors suggest may be due to the introduction of circular figures along with the rectilinear ones of VR-O.

Test characteristics. All versions of this test display steep age gradients (Wechsler, 1987; see also Mitrushina, Boone, and D'Elia, 1999; Spreen and Strauss, 1998). The score drop-off is particularly sharp in the later years: for example, on VR-R, average performances in the 30–35 point range at ages 56–66 drop to a 20–28 point range at ages 77–87 (Ivnik, Malec, Smith, et al., 1992); on the VR-III, average performances in the 43–51 point range at ages 55–64 are in the 28–34 point range at ages 70–79 (Wechsler, 1997b).

In a large-scale Australian standardization, women obtained an average score that was almost 1 point lower than that of the men (Ivison, 1977), a pattern that has been seen in other studies (Reite et al., 1993). Findings of a sex difference in VR age-related decline are inconsistent (Barrett-Connor and Kritz-Silverstein, 1999; Suutama et al., 2002). Neither the WMS-R nor the WMS-III manual reports on sex effects for VR (Wechsler, 1987, 1997b).

For a sample of subjects with educational backgrounds ranging from 0 to >12 years, education effects were prominent on both immediate and delayed recall trials ($p < .0001$) (Ardila and Rosselli, 1989). Education was also a significant variable in a study of older persons, ages 60 to 94, whose average education levels were in the $10^1/_2$ to $13^1/_2$ year range but within-group variability was large (SDs ranged from 4.78 to 6.71; Ivnik, Malec, Smith, et al., 1992c). However, no significant education effects were found on this test for any of four, mostly younger, patient groups, with education levels averaging from $11^1/_2$ to $13^1/_2$ years although the range of scores within groups was narrower (SDs from 2.34 to 3.61; Trahan, Quintana, et al., 1988). Together these studies suggest that educationally deprived persons may do poorly on this test—and perhaps any other unfamiliar task requiring paper and pencil—but beyond the level of a basic educational foundation, education effects may be small.

As with most memory tests, practise effects occur (McCaffrey, Duff, and Westervelt, 2000b). A group of older subjects (*M* age = 69.3) gained almost 2 points on retesting a year later, losing most of this gain on the next year's retesting (Kaszniak, Wilson, Fox, and Stebbins, 1986). Moreover, with only a 7 to 10 day difference between test and retest, hypertensive patients gained 1 point on immediate recall and 1.62 points on delayed recall; chronic smokers made even greater gains of 1.49 and 2.90 on immediate and delay trials, respectively, with all gains statistically significant (McCaffrey, Ortega, et al., 1992). When tested over multiple sessions, the practice effect for VR-R was less than that for other WMS-R tests (Reite et al., 1993). However, for both immediate and delayed recall trials, gains made on retesting were so small as to be practically inconsequential. Further, on immediate recall of VR-O no practice effects showed up at one year for a youthful group of TBI patients (Dikmen, Machamer, et al., 1990), which probably reflects the patients' diminished learning capacities.

An interscorer reliability coefficient of .97 was reported for VR-R, with scoring differences of 4 points or less and an average difference between two scores of 1.50 (Wechsler, 1987). Woloszyn and his colleagues (1993) report interscorer reliability coefficients of similar magnitude. Internal consistency estimates for VR-R (six age groups) ranged from .46 to .71 on immediate recall and from .38 to .59 on delayed recall (Wechsler, 1987).

For the individual components scored according to VR-R criteria, Card A (flags) is a better measure of visual memory, whereas Card B (boxes) is associated more with conceptual reasoning ability (M.A. Williams et al., 1998). VR correlates significantly with tests involving predominantly visuospatial problem solving and visual memory; the association with other visual memory tests is stronger for the delay trial (Larrabee, Kane, and Schuck, 1983; Leonberger et al., 1991; Trahan, Quintana, et al., 1988). Chelune, Bornstein, and Prifitera (1990) called attention to the consistency with which a visual construction component emerges most prominently when other tests are included in the factor analysis. That Visual Reproduction is often affected by constructional skill has been demonstrated by both factor analysis (Larrabee and Curtiss, 1995) and clinical group comparisons (Gfeller, Meldrum, and Jacobi, 1995).

Neuropsychological findings. On immediate recall, neither version discriminated between patients with right- and left-sided lesions (Chelune and Bornstein, 1988; Delaney, Wallace, and Egelko, 1980), although in one study of patients with lesions confined to the temporal lobes, those with left-sided damage performed significantly better (Jones-Gotman, 1991a). On delayed recall, patients with right temporal lesions obtained significantly lower scores on VR-O than those with lesions on the left or healthy control subjects (Delaney, Rosen, et al., 1980), but VR-R showed only a nonsignificant trend favoring patients with left hemisphere damage (Chelune and Bornstein, 1988). In a large multicenter study of over 500 patients with lateralized temporal lobe epilepsy, there was no effect of seizure onset laterality for the immediate or delayed VR-R conditions, nor for the percentage retention over 30 minutes (Barr, Chelune, et al., 1997).

The relative simplicity of the designs encourages verbal encoding and may account for the general absence of pronounced differences between performances by patients with right-sided or left-sided lesions (see also Jones-Gotman, 1991a). The stimuli were revised in the most recent version to minimize verbalization further. Certainly, this test cannot be used to aid in lesion lateralization. VR-O is sensitive to the effects of TBI, correlating significantly with ventricular enlargement (Cullum and Bigler, 1986). It even distinguished a group of patients with mild TBI from control subjects by virtue of an average 1.3 point difference that was significant (Stuss, Ely et al., 1985). While registering improvement over the first year postinjury, VR-O stabilized at that point, with no further change when these TBI patients were examined at the second postinjury year (Dikmen, Machamer, et al., 1990).

Like other memory tests, VR is very sensitive to cognitive deterioration associated with dementia (Kaszniak, Fox, et al., 1978; Laakso et al., 2000). A correlation between delayed VR-R and right hemisphere parahippocampal gyrus volume has been reported in patients with probable Alzheimer's disease (Kohler et al., 1998). D. Jacobs, Tröster et al., 1990) found that the number of intrusions from previously seen stimuli distinguished Alzheimer and Huntington patients from TBI patients who, like control subjects, made very few intrusion errors; Alzheimer patients had the most intrusions of all. In one study, VR-O surpassed the diagnostic accuracy of MRI hippocampal volume measurements (Laakso et al., 2000). "Wineglass" confabulation has been described in some alcoholic patients in which patients rotated the design on VR-R Card D to become a "bowl and stem" (L.W. Welch, Nimmerrichter, et al., 1997). It is interesting to note these patients' reports of drawing the designs as originally presented, i.e., not rotated.

Multiple sclerosis patients tend to do poorly on both immediate and delay trials (Minden, Moes, et al., 1990), although those treated with high doses of interferon beta-1b demonstrated improved VR-O performance at 2–4 years following treatment initiation (Pliskin, Hamer, et al., 1996). Solvent-exposed workers with subclinical symptoms did not give abnormal performances on VR-O (Bleecker, Bolla, et al., 1990), although meta-analysis suggests that VR is sensitive to lead exposure effects (Seeber et al., 2002).

Complex Figure Test: Recall Administration (CFT) (A. Rey, 1941; Osterrieth, 1944; Corwin and Bylsma, 1993b)

Recall of the Complex Figure typically follows the copy trial immediately, on delay, or both. The Rey-Osterrieth (or "Rey-O," see Fig. 14.2, p. 537) is the most commonly used figure, although other figures designed to be comparable have been developed for repeated assessments (e.g., Taylor figure, see Fig. 14.3, p. 538; Medical College of Georgia [MCG] figures, see Fig. 14.5, pp. 539–540; Emory figures, see Freides, Engen, et al., 1996). Because Taylor figure scores tend to run higher than R-O scores, Hubley and Tremblay (2002) modified the Taylor Figure by decreasing the number of distinctive features (e.g., star, circle in square), including additional lines to increase the complexity of the visual array and modifying the placement of other figure features (see Fig. 14.4, p. 538). A different complex figure was developed for the Repeatable Brief Assessment of Neuropsychological Status (RBANS, see pages 696–697).

In most administrations, subjects are not forewarned when given the copy instructions that they will be asked to reproduce the figure from memory. Because the four MCG figures were designed for drug trials with repeated assessments over relatively short periods of time, subjects are informed that memory will be tested upon completion of the copy trial so the task demands remain fairly constant across testing sessions.

Perhaps because of its popularity, many variations in CFT administration and scoring have been reported; precise scoring criteria are a more recent development. Even among the formal scoring systems, the criteria range from relatively liberal (Loring, Martin, et al., 1990) to strict (Jones-Gotman, personal communication, 1992 [mdl]; J.E. Meyers and Meyers, 1995b). This variability may be due to Rey's omission of scoring criteria in the original test description.

Recall trials follow either a single delay or two delays, the later delay assessing retention. The timing of the recall trials differs among examiners. The "immediate" recall trial has been given in as brief a delay as 30 sec (Loring, Martin, et al., 1990), but following Osterrieth's (1944) convention, some examiners test after a three-min (short) delay (e.g., see Table 11.11) (see also D.T.R. Berry, Allen, and Schmitt, 1991; Delbecq-Dérouesné and Beauvois, 1989). A longer delayed recall, from 30 min (D.N. Brooks, 1972; Corwin and Bylsma, 1993a; Spreen and Strauss, 1998) to 45 min or an hour (Ogden, Growdon, and Corkin, 1990; L.B. Taylor, 1979) has been obtained with or without the early recall (Spreen and Strauss, 1998). Within the limits of an hour, the length of delay appears to be of little consequence (D.T.R. Berry and Carpenter, 1992; Freides and Avery, 1991). As with the copy trial, the examiner may record how subjects go about drawing the figure, either by giving them different colored pencils to track their progress as suggested by Rey (Corwin and Bylsma, 1993b), or by having the examiner note the sequence of their drawings (Milberg, Hebben, and Kaplan, 1996). Although there are advantages and disadvantages to each of these procedures, switching pencils does not appear to distract subjects and may actually be associated with improved memory performance compared to the "flowchart" method (Ruffolo, Javorsky, et al., 2001).

Standardized systems call for scoring the number of elements correctly produced (see pp. 541–543). Adding to them are a number of systems for evaluating and scoring qualitative features of the drawings. Problems in knowing which of the various published norms to use are raised by differences in test administration and scoring, and by poor reliabilities for individual item scoring (Tupler et al., 1995). In addition, simple clerical scoring errors may further confound score validity (Charter, Walden, and Padilla, 2000).

Most studies have found the Rey figure to be harder to recall than the Taylor which typically elicits scores several points higher than the Rey (D.T.R. Berry, Allen, and Schmitt, 1991; Kuehn and Snow, 1992; Loring and Meador, 2003a; Tombaugh and Hubley, 1991). Comprehensive norms are given in Mitrushina, Boone, and D'Elia (1999) and J.E. Meyers and Meyers (1995). Spreen and Strauss (1998) provide a set of age-graded norms for the copy and 30 min recall trials. For the 16- to 30-year sample the 30 min delay norms are roughly comparable to Osterrieth's (1944) findings for 3-minute delayed recall, as were 30 min delay performances of young college students (Loring, Martin, et al., 1990; see Table 11.11). For all older age levels, the Spreen and Strauss 30 min delay scores run 2 or more points lower than Osterrieth's median score of 22, as do J.E. Meyers and Meyers'. In addition to reporting data for the 3-min recall for three subject groups (ages 45–59, 60–69, and 70–83), K.B. Boone, Lesser, and their coworkers (1993) computed a percent retention score ([recall score − copy score] × 100) for their subjects. Normative data for copy, immediate, and delayed recall are available for 211 subjects, as well as recognition and matching trials (Fastenau, Denberg, and Hufford, 1999). These norms are presented in a user-friendly table that transforms the values into the commonly used standard scores ($M = 10 \pm 3$). The MCG figures produce scores that are more comparable to the Taylor than the Rey figure. Despite some variability among the MCG figures (Loring and Meador, 2003a), they generally provide comparable results (Ingram, Soukup, and Ingram, 1997).

Immediate and delayed memory performances are usually similar. Most studies found that few performances using either the Rey or Taylor showed more than a 1- or 2-point difference between immediate and delayed recall trials (e.g., D.T.R. Berry Allen, and Schmidt, 1991; Heinrichs and Bury, 1991; Shorr et al., 1992; see Mitrushina, Boone, and D'Elia, 1999, for these and many more norm sets).

In a comparison of immediate and delayed recall scores of 40 unselected cases (27 men, age range 18–67), 30 (75%) had score differences no greater than 2 points, although four (10%) had 5-point differences. The average difference between immediate and delayed recall was .425. One-third (13) of the delay scores were higher than the immediate scores. Score distributions of ten Taylor figure protocols did not differ from that of the Rey-O. Half the cases were TBI; the others had such various diagnoses as seizure disorder, Huntington's disease, multiple sclerosis, HIV+, toxic encephalopathy, and cerebral vascular disease. Neither age nor diagnosis appeared to contribute to the higher delay scores.

It is important to note, however, that a short-term recall preceding a delayed recall trial may result in a higher delay score than if a delay trial only is given (Loring, Martin, et al., 1990; see Table 11.11). Freides and Avery (1991) reported a 4- to 5-point score increase from immediate to delay for undergraduate students, probably showing this large an increase because they gave no copy trial.

Since the presence or absence of an immediate recall trial will affect performance, this must be kept in mind when choosing a norm set. Alternative scoring systems (e.g., see pp. 461, 490) further complicate efforts to integrate findings from so many different sources. Additionally, Bennett-Levy (1984a) noted that some examiners tend to score recall trials less strictly than the copy trial, based on the rationale that subjects often do not exercise the same degree of care as when copying so that small lapses in precision probably do not represent lapses in memory. He therefore scored both strictly (following the Montreal Neurological Institute standards) and with more lax criteria. He found that although the correlation between these two scoring methods was high (.94), scoring differences amounted to an average of more than 4 points.

TABLE 11.11 Percentiles for Adult Accuracy Scores on Memory Trials of the Complex Figure Test (Rey-O)

	PERCENTILE										
	≤5	10	20	30	40	50	60	70	80	90	99+
Osterrieth*	—	15	17	19	21	22	24	26	27	28	31
Loring†	15	21	24	26	28	29	29.5	30.5	32	33	36
Loring‡	16	22.5	25	28	29	30.5	31	32	33	35	36
Loring§	13.5	16	19.5	20	21	23	24.5	26	28	30.5	32

*$n = 60$.
†$n = 49$, 30 sec recall.
‡$n = 49$, 30 min recall following 30 sec trial.
§$n = 38$, 30 min recall with no prior recall trial.
From Loring, Martin, et al. (1990)

The role of strategy. How the test-taker goes about copying the complex figure will bear a significant relationship to figure recall (Bennett-Levy, 1984a; Heinrichs and Bury, 1991; Shorr et al., 1992). By and large, persons who approach the copying task conceptually, dealing first with the overall configuration of the design and then—only secondarily—with the details, recall the figure much better than subjects who copy the details one by one, even if they do so in a systematic manner (such as going from top to bottom or left to right). The organizational strategy or lack thereof employed during the copy trial is often a strong predictor of subsequent recall (L.K. Dawson and Grant, 2000; Deckersbach et al., 2000; P.D. Newman and Krikorian, 2001), particularly for subjects at lower mental ability levels (Fujii et al., 2000). This difference may be due to the need to recall many more items when they are processed in individual pieces rather than combined into conceptually meaningful units (e.g., see Ogden, Growden, and Corkin, 1990). Somewhat surprisingly, the orientation of the figure during copy (0°, 90°, 180°, or 270°) is not related to recall success (Ferraro et al., 2002). Thus, the CFT may still be a useful test of visual memory when a fixed stimulus position is not possible, such as in bedside assessment.

Applying Osterrieth's system to scoring copying strategies (pp. 544–545), Ska and Nespoulous (1988a) found that until age 74 the usual relationship between strategy and recall level held but that their 75+ group showed a marked decline in both copy ($M = 30.8 \pm 4.1$) and recall ($M = 13.3 \pm 5.4$), although overall, the older subjects' strategic approaches did not differ significantly from those of the younger groups. Moreover, from 41% to 50% of their younger groups of healthy subjects used Osterrieth's level IV, additive details approach (as did six of the ten persons in the 75+ group).

A "perceptual cluster ratio" devised by Shorr and her coworkers (1992, see p. 546) demonstrated this phenomenon. This score correlated significantly with both the copy score (.55) and an "encoding score" (obtained by dividing the immediate recall score by the copy score) (.55) at a much higher level than the correlation between the usual copy score and the encoding score (.35). In regression analyses, the "strategy total" score calculated by Bennett-Levy (1984a, see p. 546) proved to be the first "of the major determinants of copy scores" (sharing this honor with copy time and age) and the first of three "best predictors of later recall" (along with copy score and age).

In an investigation of the role of verbalization versus visualization strategy and the verbalizability of the Rey-O and Taylor figures, those college students who generally tend to use visual strategies recalled both figures better than those who rely on verbal strategies (M.B. Casey et al., 1991). The visualizers were at a greater advantage on the Rey-O figure, but no differences between these two strategy groups obtained for the Taylor figure.

Test characteristics. Significant age effects on recall trials show up consistently (Delbecq-Dérouesné and Beauvois, 1989; Fastenau, Denburg, and Hufford, 1999; Spreen and Strauss, 1998). Spreen and Strauss' data based only on the 30 min delayed recall suggest that decline begins in the 30s, continuing fairly steadily until the 70s when a larger drop in scores appears. On three-minute short-term recall, however, a tendency to an average decrease in scores was first shown by a 41–55 age group, but it did not become pronounced until around age 60, with marked decline continuing into the 65+ ages (Delbecq-Dérouesné and Beauvois, 1989). For relatively well-educated subjects (averaging $14^1/_2$ years of schooling), on 3 min delay recall scores did not decrease notably until after age 69 (K.B. Boone, Lesser, et al., 1993). The ubiquitousness of the late age decline is seen on the Medical College of Georgia figures (see Table 11.12).

Some studies have reported that men tend to recall the figures better than women (Bennett-Levy, 1984a; M.B. Casey et al., 1991; Rosselli and Ardila, 1991). However Freides and Avery's (1991) college students showed no sex differences, nor did the large sample of 211 subjects across different ages (Fastenau, Denburg, and Hufford, 1999). No sex differences were found for recall of the MCG figures (Ingram et al., 1997).

A "cultural level" score derived from education levels contributed significantly ($p < .05$) to recall of the Rey figure (Delbecq-Dérouesné and Beauvois, 1989). Rosselli and Ardila (1991) reported a significant correlation between recall scores and education (.37, $p < .001$), but the inclusion of persons with less than six years of schooling in a sample also containing about equal numbers of persons with more than 12 years of schooling probably exaggerates the contribution of education,

TABLE 11.12 Medical College of Georgia Complex Figure (MCGCF) Data for Two Older Age Groups

Age	N	*Copy*	*Short Delay*	*Long Delay*
55–64	48	35.7 ± .61	30.3 ± 3.8	29.6 ± 4.1
65–75	29	35.6 ± .87	26.9 ± 7.1	26.4 ± 7.2

at least for application to populations with a generally higher average educational level.

Interscorer reliability is good ($r = .91$ to .98) (D.T.R. Berry, Allen, and Schmitt, 1991; Loring, Martin, et al., 1990; Shorr et al., 1992). Test–retest reliabilities using alternate forms (CF-RO, CF-T) were in the .60 to .76 range (D.T.R. Berry, Allen, and Schmitt, 1991). The Rey-O figure is a little more difficult to remember than the Taylor figure (M.B. Casey et al., 1991; Duley et al., 1993; Kuehn and Snow, 1992; Tombaugh and Hubley, 1991) or the MCG figures (Meador, Loring, Allen, et al., 1991) such that CFT-RO scores run a little lower than CFT-T scores. Both immediate and delayed recall trials have a strong visual memory component (Baser and Ruff, 1987; Loring, Lee, Martin, and Meador, 1988) and an almost as strong visuospatial component (D.T.R. Berry, Allen, and Schmitt, 1991).

Neuropsychological findings. Performance on the two recall trials helps the examiner sort out different aspects of the constructional and memory disabilities that might contribute to defective recall of the complex figure. Patients whose defective copy is based more on slow organization of complex data than on disordered visuospatial abilities (more likely with left-sided lesions) may improve their performances on the immediate recall trial (Osterrieth, 1944), and improve further with a second, later trial. These patients tend to show preserved recall of the overall structure of the figure with simplification and loss of details. Patients with right-sided lesions who have difficulty copying the figures display even greater problems with recall (L.B. Taylor, 1979). As a result of the distortions made by patients with right temporal lesions and of loss of details by those whose lesions involve the left temporal lobe, these two groups could not be discriminated on the basis of delayed recall scores alone, although a qualitative error score did differentiate them (Loring, Lee, and Meador, 1988; Piguet et al., 1994). The Loring group cautioned against relying on just one material-specific memory test when attempting to make such an identification. Although both figural and spatial features of the CFT are affected by right medial temporal impairment, the effect is greater for the spatial components, which may be less verbalizable than figural features (Breier, Plenger, et al., 1996).

In addition to scoring in the "traditional" manner (i.e., following Osterrieth's 1944 guidelines), a system for scoring qualitative errors was developed for distinguishing performances by patients with right or left temporal lobe damage (Loring, Lee, and Meador, 1988; Piguet et al., 1994; see Table 14.5, p. 544). These qualitative errors are most likely to occur in recall drawings of patients with right-sided temporal lobe lesions, but they may be applicable to drawings by patients whose right-sided dysfunction is not confined to the temporal lobe, and to TBI patients and those with frontal damage as well.

Patients with right hemisphere damage also tend to lose many of the elements of the design, making increasingly impoverished reproductions of the original figure as they go from the immediate to the delayed recall trial. Those right hemisphere damaged patients who have visuospatial problems or who are subject to perceptual fragmentation will also increasingly distort and confuse the configurational elements of the design.

This showed up in the three trials—copy (a), immediate recall (b), and (approximately) 40 min delayed recall (c)—drawn by a 50-year-old graduate civil engineer 12 years after suffering a ruptured aneurysm of the right anterior communicating artery, which resulted in left hemiparesis, significant behavioral deterioration, and pronounced impairment of arithmetic and complex reasoning abilities along with other cognitive deficits (see Fig. 11.3, p. 461).

CFT recall is sensitive to mild neuropsychological impairment in a variety of clinical populations. Alcoholic patients perform more poorly on recall than controls (L.K. Dawson and Grant, 2000; E.V. Sullivan, Mathalon, et al., 1992), and recall of the CFT continues to be impaired for a longer period following abstinence in older alcoholics than in younger patients (Munro, Saxton, and Butters, 2000). The magnitude of severe postoperative pain was found to be inversely related to CFT recall (Heyer et al., 2000), although the independent contribution of analgesia (i.e., morphine) is difficult to determine since patients experiencing greater pain receive more aggressive pain treatment.

Traumatically brain injured patients also tend to have difficulty on CFT recall trials. Patients with mild TBI showed significant deficits on 3 min recall trials within the first 21 months postinjury (Leininger, Gramling, et al., 1990). Two to five years posttrauma, moderately injured patients (PTA $<$ 3 weeks) achieved significantly higher delayed recall scores than those whose injuries were severe (Bennett-Levy, 1984b). D.N. Brooks' (1972) TBI patients did as well as control subjects on immediate recall but gave impaired performances after a 30 min delay.

Following generally piecemeal copy trials, Parkinson patients had very poor recall scores ($M = 7.55$) (Ogden, Growdon, and Corkin, 1990), as might be expected from other studies, demonstrating the inefficiency of a fragmented copy approach for memory storage. Even after being asked to remember the design before beginning the copy trial, Huntington patients recalled significantly fewer elements than did either control subjects or persons at risk for the disease (whose

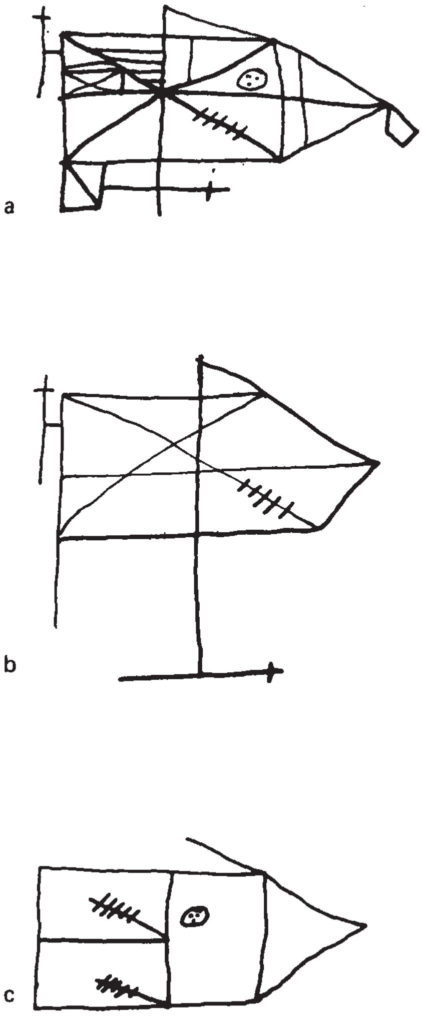

FIGURE 11.3 Complex Figure Test performance of a 50-year-old hemiparetic engineer with severe right frontal damage of 14 years' duration (see text; Chapter 9, Fig. 9.8). (*a*) Copy trial. (*b*) Three-minute recall with no intervening activities. (*c*) Recall after approximately 40 minutes of intervening activities, including other drawing tasks. This series illustrates the degradation of the percept over time when there is a pronounced visual memory disorder.

average scores on both copy and recall trials exceeded those of the control group by a nonsignificant bit) (Fedio, Cox, et al., 1979).

Patients with gliomas who survived at least 4 years after diagnosis differed in CFT recall according to their treatment (Gregor et al., 1996). Patients receiving whole brain irradiation and surgery displayed poorer CFT recall than those with focused irradiation and surgery. Children with acute lymphoblastic leukemia who were treated with intrathecal methotrexate therapy or whole brain irradiation performed more poorly on CFT recall (Lesnik et al., 1998; Waber, Shapiro, et al., 2001).

Complex figure modifications

Several modifications to the Complex Figure test have been developed to overcome limitations in the procedure as originally presented. Patterning their procedure after the Babcock-Levy story recall (see pages xx-xx), Freides and Avery (1991) had subjects study the Taylor figure for 60 sec, recall it, and then gave a second presentation of the figure for additional study with recall following a 20 min delay. Using their two new figures, they decreased exposure time to 30 sec to avoid ceiling effects (Freides, Engen, et al., 1996). Expected age declines appeared. The authors cite Erickson and Scott (1977) in support of using repeated learning trials:

> Basing one's inferences about learning and memory capabilities on immediate recall or recognition of material that has been presented one time seems a poor way of assessing memory. (p. 1144)

Tombaugh, Faulkner, and Hubley (1992) also used the Taylor figure in a learning paradigm with four learning trials, a 30 sec exposure on each trial, and a 2 min limit to recall time. Delayed recall is requested 15 min later, followed by a copy trial that lasts for only 4 min. This technique was sensitive to age differences over a 20- to 79-year range, with prominent score decrements beginning in the 50s for all the memory and learning measures. An apparently faster rate of learning for older subjects simply reflected the very much lower scores made by them on the first trial; even by the fourth learning trial, subjects over 50 never caught up with the younger ones and retained less. In providing a learning curve, this method adds potentially important information not obtainable by standard administration of either Verbal Reproduction or the CFT. It is a somewhat lengthy and possibly tedious procedure for which they developed a 69-point scoring system that greatly increases scoring time and effort; Freides, Engen, et al. (1996) report no psychometric benefit using Tombaugh's system in comparison to traditional scoring methods. In deciding whether to use this technique, the clinician must weigh its potential benefits against the suspected drawbacks of time (for administration and scoring), patient discontent, and examiner impatience with all that scoring.

Complex Figure Test recognition formats. J.E. Meyers and Meyers (1995) devised a recognition trial. Fastenau (1996a; with Denburg, and Hufford, 1999) supplemented the CFT by adding a recognition and a matching trial following delayed free recall (Extended Complex Figure Test, see below). Important differences distinguish these two recognition formats.

J.E. Meyers and Meyers' *Complex Figure Recognition Trial* (1995) presents 12 items for the figure along with 12 foils. The items are copies of internal details from the Rey-O and Taylor figures, both small (e.g., R-O: circle with dots, Taylor: wavy line) and large (the structure of each figure). The subject is asked to encircle each figure that belongs to the "whole design" just drawn. Norms were compiled from performances by 208 intact subjects in the 14 to 60 age range; their average age of 26.55 ± 8.62 attests to the relative youth of this group. Neither age nor education contributed significantly to these scores. This technique distinguished brain injured patients, psychiatric patients, and healthy subjects effectively. Brain injured patients identified more CFT parts than they recalled after either a 3 min or a 30 min delay, although healthy control subjects' recall exceeded recognition (J.E. Meyers and Lange, 1994).

In the *Extended Complex Figure Test (ECFT)* (Fastenau, Denburg, and Hufford, 1999; Fastenau, 2003), for each of 30 items, five stimulus figures are presented vertically, decreasing the effect of bias due to response preference associated with visual field or inattention defects. Each target contains one of the original 18 elements from the figure with four distractor elements. In addition to assessing recognition of the different elements, elements recalled are evaluated in different sets to provide a *global score* (the large rectangle, diagonal cross, and horizontal and vertical midlines), a *detail score* (the cross at the far left of the figure, diamond at the far right, circle with three dots, and five horizontal lines), and *left* and *right element scores*. The detail score was divided so that right- and left-sided elements could be considered separately. The foils for these details also have distractor elements in either the left or right portion of the figure. Normative data on 211 healthy subject ranging from 30 to 85 years of age are presented as scaled scores. The mean age of this group (62.9 years) is older than that of the J.E. Meyers and Meyers (1995) sample indicating its appropriateness for a larger range of patents. Sex effects on the supplemental recognition and matching trials are negligible. "This test adds a little more time but it will have significant yield for some patients" (Fastenau, personal communication, April 2003 [mdl]).

Benton Visual Retention Test (BVRT) (Sivan, 1992)

This widely used visual recall test is often called by its originator's name alone, "the Benton." It owes its popularity to a number of virtues. It has three forms that are roughly equivalent; some studies demonstrate no differences in their difficulty level and other studies indicate that Form D may be a little more difficult than Forms C or E (Benton, 1974; Riddell, 1962), or that Form C is a bit easier than the other two forms (Sivan, 1992). Its norms include both age and estimated original mental ability.

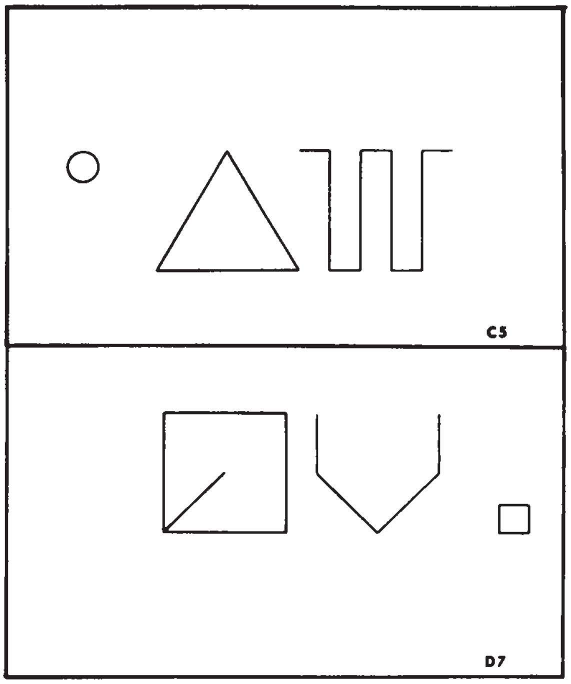

FIGURE 11.4 Two representative items of the Benton Visual Retention Test. (© A.L. Benton. Courtesy of the author)

The three-figure design format is sensitive to unilateral spatial neglect (see Fig. 11.4). All but two of each ten-card series have more than one figure in the horizontal plane; most have three figures, two large and one small, with the small figure always to one side or the other. Besides its sensitivity to visual inattention problems, the three-figure format provides a limited measure of immediate span of recall since some patients cannot keep in mind the third or both of the other figures while drawing a first or second one, even though they may be able to do a simple one-figure memory task easily. Further, spatial organization problems may show up in the handling of size and placement relationships of the three figures.

Both the number of *correct* designs and the number of *errors* are scored. The complex but easily learned scoring system helps the examiner identify error patterns. The manuals furnish adult norms for two administration procedures, Administrations A and C. Administration A allows a 10 sec exposure to each card with immediate recall by drawing (see Table 11.13, p. 463 for adult norms; norms for children ages 8 through 14 can be found in the manuals and in Spreen and Strauss, 1998). Administration B, like A, is also a simple recall test but follows a five-second exposure. Administration B Number Correct norms run about an

TABLE 11.13 BVRT Norms for Administration A: Adults Expected Number Correct Scores, by Estimated Premorbid IQ and Age*

	EXPECTED NUMBER CORRECT SCORE, BY AGE		
Estimated Premorbid IQ	*15–44*	*45–54*	*55–64*
100 and above	9	8	7
95–109	8	7	6
80–94	7	6	5
70–79	6	5	4
69 and below	≤5	≤4	≤3

BVRT Norms for Administration A: Adults Expected Error Scores

	EXPECTED ERROR SCORE, BY AGE			
Estimated Premorbid IQ	*15–39*	*40–54*	*55–59*	*60–64*
110 and above	1	2	3	4
105–109	2	3	4	5
95–104	3	4	5	6
90–94	4	5	6	7
80–89	5	6	7	8
70–79	6	7	8	9
69 and below	≥7	≥8	≥9	≥10

*These data are identical to those given in Sivan's 1992 test manual except for slight differences in age range: The three new age ranges for Number Correct scores at 15–49, 50–59, and 60–69; for Error scores they are 15–44, 45–59, 60–64, and 65–69.

average of 1 point below those reported for Administration A. Administration C is a copying test in which the subject is encouraged to draw the designs as accurately as possible. On Administration D, which requires the subject to delay responding for 15 sec after a 10 sec exposure, the average Number Correct score may be lower than that for Administration A by 0.1 to 0.4 points (Sivan, 1992); however, intersubject variations can be great as some patients improve with delay while others' scores drop.

Mitrushina, Boone, and D'Elia (1999) have compiled a comprehensive collection of norms for this test. Spreen and Strauss (1998) present a data set of error norms organized by sex, seven age levels (from 10–19 to 80–89), and two levels of education (with or without college degree) that may be useful in evaluating performances of educationally advantaged subjects. Also focusing on better educated subjects, Youngjohn, Larrabee, and Crook (1993) give norms for five age groups (18–39, each of the next three decades, and 70+) and three levels of education (12–14, 15–17, and 18+). Extensive norms based on 156 healthy volunteers between 61 and 97 years of age, in addition to 625 subjects with memory concerns and 196 patients with mixed etiology are presented by Coman et al. (1999).

The examiner should give the patient a fresh sheet of paper, approximately the size of the card for each design. The test publisher sells a response booklet, but half sheets of letter-size paper work well. To avoid the problem of a patient "jumping the gun" on the memory administrations—and particularly on Administration D—the pad of paper may be removed after completion of each drawing and not returned until it is time for the patient to draw the next design.

When the copy administration is given first, the examiner is able to determine the quality of the patient's drawings per se and also familiarize the subject with the three-figure format. Well-oriented, alert patients generally do not require the practice provided by administration C, so it need not be given if there is another copying task in the battery. Patients who have difficulty following instructions and lack "test-wiseness" should be given at least the first three or four designs of a series for copy practice.

Six types of errors are recognized for scoring purposes: *omissions, distortions, perseverations, rotations, misplacements* (in the position of one figure relative to the others), and errors in *size.* Thus, there can be, and not infrequently are, more than one error to a card.

Interpretation of performance is straightforward. Taking the subject's age and "estimated premorbid" ability into account, the examiner can enter the normative tables for Administration A and quickly determine whether the Number Correct or the Error score falls into the impairment categories. On Administration B, the normal tendency for persons in the age range 16–60 is to reproduce correctly one design less than under the 10-second exposure condition of Adminis-

Biber Figure Learning Test (BFLT)[1] (Glosser, Goodglass, and Biber, 1989)

This test consists of ten test items each composed of two geometric figures (e.g., two truncated rectangles back to back, a triangle with an inside triangle rotated 180°) and 30 distractors (three for each target figure, of which one differs in orientation, one in shape, and one in being quite different from the target). Following exposure to all ten items, shown one by one for two seconds, the subject is asked to draw them from memory for five learning and recall trials. An immediate recognition trial follows the free recall trials, then a 20-minute delay period ends with Delayed Free Recall and Delayed Recognition. A final two trials involve Immediate Reproduction of each design after a 3 sec exposure, and then a Copy trial for all designs just drawn "with any error."

For four age groups from 40–49 through to 70–79, a significant age effect appeared on free recall but not on recognition trials. Test–retest reliabilities were in the .79 to .91 range, with a year-and-one-half wait between tests. Correlations of the various BFLT scores with visual reasoning, construction, and other visual memory tests were in the .46 to .71 range; the highest correlation was with Wechsler's Paired Associate Learning test. Free recall on learning trials did not differentiate patients with right hemisphere disease from those whose lesions were on the left, and both groups performed significantly below the level of healthy subjects. Although delayed recall for both patient groups was also below that of the healthy subjects, the right hemisphere group had a retention score (Delayed Recall divided by Trial 5) proportionally like that of the healthy subjects and significantly higher than the retention score made by the left hemisphere patients. However, on recognition trials the relationship between the two lateralized lesion groups was reversed in that patients with left hemisphere involvement performed like the control subjects but the right brain injury group recognized significantly fewer designs. The lower retention score made by left brain injury subjects and the high correlation with a verbal learning test together implicate a significant verbal learning component for this test.

A 15-item format—the *Biber Figure Learning Test–Extended* (*BFLT-E*)—is patterned after the AVLT (Glosser, Cole, et al., 2002). It includes two parallel versions for repeat assessment. After five trials in which the original 15 designs are shown at a rate of one every 3 sec, each trial followed by a drawn recall, a second set of designs is presented also followed by an immediate recall trial. Then, without additional exposure, patients are asked to draw the original 15 items. After 20 to 30 min, delayed recall and recognition memory are tested, the latter with a 45-item, yes/no procedure displaying the original 15 items, seven items from the distractor set, and 23 foils.

With this longer version of the test, significant left vs. right temporal lobe epilepsy focus differences showed up on Learning trial 5, a score for learning across trials, immediate and delayed memory, as well as recognition measures (Glosser, Cole, et al., 2002); poorer performances were associated with right seizure-onset patients. For discriminating lateralized seizure onset, the BFLT-E was superior to traditional memory measures (Rey Auditory Verbal Learning Test and Wechsler Memory Scale-Revised) which failed to identify to group differences.

Visual Spatial Learning Test (VSLT)[2] (Malec, Ivnik, and Hinkeldey, 1991)

These authors sought a visuospatial learning task suitable for patients with movement disorders. The test consists of a 6 × 4 grid and seven different nonsense designs that are (truly) difficult to verbalize. After seeing the designs placed on squares on the grid, subjects are given an empty 6 × 4 grid and 15 designs with the task of selecting the target seven and placing them as they were when seen on the grid. Five learning trials are followed by a 30 min delayed recall trial. Performance is scored for recognition learning of the designs, recall of the target positions on the grid, and recall of designs in their proper places on the grid.

The usual age gradient appeared for both learning and delay trials. Highest correlates of VSLT scores were with Visual Reproduction scores (in the .29 to .46 range), but the VSLT also had correlations with some verbal memory tests that fell within this range. Before temporal lobectomies, both right and left temporal lobe seizure patients performed at similar levels on all VSLT measures. After surgery these two groups tended to diverge, with right-resected patients performing less well (see also Ivnik, 1991). Nevertheless, a factor analytic study failed to demonstrate the VSLT as a measure of nonverbal memory distinct from verbal memory (G.E. Smith, Malec, and Ivnik, 1992).

Normative data are available for elderly patients ranging from 56 to 97 years of age (Malec, Ivnik, Smith, et al., 1992). The VSLT is a good discriminator of de-

[1]Dr. Glosser provides her fax (215-349-5579) and e-mail (glosser©mail.med.upenn.edu) addresses.

[2]This material may be requested from James F. Malec, Ph.D., PM & R-1D-SMH, Mayo Medical Center, Rochester, MN 55905. (e-mail: malec.james©mayo.edu)

mentia patients from intact elderly persons, reportedly classifying 78.9% of healthy subjects and 87.9% of dementia patients correctly (Malec et al., 1992).

Diagnosticum für Zerebralschäde–Revision (Diagnostic workup for cerebral disease–Revised) (Helmstaedter, Pohl, Hufnagel, and Elger, 1991)

This visual memory test asks the patient to learn nine simple geometric designs, all of which can be produced by connecting five separate lines. The designs are individually presented to the subject at the rate of one design every 2 sec for the patient to recall by using 5 wooden sticks of equal length to reconstruct the designs. A maximum of six trials are given but testing is stopped when the patient reproduces all items. Performance feedback is not given.

In reports about this test several different measures have been scored. The primary measures include a learning score summing performance over trials, total number of rotations (at least 30°) or mirror inversions, failures (false positive reproductions), and a 30 min delayed recognition with drawings of the target figures interspersed with 12 foils.

Sex differences have been reported, with men displaying an advantage (Helmstaedter, Kurthen, and Elger, 1999). This test has predictive value for everyday memory performance in healthy controls in that performances was directly related to subjects' ability to recall specifics of the assessment examination when questioned one week later (Helmstaedter, Hauff, and Elger, 1998).

Neuropsychological findings. Patients with right temporal lobe epilepsy and epilepsy patients with bitemporal EEG abnormalities performed significantly poorer on immediate recall and learning compared to either healthy controls or patients with left temporal lobe seizures (Helmstaedter et al., 1991). Patients with right temporal lobe epilepsy had a greater number of rotated or mirrored responses. In a subsequent study in which hippocampal sclerosis was identified by MRI, patients with right TLE with hippocampal sclerosis performed more poorly on learning, recognition, rotations, and failures than those without MRI evidence of sclerosis; the latter did not differ from healthy controls (Gleissner, Helmstaedter, and Elger, 1998). These findings suggest that the right–left differences reported in the early 1991 study resulted from the inclusion of large numbers of right TLE patients in whom hippocampal atrophy or sclerosis was present. This test also appears to be sensitive to "crowding," in which language shifts from the left to the right hemisphere following an early left hemisphere lesion (Helmstaedter, Kurthen, Linke, and Elger, 1994).

Heaton Figure Memory Test (Heaton, Grant, and Matthews, 1991)

This test uses the Visual Reproduction stimuli from the original Wechsler Memory Scale but introduces several important changes. Each card is shown for 10 seconds, although the patient does not reproduce the designs until all three cards (four designs) have been presented. The three cards are again shown up to a maximum of five trials or until a criterion of 15 points is reached. Thus, this format incorporates a learning component in which acquisition rate can be examined. In addition, a four hour delay trial is obtained.

In a mixed patient sample, factor analysis has generally linked the initial trial with immediate visual memory (DiPino et al., 2000). Although a distinct visual memory factor was not extracted, overall performance was related to the California Verbal Learning Test, supporting its use as a memory measure. In elderly patients with age-associated memory impairment, immediate recall was correlated with left amygdala volume while delayed recall correlated with both right and left amygdala volume (Soininen et al., 1994). No correlation with hippocampal volume emerged in this study.

Ruff Light Trail Learning Test (RULIT) (Ruff, Light, and Parker, 1996)

This nonverbal learning task asks the subject to learn a specific pathway or trail through circles spread over a piece of paper; the circles are interconnected by lines (see Fig. 11.5, p. 468), and the "start" and "end" circles are labeled as such. Subjects are told that they will be learning a 15-step path that is not the shortest route between the start and end circles. Subjects are not told how to choose the correct response from the different alternatives but must find it by trial and error. On moving their finger from one circle to the next, subjects are informed whether they are correct or not; if incorrect, they go back to the last correct circle and make a different choice. A trial is completed when the end point is reached following the correct path, regardless of the number of digressions made on the way. The task is discontinued after two consecutive trials in which the 15-step sequence is correctly traced, up to a maximum of 10 trials. Performance can be evaluated across learning trials and for "step errors," errors that are repeated at the same circle. Factor analysis of a group of healthy subjects suggests that the RULIT assesses visual learning/

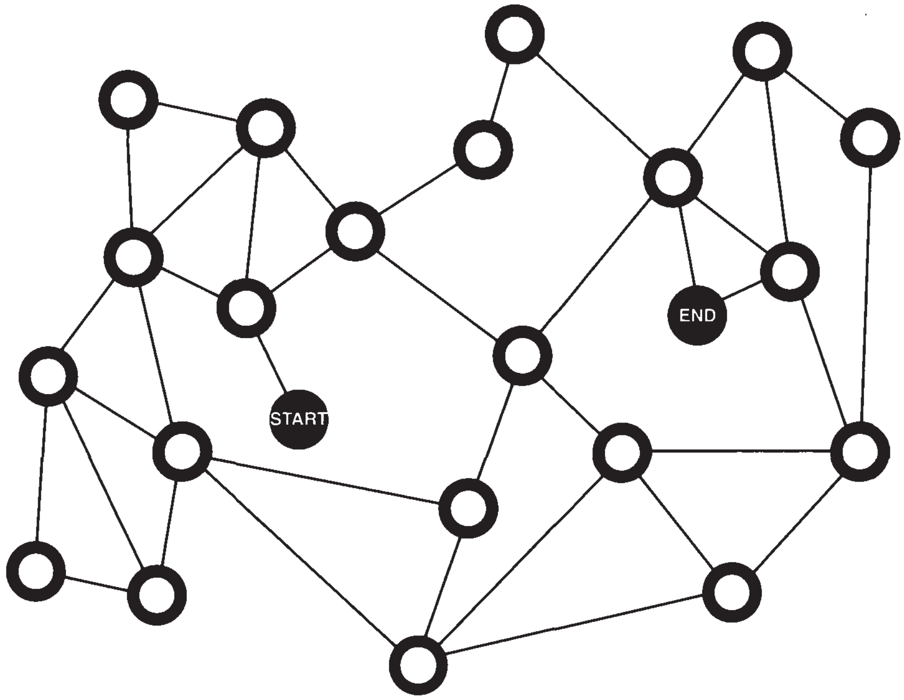

FIGURE 11.5 Ruff-Light Trail Learning Test (RULIT) (reduced size). Reproduced by special permission of the publisher, Psychological Assessment Resources, Inc., from the Ruff-Light Trail Learning Test by Ronald M. Ruff, Ph.D., and C. Christopher Allen, Ph.D. © 1999 PAR, Inc. Further reproduction is prohibited without permission of PAR, Inc.

memory and is distinct from its verbal counterpart, the Selective Reminding Test (C.C. Allen and Ruff, 1999).

Shum Visual Learning Test (SVLT)[1] (Shum, O'Gorman, and Eadie, 1999)

This test uses Chinese characters as stimuli since they are not easily verbalized (by non-Chinese, that is!). The format is similar to the Rey AVLT: five learning trials, a second character set designed to measure interference, then a recognition trial for the original stimulus set. The addition of the second set is a slight modification of the test from its original presentation (Eadie and Shum, 1995), and the new stimulus set is included to measure interference effects in retention.

The ten target stimuli are displayed one-by-one for 2 sec each on each of the first five learning trials. Retention is tested by means of a recognition format, with the ten stimuli interspersed with ten distractors. These foils were created by modifying the target stimuli through adding, deleting, or relocating a stroke from the Chinese character. The same set of distractors is used for each recognition memory trial. Subjects are giving up to 5 sec to respond. After the fifth learning trial, the second stimulus set is presented and recognition memory is tested with a new distractor set; this is followed by a recognition trial of the original stimulus set. After a 20-min delay, recognition of the first set is again tested.

Normative data are available for three major indices: *learning* (difference between the first and fifth trials), *retention after interference* (difference between trials 5 and 7), and *delayed retention* following a 20-min delay (difference between trials 7 and 8). Unlike most clinical memory measures, however, performance criteria are reported using a nonparametric measure of recognition memory based on signal detection theory which incorporates hit rate and false alarms for descriptive purposes. Correct responses for individual trials, the sum of correct responses for the five learning trials, and the number of false positives can be scored and evaluated.

Neuropsychological findings. TBI patients with severe injuries performed more poorly than healthy controls both during the first year after injury and later (D.H. Shum, Harris, and O'Gorman, 2000) although

[1]Dr. Shum can be reached at School of Applied Psychology, Griffith University, Nathan, Queensland, Australia 4111. (e-mail: d.shum©mailbox.gv.edu.au)

no group effect for time after injury appeared. In addition to poorer learning scores, TBI patients made more false positive recognition errors. No effect of TBI compared to healthy controls was present on the Retention after Interference or Delayed-Retention trial, an unexpected finding.

Hidden Objects

Testing the patient's immediate memory and learning for spatial orientation and span of immediate memory by asking for recall of where and what objects have been hidden is an examination technique in the Terman and Merrill (1973) Stanford-Binet tests and in mental status examinations (e.g., Strub and Black, 2000). Strub and Black hide four common objects, such as a pen, keys, watch, or glasses, in the examining room while the patient observes, naming each object as it is hidden. The patient's task is to find or point out each hiding place after at least ten minutes of interpolated activity. Adults with unimpaired visual learning remember all four objects and hiding places. Barbizet and Duizabo (1980) also used familiar objects (e.g., pen, button, cork) in their version of the hidden objects test: The examiner gives the patient five objects to name and place in a box, which is then hidden from view. After 15 min, the examiner asks the patient which objects had been hidden, where, and to describe them. Recall is tested again at one and 24 hours. Barbizet and Duizabo pointed out that the technique of asking for immediate recall and then delayed recall at two subsequent times helps to differentiate among conditions in which memory disorders occur. They demonstrated this by describing a jovial *grand alcoolique,* who found a bottle of wine that had been hidden behind him 3 minutes earlier but after 10 minutes more recalled neither the hiding place nor what had been hidden.

TACTILE MEMORY

Tactual Performance Test (TPT)

Like the Knox Cubes, the material for this test came from the Arthur (1947) battery of tests (*Seguin Formboard,* see Fig. 11.6). Although originally administered as a visuospatial task, Halstead (1947) converted it into a tactile memory test by blindfolding subjects and adding a drawing recall trial. Halstead incorporated this version of the test into his battery which has continued to be used for neuropsychological testing (Reitan and Wolfson, 1993). Halstead's administration requires three trials, the first two with the preferred and nonpreferred hands, respectively, and the third with both hands. The score for each trial is the time to completion, which Halstead recorded to the nearest tenth of a minute. Their sum is the *Total Time* score.

Differences in administration time were reported by Snow (1987b) who noted that Reitan (1979) suggested ending a trial after 15 minutes for patients who are "getting discouraged and . . . making very slow progress" unless they are close to a correct performance; but other workers discontinue at 10 minutes routinely or at the examiner's discretion.

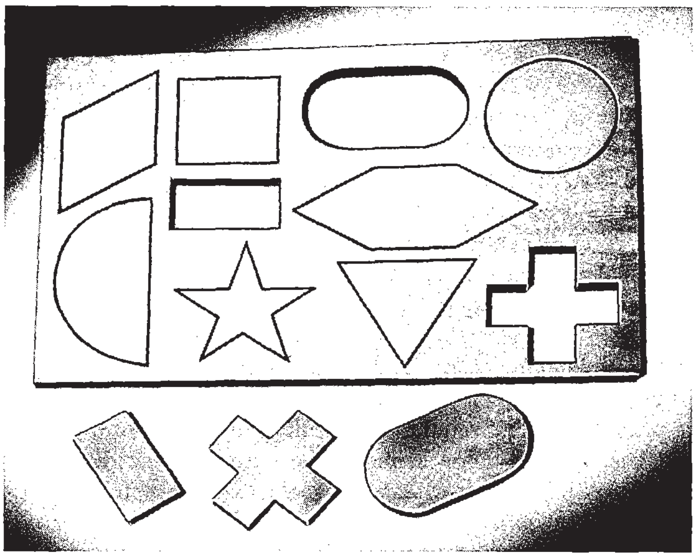

FIGURE 11.6 One of the several available versions of the Sequin-Goddard Formboard used in the Tactual Performance Test. (Courtesy of the Stoelting Co.)

On completion of the formboard trials it is concealed, the examiner removes the blindfold, and instructs the subject to draw the board from memory indicating the shapes and their placement relative to one another. Thus two measures of incidental memory are obtained (Hom and Reitan, 1990): the *Memory* score is the number of shapes reproduced with reasonable accuracy; the *Location* score is the total number of blocks placed in proper relationship to the other blocks and the board. Scoring of Memory and Location is at the examiner's discretion.

Cutting scores, developed by Halstead on a limited and skewed subject sample (see p. 671) were retained by Reitan to indicate "organic" impairment. They are now superseded by adequate normative data (Mitrushina, Boone, and D'Elia, 1999; Heaton, Grant, and Matthews, 1986, 1991; Spreen and Strauss, 1998).

Test characteristics. Age contributes significantly to TPT performances, so much so that when using Halstead's cutting scores for older persons, many will fall into the impaired range (Ernst, 1987; L.L. Thompson and Parsons, 1985). For example, Heaton, Ryan, et al., (1986) found that 30% and 58% of their healthy subjects in the 40-59-year age range had Total Time and Location scores (respectively) in Halstead's *defective* range; by ages 60+, 77% and 91% "failed" on their Total Time and Location scores.

Whether sex makes a difference has not yet been satisfactorily determined differences have shown up on one or more of the three scores (Ernst, 1987; Heaton, Ryan, et al., 1996; L.L. Thompson and Parsons, 1985) but not always (Filskov and Catanese, 1986). Where differences appear, men tend to perform better than women on Total Time while women make higher scores on one or both of the recall drawing measures. A small education contribution showed up for all three scores on the large-scale normative study by Heaton and his colleagues (1996) but Ernst (1987) examined a group of elderly Australians and found none. Bernard (1989) reported that, using the Halstead cut-off scores, 36% of a relatively poorly educated group of healthy young men had Location scores in the impaired range, and 18% were classified as "impaired" on their Total Time scores.

Reviewing reliability studies, L.L. Thompson and Parsons (1985) concluded that with between-test delays of three months to one year, test–retest reliability was generally adequate, and best (.68–.93) for Time (see also Spreen and Strauss, 1998). On three week retesting the reliability correlation for Time was .69, for Memory it was .80, and .77 for Location (Bornstein, Baker, and Douglass, 1987). With few exceptions, performances improve for control subjects after as much as a two-year interval between tests (McCaffrey, Duff, and Westervelt, 2000b). Some score gains on each of the TPT measures have also been documented for a variety of patient samples. However, E.W. Russell (1985) found the Location score to be particularly unreliable. Moreover, Snow (1987b) noted a report (P.W. Martin and R.L. Green, cited in Snow) that scoring for Memory produced agreements in 71% to 76.3% of cases, but for Location the agreement between judges dropped to 56.7% and 63.8%, with some judge pairs agreeing as little as 36% on Memory and 29% on Location, which puts the reliability of these measures on somewhat shaky ground.

Internal consistency coefficients of the Memory score ranged from .64 for a group of healthy subjects to .74 for a group of patients with diffuse injuries (Charter and Dutra, 2000). Internal consistency coefficients for Location were higher, ranging from .77 (the healthy subjects) to .82 (the patient group). As observed by Charter and Dutra, these scores fall into the border zone of the "unacceptable/fair" and "fair/good" classifications describing strength of internal consistency according to Cicchetti's 1994 guidelines. The rhombus is the most difficult block to remember and the circle is the easiest (Charter and Dutra, 2001). Using discrimination indexes which vary from -1.00 to $+1.00$ depending upon how well performance by an individual block discriminates between groups, three blocks for Memory (diamond, oval, and rhombus) and one block for Location (oval) were considered unacceptable (discrimination values $\leq .3$) (Charter and Dutra, 2001).

L.L. Thompson and Parsons (1985) remarked on the relatively high level of intercorrelations between the three TPT measures, especially between Memory and Location (.56 to .71), with TPT Time correlating in the .32 to .72 range with Memory, in the .26 to .62 range for Location. Out of one large group of tests, all TPT measures loaded on the same factor with the most commonly scored five measures showing a very narrow range of correlations (.50 to .69) (P.C. Fowler, Richards, et al., 1987). This further supports the impression that these different trials are essentially measuring much the same thing.

Neuropsychological findings. Although there appears to be little doubt that markedly slowed or defective performances on the formboard test or the recall trial are generally associated with brain damage, the nature of the defect remains in dispute. Most investigators have found a right–left hemisphere differential favoring performance by patients with left hemisphere lesions (Heilbronner and Parsons, 1989; Teuber and Weinstein, 1954; L.L. Thompson and Parsons, 1985). Halstead (1947) considered this test to be particularly sensitive to frontal lobe lesions; yet Teuber and Weinstein's posterior brain injured patients performed least well, and their anterior brain injured patients made the best scores of their three brain injured subgroups (1954; Teuber, 1964; see also L.L. Thompson and Parsons, 1985). Teuber (1964) noted that their findings are "not unreasonable, in view of the known symptomatology of parietal and temporal lesions. What is difficult to understand is that this formboard task should have been considered a test of frontal pathology at all" (p. 421).

The difference between the time taken on the preferred hand and that on the nonpreferred hand trials may provide a clue as to the side of the lesion. Normally, as learning takes place, trial II takes a little less time than trial I even though it is performed with the nonpreferred hand, and trial III takes the least amount of time. When this pattern is reversed in right-handed subjects (i.e., Trial II with the left hand takes longer than Trial I, the right-handed trial) depressed functioning of the right hemisphere may be suspected (L.L. Thompson and Heaton, 1991).

Like other tests calling upon a complex of neuropsychological functions for optimum performance, TPT scores are typically lower for brain damaged populations than for intact persons (e.g., Spreen and Strauss, 1998; L.L. Thompson and Parsons, 1985). A fairly consistent pattern of dysfunction on this test has emerged in studies of chronic alcoholics (Fabian, Jenkins, and Parsons, 1981; W.R. Miller and Saucedo, 1983; Parsons and Farr, 1981). When administered in the Halstead-Reitan format, right-handed alcoholics showed the most slowing on the nonpreferred hand trial, significant slowing on the preferred hand trial, impaired performance with both hands, and an abnormally low Location score with a normal or near normal Memory score. This pattern was essentially the same for both male and female alcoholics, although women (both alcoholics and controls) tended to outperform men on the Memory score but do relatively poorer than their male counterparts on the Location task (Fabian et al., 1981).

Shortcomings of the TPT. Probably because of its inclusion in the popular Halstead-Reitan Battery—and perhaps because this battery is so often administered by a technician rather than by the clinician responsible for deciding which tests to give—the TPT continues to enjoy wide usage despite several important drawbacks. The chief clinical drawback is the enormous discomfort experienced by many patients when blindfolded, which, when added to their frustration in performing a trial that may take as many as ten or even more minutes for some to complete, creates a degree of psychological distress that does not warrant use of an instrument that may give very little new information in return. The other major problems are the amount of time consumed in giving this test to older and brain injured patients, and the equivocal and often redundant nature of the data obtained. It is most appropriately used with visually compromised persons [mdl].

TPT variants. De Renzi (1968) used a six-figure rather than the ten-figure formboard in his studies. The smaller board may reduce time about one-third, making this test feasible for ordinary clinical use (C. Clark and Klonoff, 1988). Certainly the reduction in number of forms does not seem to have reduced the discriminating power of this technique (see De Renzi, 1968; also, E.W. Russell, 1985, Spreen and Strauss, 1998). Rather, Clark and Klonoff found that both Memory and Location scores are higher with the six-hole board, improving the discrimination potential of these scores. They also found that the six-hole formboard showed good reliability over two years and four testing sessions, with essentially no gain from session to session. Further, construct validity was comparable to that of the ten-hole board. In comparing the six and ten hole formboards, E.W. Russell (1985) concluded that the six hole board was actually more sensitive to performances by severely impaired patients because the ten hole board was so difficult for them that their very large time scores registered neither differences in severity of dysfunction nor lateralization deficits. Russell suspected that the larger number of blocks contributes to disrupted performances of very impaired patients by confusing them.

Teuber and Weinstein (1954) administered this test somewhat differently than Halstead and Reitan. They gave only two trials to blindfolded subjects, one with the board in the usual position and one with the board rotated 180°. Like Halstead and Reitan, they followed the formboard task with a drawing recall, but scored only for memory, not for location. Performances of their frontal lobe injured patients were consistently superior to those of patients whose injuries involved other cortical areas (Teuber, 1964). The frontal lobe patients also recalled more forms on the drawing trial than any other group, and the occipital lobe patients recalled the fewest.

Tactile Pattern Recognition (B. Milner, 1971)

Four pieces of wire, each twisted into a distinctly different nonsense shape (see Fig. 11.7, p. 472), comprise the material for a tactile test of immediate memory. Subjects never see the wire figures. After several training trials on matching the figures with no time delay,

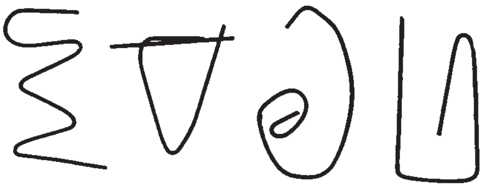

FIGURE 11.7 Tactile nonsense figures (Milner, 1971. © 1971, Pergamon Press. Reprinted by permission)

matching follows an increasing delay length up to two minutes. During 30 sec delay trials, a distractor task of copying matchstick patterns was introduced (B. Milner and Taylor, 1972).

Six out of seven commissurotomized patients performed best with their left hand, indicating that complex perceptual learning can take place without words and that it is mediated by the right hemisphere. B. Milner and Taylor (1972) found little difference between patients with unilateral surgical lesions and intact interhemispheric connections, and control subjects on this task. Both groups performed rapidly and virtually without errors except for a single error made by each of two left temporal lobectomy patients using their right hands. These findings suggest that, even after delay with an intervening distractor task, this test may be too easy for patients whose lesions are as localized and circumscribed as temporal lobectomy patients.

INCIDENTAL LEARNING

The virtue of testing for incidental learning is that this technique looks at learning as it occurs naturally in the course of events. The WIS-A provides an opportunity for assessing incidental learning. E. Kaplan, Fein, and their colleagues (1991) devised a technique for using the WAIS-R Digit Symbol (see pp. 368–369) to measure incidental learning in addition to obtaining a score on a standardized coding performance. They note which square the patient filled in at 90 seconds (the time allotted for the test) but allow the patient to continue until the end of the next-to-the-last row. They then fold the test sheet under so that only the unmarked last row shows and request subjects to fill in from memory as many of the symbols as can be recalled. After this, subjects are asked to write out as many of the symbols as they can remember. A recall of six of the nine symbol pairs is at the low end of the range of normal recall. Patients who cannot place seven or more correctly are encouraged to write as many of the symbols as they can recall in the margin below. Significant age and education effects were seen. As a guideline, the authors state that 80% of healthy older adults will correctly match at least three symbols with digits; scores below three should be noted but only scores of zero may be regarded as definitely abnormal. The exception is participants in their 80s with no more than a high school education, who recalled an average of two associations. Over 85% of adults over 50 years of age will produce at least six correct symbols, and any score below five may be treated as abnormal except for the group noted above. Age and education norms are provided for 177 adults ranging in age from 50 to 90 using the Kaplan procedure for administering the WAIS-R (Joy et al., 2000). Age-related decline was found on a shortened version of this task with 131 South African adults of Anglo-Saxon origin ranging in age from 20 to 89 (Shuttleworth-Jordan and Bode, 1995). Participants completed 42 coding pairs rather than the 68 pairs used in the Kaplan version. Age-related decline was greatest in participants 70 years and older.

The WAIS-III changed the name of this test to Digit Symbol-Coding, increased the time to completion to two minutes, and included a procedure for assessing incidental learning similar to the one proposed by E. Kaplan and her colleagues (Wechsler, 1997a). The Administration and Scoring Manual does not provide normative data for this test, but the Technical Manual (The Psychological Corporation, 1997) provides cumulative percentages of participants performing at different levels up to the 50th percentile.

With 110 items in $7^1/_2$ rows, the Symbol Digit Modalities Test (SDMT) (A. Smith, 1982; see pp. 370–371) has both more items and several more rows than Digit Symbol. Its length usually allows for both immediate and delayed recall trials to be examined on the same test form, as few patients get as far as the next-to-last row. Thus, in most cases, immediately upon completion of the test proper the examiner can fold the last two rows back and ask the subject to fill in only the now top row from memory, marking it at the left with a big X. Folding this row back makes it possible to give the delayed trial on the last row. In addition to the examination of both incidental learning and retention, impaired self-regulation may show up in a patient who continues the immediate recall trial on the bottom line, despite specific instructions to fill in only the line marked with an X. Impaired self-regulation may also become evident when patients pair up two different symbols with the same number, or write in different numbers for two of the same symbols. Since incidental recall on SDMT is based on two 90 sec trials, the cutoff score between 5 and 6 recommended for Digit Symbol may be a little generous. However, in clinical practice it is usefully discriminating. Most patients recall as many or almost as many digit–symbol pairs correctly on delayed recall as they had recalled immediately. Patients with significant retention problems recall fewer—sometimes only one or even none—on delayed recall. A few patients will improve from immediate recall to delayed recall. A close inspection of their test performances typically provides other indications that theirs is a slowed processing problem.

A procedure for examining incidental memory using WIS-A Similarities (Vilkki, Holst, Ohman, et al., 1992) gives a nice example of such an examination and how it can be incorporated into the usual test proceedings. Upon completing Similarities, subjects were asked for both free recall and cued recall of test items, the latter trial presented as a paired associates test. When given to patients who had undergone surgical repair of a subarachnoid hemorrhage a year earlier, free recall but not cued recall discriminated between patients functioning normally and patients with obvious neuropsychological deficits.

A 15-item version of the Boston Naming Test has also been used for incidental recall of the items (Bryan and Luszcz, 2000). Because the task was used to assess incidental learning rather than confrontation naming only, correct names were given for incorrect or unnamed items. On this task, age accounted for only 8% of the variance and speed. Performance was predicted by scores on verbal knowledge, speed, and retrieval on other tasks.

A female superiority has been observed for verbal memory on incidental learning tasks too (McGivern et al., 1998). For example, using pictorial material with varying degrees of verbalizable stimuli (e.g., complex scenes vs. easily labeled stimuli), Chipman and Kimura (1998) found that women performed better than men when the stimuli were verbalizable. Incidental learning is not often examined in patients with brain disorders. For Parkinson patients the findings are mixed. Ivory and colleagues (1999) reported that patients with Parkinson's disease without dementia were impaired in learning verbal material under incidental learning conditions; but another study found that incidental learning of Parkinson patients without dementia equalled that of control subjects, while only Parkinson patients with dementia were impaired (Azuma et al., 2000).

PROSPECTIVE MEMORY

Remembering to execute an action planned for the future requires a mechanism for signaling when the appropriate time has arrived and recall of the nature of the intent. Signals may be either time or event based (Einstein and McDaniel, 1990). An alarm clock may serve as a signal for a *time-based intention,* while remembering to tell a friend something at the next encounter would be an *event-based intention.* Most people depend on prospective recall of numerous tasks in their daily activities. Investigators are beginning to explore ways to examine prospective memory systematically (see Brandimonte et al., 1996). The Rivermead Behavioral Memory Test (RBMT) contains several tasks designed to measure prospective memory (Wilson, Cockburn, and Baddeley, 1985; see pp. 491–493). The *Cambridge Behaviour Prospective Memory Test* (Groot et al., 2002) has four time-based items such as, "In 20 minutes please ask me for a copy of the newspaper;" and four event-based prospective memory tasks, one of which is, "When the alarm rings, please put this briefcase under the desk."

Many studies of prospective memory have looked at the influence of age, prompted by knowledge of age-related decline on tests of episodic memory and the frequent complaints of elderly persons that they forget to carry out actions or go to a room and cannot recall what they intended to do while there. Findings of an age-related decline in prospective memory have been mixed, dependent on the demographic characteristics of the subjects and the type of prospective memory studied. Einstein and McDaniel (1990) instructed subjects to press a response key every time a designated word appeared in a word list task. Performance was examined under both external-aid and no-aid conditions. No age effect was found for well-educated younger and older subjects (mean age 69 years) on this prospective memory task, although the younger group outperformed the elderly on the more traditional tasks of free recall and recognition of word lists. Using a memory aid facilitated prospective memory at similar levels for both groups. While no age differences were found on an event-based task, such as the one above, age differences were found with a time-based task in which subjects were instructed to perform an action every 10 minutes (Einstein, McDaniel, et al., 1995). Studying subjects with more varied educational attainment, occupational status, and verbal ability, K.E. Cherry and LeCompte (1999) found a significant age-related decline in older adults (mean age 70 years) with lower ability level but not for higher ability older adults. Studying prospective memory in subjects up to 92 years of age, Bisiacchi (1996) found age effects only for subjects over 70 years. Other data suggest that age-related decline in prospective memory occurs when conditions of learning and recalling are sufficiently complex (Einstein, Smith, et al., 1997).

The roles of "hippocampal" memory and "frontal" functioning in prospective memory were examined in elderly (64 to 85 years) adults who showed differential performances on these functions (McDaniel et al., 1999). Hippocampal function was assessed with traditional memory tests and the Wisconsin Card Sorting Test, Controlled Oral Word Association Tests, WAIS-R Arithmetic, and WMS-R Mental Control and Digits Backward were used to evaluate frontal competence. The prospective memory task involved pressing a response key every time a target word appeared in a multiple-choice test of general knowledge. The subjects with the best performance on "frontal" tasks responded significantly more often to the target words than those who did poorest on these tasks. The "hippocampal" memory factor did not produce a significant effect. The authors concluded that the results were in line with theoretical speculation that prefrontal systems subserve significant processes in prospective memory.

A few studies have examined prospective memory in patients with brain disorders. Cockburn (1996a) asked patients with mixed brain disorders and controls to indicate when a task duration reached five minutes (time-based intention) and to change color pens when numbers on a cancellation task changed from two-digit to three-digit and to initial the last page of the test booklet (event-based intentions). Subjects also did other memory tasks and tests of executive function. Control subjects generally performed better than patients, particularly on the time-based task. No relationship was found between performance on executive and memory measures and success on the time-based prospective task. Patients who had poorer prose recall were more likely to fail event-based tasks.

Prospective memory is impaired in many severely injured TBI patients (Kinsella, Murtagh, et al., 1996). For a group of patients with brain injury, mostly TBI, the total perspective memory score for both time-based and event-based tasks correlated significantly with scores on traditional episodic memory tests and executive function tests (Groot et al., 2002). No significant correlations were found for attention, speed of processing, or years of education. Event-based tasks were easier than time-based tasks for patients and control subjects. Note-taking improved performance. In a study comparing patients with focal lesions in various sites, P.W. Burgess, Veitch, and their colleagues (2000) found that patients with lesions of medial regions in the left hemisphere had problems in prospective memory as measured by ability to follow self-generated plans for completion of multiple tasks. All six patients with severe memory impairment from herpes encephalitis failed prospective memory tasks, both time-based and event-based, even though the patients showed variable performance on traditional verbal memory tests (Sgaramella et al., 2000). Huppert and Beardsall (1993), using the Rivermead Behavioural Memory Test, found that impairment in event-based prospective memory was more prominent than impairment in learning and recall of words and objects in patients with very mild dementia.

REMOTE MEMORY

The need to assess very long-term memory arises particularly when retrograde amnesia is present and the examiner wants to know how far back it extends. Thus, testing for the integrity of remote memory usually concerns persons with brain conditions that result in retrograde amnesia, such as Korsakoff's disease, temporal or frontal lobe pathology, and those with memory problems incurred in special circumstances, such as treatment with electroconvulsive therapy (ECT). Usually, the retrograde amnesia shows a temporal gradient in which it is most severe for the period preceding the precipitating event. Several strategies for measuring retrograde amnesia involve recall or recognition of information that is commonly held. Unfortunately, in using test items that range from recent to remote topics, an instrument developed to assess gradients of long-term memory must be constantly updated or it will soon become obsolete. This precludes the development of a well-standardized test of remote memory because of the impossibility of going through elaborate standardization procedures every few years. Bahrick and Karis (1982) described methods for assessing remote memory (what they call "long-term ecological memory") and discussed some of the attendant methodological problems.

The interpretation of data from remote memory studies has been questioned since some of the material may have been relearned years after the event (when presented in an article, a book, or a television program, e.g.) (M.G. O'Connor et al., 2000; Sanders, 1972). Sanders also wondered how even-handed this examination technique was since the amount of interest in events such as the death of a prime minister of another country or in personalities such as politicians or movie stars varies so greatly from person to person. McCarthy and Warrington (1990) point out that various items of current salient information will acquire differing retention values as decades pass. Scores on a test of familiarity with television program titles, for example, were positively related to the amount of time the subjects watched television (Harvey and Crovitz, 1979). These tests also presuppose a degree of nationwide cultural homogeneity that one may no longer be able to count on, not just in the United States but in most sizable English-speaking countries and perhaps in some continental European countries too.

Recall of Public Events and Famous Persons

News events tests

A variety of remote memory tests have been based on information from well-known news events. Recall and recognition of public events were investigated in Great Britain by Warrington and Silberstein (1970), who examined the usefulness of both a recall and a multiple-choice questionnaire for assessing memory of events that had occurred in the previous year. Subjects took this test three times at six-month intervals. This technique showed that both age and the passage of a year's time affected recall and recognition of once-known information, and that recall was much more sensitive to age and time effects than was recognition. This method was then extended over longer periods with the development of a multiple-choice *Events Questionnaire* using events for the four preceding decades selected to give even coverage over the 40-year span (McCarthy and Warrington, 1990). A companion test of "well-known" faces covering the previous (approximately) 25 years was also developed in both free recall and multiple-choice versions. With long time periods, both recognition and recall techniques registered significant decrements for age and the passage of time.

Hodges and Ward (1989) used the *Famous Events Test,* consisting of 50 famous events from the 1930s to the 1970s randomly interspersed with 50 made-up events. The subject must identify the true event and, if correct, tell the decade in which it occurred. Patients with transient global amnesia could adequately identify true events but were deficient in dating those in the two decades prior to the amnesic episode. Squire, Haist, and Shimamura (1989) developed a later updated (Reed and Squire, 1998) *Public Events Test* composed of 145 events reported in the news from the 1940s through the 1990s that asks first for free recall and then provides a four-choice format. Patients with the most extensive temporal lobe damage showed the greatest memory loss for facts and events in the most recent decades preceding their brain injury. Kopelman (1989) developed a now updated *News Events Test* (Kopelman, Stanhope, and Kingsley, 1999). Memory for events familiar to British subjects is tested by asking them to identify what was happening in 40 pictures from each decade from the 1960s to the 1990s. Partial credit is given for incomplete identification. Patients with temporal lobe lesions (mostly herpes encephalitis), frontal lobe lesions, and Korsakoff's syndrome performed poorly, showing temporal gradients such that items from the 1960s were recalled better than items from the 1970s and 1980s. The *Dead-or-Alive* memory test (Kapur, 1989) asks subjects to indicate whether a famous personality from the past is dead or alive, whether the cause of death was natural, and what the year of death was, expressed in five-year bands. The items consist of 30 famous people who have died and 10 who are still living. This test demonstrated the remote memory deficits of a patient with severe TBI (Kapur, Scholey, Moore, et al., 1996).

The *Transient News Events Test* (M.G. O'Connor et al., 2000) tests recall of remote events that had time-limited media exposure. Selected news items had had extensive popular appeal but discontinued coverage. Amounts and durations of news coverage were obtained from *New York Times*' records; to be considered, an item had to radically decrease in frequency of mention over a three-year period. The resulting 40 items cover the years from 1952 through 1992. Free recall is scored for both correct and partially correct answers. Recognition probing is used for any items that the subject does not recall. No age-related decline was observed for the 20 to 80 year-old range. All ages demonstrated a recency effect in that events from the recent past were recalled better than remote ones. Younger subjects were unable to tell about events that predated their birth, as expected. Men had superior recall for the remote time periods compared to women, but no other sex differences were observed.

Famous people tests

As part of a series of studies of memory disorder in Korsakoff's disease, several tests involving recall or recognition of famous people were developed (M.S. Albert, Butters, and Levin, 1979; M.S. Albert, Butters, and Brandt, 1980). The *Famous Faces Test* consists of black and white photographs of people who achieved fame in each of six decades (1920s to 1970s) and, in the free recall portion, subjects are asked to name the person shown in the photograph. *Facial Recognition Test* is a multiple-choice recognition of the sample people. Twenty-nine photographs, taken when the subjects were young, were paired with photographs of these same people who were still famous when they were old (e.g., Charlie Chaplin) and presented in randomized order to make up the *Old–Young Test.* In addition, two questionnaires about famous people from these decades were constructed, one testing recall, the other, recognition. Patients with Korsakoff's disease averaged almost 60 years of age and 12 years of education. Control subjects were matched on these variables. This latter group did not show a gradient of loss of information with the passage of time. On most analyses of the data from these tests, the patients showed a marked gradient, from low scores for recent material to scores approaching normal for material from early decades. When this set of tests was given to patients recently diagnosed with Huntington's disease and to advanced Huntington patients, no temporal gradient was found for either of the patient groups, as both sets of patients performed poorly on material from all decades (M.S. Albert, Butters, and Brandt, 1981). However, both patients with Alzheimer's disease and control subjects recognized photographs of familiar faces taken when the person was old more readily than those taken earlier (R.S. Wilson, Kaszniak, and Fox, 1981).

The Famous Faces Test was updated to study remote memory in Alzheimer patients (Hodges, Salmon, and Butters, 1993). The new version has photographs of 85 people who were famous from the 1940s through the 1990s. Both semantic and phonemic prompting cues are given along with a four-choice recognition format. Alzheimer patients were significantly impaired in all test conditions, displaying a temporal gradient for recognition, identification and naming with phonemic cues, but impairments for spontaneous naming or naming with semantic cues did not vary with decades. Thus, they appeared to have lost stored knowledge about the person and not simply a naming deficit. Patients with dorsolateral frontal lobe lesions were impaired on free recall of famous faces but less impaired when given multiple-choice recognition alternatives (Mangels et al., 1996). Unlike Alzheimer patients, they appeared to have knowledge of the person but were unsuccessful in retrieving the information in the free recall condition.

Presidents Test[1] (Hamsher and Roberts, 1985)

While this test requires updating every four to eight years, barring unfortunate circumstances, the update involves only the addition of the photo of the new president and discarding that of the last of the six which had been serving as items in this test. Four different administrations examine: (1) *Verbal Naming* (*VN*), which asks for free recall of the current president and his five immediate predecessors; (2) *Verbal Sequencing* (*VS*), in which six cards with the presidents' names are handed to subjects in a "fixed, quasi-random" order with instructions to arrange them chronologically; (3) *Photo Naming* (*PN*), which shows the presidents' pictures in the same order as the VS cards for the subject to name; and (4) *Photo Sequencing* (*PS*), which asks for a chronological sequencing of the photos. The naming tests each have a maximum score of 6. The sequencing tasks are scored by rank order correlation (Spearman's *rho*) between the correct sequence and the one given by the subject.

Test characteristics. For VN scores, an age effect was found only for subjects with 12 or fewer years of schooling. For PN and PS, only educational differences showed up, while neither age nor education affected VS performances. Score corrections are provided for VN and PN but not PS because of the great variability within the lower education group. VN, PN, and VS each have one cutting score, but two were determined for PS for $\geq$13 and $\leq$12 years of education. In one factor analytic study this test loaded on a remote memory factor which, interestingly, also included a significant weighting (.42) for Digit Symbol (Larrabee and Levin, 1986).

Neuropsychological findings. No control subjects failed more than two tests, and only 8% failed one or two although only 33% of brain damaged patients succeeded on all four tasks (Hamsher and Roberts, 1985). A comparison of patients with lateralized damage found that significantly more with right-sided lesions failed the sequencing tasks than those with left-sided involvement (R.J. Roberts, Hamsher, et al., 1990). Patients with bilateral/diffuse damage or dementia are likely to have memory failures, but few with lateralized lesions fail the memory parts of the test. A significant relationship was found between general cognitive deterioration and number of task failures.

Black and white photographs of the seven most recent U.S. presidents before George Bush and seven other well-known figures, e.g., Neil Armstrong, were used for testing recognition memory of the presidents and for obtaining temporal order of their terms (Storandt, Kaskie, and Von Dras, 1998). Both healthy older persons and patients with mild dementia of the Alzheimer type correctly recognized the presidents and both groups produced U-shaped patterns of errors in temporal ordering, although the dementia group produced more overall errors.

Fama, Sullivan, Shear, et al. (2000b) made the task more difficult by asking patients to write down all the presidential candidates since 1920 and to identify their party affiliation and the years they ran for office. They were then asked to: select from six choices which candidates ran against each other in a particular election and choose the year that the election occurred; sequence the names written on cards of presidential candidates within a single political party from 1920 to 1988; and, identify photographs of presidential candidates from the elections of 1920-1980, giving the candidate's name, party affiliation, and year(s) of the election. Con-

[1]This test can be obtained from Kerry de S. Hamsher, Ph.D., Neuropsychology Clinic, 1218 W. Kilbourn Ave., #415, Milwaukee, WI, 53233-1325.

trol subjects had correct free recall of approximately 65% of the elected candidates since 1920 and less than 25% of the defeated candidates. Parkinson patients without dementia performed as well as controls, while Alzheimer patients were severely impaired. A similar pattern was seen on free recall photo naming. Although the Alzheimer patients performed better on recognition tasks compared to free recall, they were impaired compared to Parkinson patients and control subjects except for recognition of names of elected candidates.

Autobiographic Memory

Another aspect of remote memory is the ability to recall one's own history. Rarely, retrograde amnesia involves autobiographical events and even more rarely affects autobiographical mempory more than knowledge of public events and people (J.J. Evans et al., 1996), although some amnesic patients have greater loss of personal memory than general information (Kopelman, 2002). The virtue of examining autobiographic recall is that all people have had full exposure to their history, making it a rich and culture-fair examination resource. It taps a different aspect of memory, and may be useful for patient counseling (B.A. Wilson, 1993). The drawback, of course, is the difficulty of verifying someone else's personal names, dates, and events. I can tell you that Miss Donovan was my first grade teacher [mdl], but how can you check up on me? Only the exceptional case, such as the prominent scientist who had written an autobiography just two years before succumbing to Korsakoff's psychosis, provides a reliable history for the examination of remote memory (N. Butters and Cermak, 1986). For the rest, the examiner can only test the validity of a patient's self-report by interdata comparisons (e.g., do dates and events make chronological sense?), clinical judgment of the clarity and integrity of the patient's responses and, where possible, by reports of others.

The Crovitz Test (Crovitz and Schiffman, 1974)

This test or its modification (Sagar, Cohen, et al., 1988) has been used to examine the efficiency of autobiographic recall from specific time periods. Subjects are asked to describe from any time period personal experiences of a unique episode associated with each of ten common nouns (e.g., car, bird), and estimate the date of occurrence. Four minutes is allowed for each noun, and if necessary, prompts or cues are offered after two minutes. Responses are scored according to the specificity in time and place of the recalled memory and the richness of details. Recollection of memories from various ages is noted. In a restrained time condition subjects are asked to recall memories from particular ages, such as "before the age of 17" (N.E. Kroll, Markowitsch, et al., 1997). Healthy subjects tend to produce memories from all decades, most from the previous decade (Hodges and Oxbury, 1990). Patients tested during transient global amnesia episodes showed impaired uncued recall and a virtual absence of recent memories (Hodges and Ward, 1989). Six months after resolution of an episode of transient global amnesia, patients' recall of autobiographical memories was impaired in both cued and uncued conditions. Patients with severe TBI produced very few memories from before the onset of their injuries (N.E. Kroll, Markowitsch, et al., 1997).

Autobiographical Memory Interview (AMI) (Kopelman, Wilson, and Baddeley, 1989, 1990)

This questionnaire was developed to standardize the collection of autobiographical data and to provide a range of time spans and item types. It contains two sections: an *Autobiographical Incidents Schedule* and a *Personal Semantic Memory Schedule*. Each schedule asks three questions from each of three time blocks: Childhood (e.g., preschool, primary school), Early Adult Life (e.g., first job, courtship, marriage in 20s), and Recent Events (e.g., a recent visitor, an event in place where interviewed). Patients who cannot respond to a question are given prompts (e.g., for childhood block, first memory? involving brother or sister? etc.). Responses are graded on a 0–3 scale which takes into account the clarity and specificity of the response so that the maximum score for each time block is 9. The Personal Semantic Memory Schedule has four parts, inquiring into Background Information, Childhood, Early Adult Life, and Recent Information. Here the three questions in each part concern the specifics of names, dates, and places. Background Information was allocated a maximum of 23 points, each other section has a maximum score of 21 points. Questionnaire scores were examined in a correlational study with other remote memory tests, producing coefficients in the .27–.76 range, with most .40 or above. Interrater reliability coefficients were satisfactory ($r = .85$). While full confidence in patients' veracity cannot be achieved, this technique appears to satisfy practical requirements as a test of remote memory.

Neuropsychological findings. Amnesic patients performed significantly below control subjects on all variables, with the greatest difference between these groups occurring on recent memory as the controls made almost perfect scores here while amnesics' recent recall (both semantic and event) was poorest (Kopelman, Wilson, and Baddeley, 1989). Examples provided by the authors show how patients' confusion or acuity tends to relate to their performance on the AMI. Patients with some conditions show impairment with a temporal gradient: for Korsakoff patients it is steep with relative sparing of earliest memories (Kopelman, Stanhope, and Kingsley, 1999); of two herpes encephalitis patients, the one with the more extensive temporal lobe lesion did show a temporal gradient on the AMI, but remote memories of the other were relatively spared, as were the temporal lobes (J.M. Reed and Squire, 1998). When multiple sclerosis is severe, overall performance is poor with a slight temporal gradient (Kenealy et al., 2002). Eslinger (1998a) studied the effect of focal temporal lobe lesions in eight patients. Four with bilateral lesions had more severe retrograde memory deficits on the AMI than one patient with unilateral lesions; the other three with unilateral lesions were entirely normal on this inventory. In Eslinger's study a patient with a bilateral prefrontal lesion also showed striking defects in autobiographical memory, except for childhood personal–semantic memories.

FORGETTING

Forgetting involves memory decay over time. Inability to retrieve information either by free recall or recognition suggests that forgetting has occurred. Most techniques for measuring learning can be used to examine forgetting by adding recall or recognition trials spaced over time. Examining Korsakoff patients, Talland (1965a) used a delayed recall format—for example, with recall trials of hours, days, and up to a week to establish forgetting curves for many different kinds of material.

The *savings* method provides an indirect means of measuring the amount of material retained after it has been learned (H.S. Levin, 1986). This method involves teaching the patient the same material on two or more occasions, which are usually separated by days or weeks, but the second learning trial may come as soon as 30 min after the first. The number of trials the patient takes to reach criterion is counted each time. Reductions in the number of trials needed for criterion learning (the "savings") at a later session are interpreted as indicating retention from the previous set of learning trials. Warrington and Weiskrantz (1968) demonstrated some retention in severely amnesic patients over one- and four-week intervals by using the savings method with both verbal and nonverbal material. No other method they used gave evidence that these patients had retained any material from their initial exposure to the tests. In another application of the savings technique, both postacute brain damaged patients and control subjects were given Logical Memory and Verbal Paired Associate Learning (WMS-R) 24 hours apart and then compared for savings on the second administration (B. Caplan, Reidy, et al., 1990). Their savings scores "produced the sharpest differentiation" between the two groups.

The score devised by D.N. Brooks (1972) to document the relative amount of information lost between the various trials of the Complex Figure Test can be applied to other tests as well:

$$\frac{\text{CFT-I} - \text{CFT-D}}{\text{CFT-I}} \times 100$$

Brooks demonstrated this by also using his "% Forgetting" score to compare performances on immediate and delayed trials of the Logical Memory and Associate Learning subtests of the Wechsler Memory Scale. Tröster, Butters, Salmon, and their colleagues (1993) devised another formula for calculating savings:

$$\%\ \text{savings} = \frac{\text{Delayed recall}}{\text{Immediate recall}} \times 100$$

Using this technique for Logical Memory and Visual Reproduction (WMS-R), these authors demonstrated that for both memory tasks older normal subjects had a somewhat higher rate of forgetting than younger ones; Huntington patients' rate of forgetting was higher than elderly subjects', and Alzheimer patients' rate of forgetting greatly exceeded that of the Huntington patients. The WMS-III uses this formula for calculating the "% Retention" for delayed recall measures.

These formulas work well for patients whose recall scores are not at the extremes of the distribution. Patients who have very little immediate recall may achieve a good savings score based on equally scanty delayed recall.

On WMS-III, for example, a 45-year-old patient whose Logical Memory I (immediate recall) raw score is 9 and LM-II (delayed recall) is 7 will have a good % Retention Score of 78, even though the delayed recall score is only *borderline defective* for this age. Another patient of the same age whose LM-I raw score is 65 and LM-II is 40 will have *superior* performances on both recall trials, but the % Retention Score will be 62.

Therefore, forgetting scores should never be interpreted in isolation from initial learning levels.

When the effects of acquisition differences were controlled, rates of forgetting verbal material over a 20-min interval were similar for ages 20–79 (Tombaugh and Hubley, 2001). These authors found that when the retest interval was increased to one day, more forgetting occurred with increasing age. No further differential decline appeared for delays up to 62 days. Age-associated differences were greatest on a word list task, which provides limited opportunity for encoding, and least on paragraph and word pairs in which the format provides a more associative structure. The standardization sample from the WMS-III shows that forgetting rates are similar for ages 16 to 89 on Logical Memory and Visual Reproduction (Haaland, Price, and La Rue, 2003). Age-related decline on these tests was mostly due to poor immediate recall. Nonverbal forgetting may have a steeper age-related decline. For example, older subjects exhibited faster rates of forgetting than younger ones when tested for recognition of magazine photographs one day and one week after initial viewing (Huppert and Kopelman, 1989).

Some patients with degenerative disorders, such as Huntington's disease, have normal retention but forgetting is accelerated in Alzheimer's disease (Massman, Delis, and Butters, 1993). Reports of forgetting rates in other memory disorders are inconsistent and probably relate to the length of delay intervals and the method of study. By extending exposure to the material to be learned and with a 10 min delay, Huppert and Piercy (1979) found normal rates of forgetting for Korsakoff and Alzheimer patients. Patients who have undergone electroconvulsive shock treatments have accelerated forgetting (P. Lewis and Kopelman, 1998; Squire, 1981), and TBI patients have accelerated forgetting during posttraumatic amnesia (H.S. Levin, High, and Eisenberg, 1988). Rapid rates of forgetting new information can continue for months and years following moderate to severe TBI (Vanderploeg, Crowell, and Curtiss, 2001).