Journal of Clinical and Experimental Neuropsychology
1987, Vol. 9, No. 4, pp. 340-358

0168-8634/87/0904-0340 $3.00

Memory Assessment in Neuropsychology: Theoretical Considerations and Practical Utility*

David W. Loring
Medical College of Georgia

and

Andrew C. Papanicolaou
The University of Texas Medical Branch

ABSTRACT

Memory assessment is one of the principal objectives of neuropsychological evaluation. Yet, careful examination reveals very clear shortcomings in the memory tests employed by neuropsychologists. Specifically, most procedures are selected on the basis of their ability to detect structural brain pathology rather than their ability to assess memory performance *per se* or the constituent operations that underlie it. This paper addresses the shortcomings in the structure of several representative memory tests in neuropsychology, how some of these limitations have been overcome with newer scales, and presents practical and theoretical considerations for the development of new clinical memory measures.

With the introduction of sophisticated brain-imaging techniques over the past several decades, the clinical and research functions of the clinical neuropsychologist have shifted from establishing the cerebral localization of lesions and the localization of brain function. Two aspects of this shift, one more theoretical and one largely applied, are seen. Presently, there has been a theoretical shift toward integration with cognitive psychology and analysis of cognitive "components," although localization is still of importance, and an applied shift toward rehabilitation involving the analysis of more pragmatic behavioral "components" (Hart & Hayden, 1986). However, when performing an evaluation of cognitive functioning, the neuropsychologist must first ensure that the tests employed genuinely assess the cognitive operations of interest in order to

* Preparation of this article was supported in part by Department of Education Grant G008435031 to Andrew C. Papanicolaou. We gratefully acknowledge the assistance of George M. Faibish and Tessa Hart, both of whom critically read earlier versions of this manuscript, and the helpful suggestions of our anonymous reviewers. We thank Patricia Downs for her help in manuscript preparation.

Requests for reprints should be addressed to David W. Loring, Department of Neurology, Medical College of Georgia, Augusta, Georgia 30912.

Accepted for publication: May 15, 1986.

avoid making erroneous conclusions regarding the behavioral sequelae associated with brain pathology.

Given the importance of memory evaluation in contemporary neuropsychology, one would expect a considerable degree of sophistication on the part of the professionals performing these assessments, both in their approach to the phenomena of memory and in the assessment procedures they select. Regrettably, however, many neuropsychologists appear to believe tacitly a variant of a familiar theme; that is, memory is that quantity that the Wechsler Memory Scale (WMS) measures. Similarly, when employing newer measures attempting to parcel memory into different components such as storage vs. retrieval (e.g., Selective Reminding Procedure), they fail to remember that an unambiguous operational definition of a psychological construct is completely independent of construct validity.

This paper is not intended as a survey of research performed with various memory tests, since excellent reviews exist already (e.g., Erickson & Scott, 1977; Prigatano, 1978), nor is it meant to update those reports. Similarly, we do not attempt to describe the many techniques available to the neuropsychologist to assess memory since this information is readily available (e.g., Lezak, 1983). Rather, our intention is to evaluate several representative approaches to memory assessment currently in widespread use by clinicians and researchers and to examine their theoretical assumptions and construct validity. These include the Russell and Boston revisions of the WMS and the Selective Reminding Procedure of Buschke. In addition, we discuss the theoretical and practical considerations necessary for the construction of adequate memory tests.

WECHSLER MEMORY SCALE REVISIONS

The Wechsler Memory Scale (WMS) is the most widely used clinical test for memory assessment (Erickson & Scott, 1977). The validity, standardization, and general psychometric properties of the WMS, however, have been extensively criticized (e.g., Erickson & Scott, 1977; Prigatano, 1977, 1978). Briefly, these criticisms refer to inadequate normative information, the assumption that memory is a unitary phenomenon (i.e., the concept of MQ), that the test assesses constructs that, although perhaps necessary for successful memory performance, are not genuine measures of memory (e.g., Orientation and Mental Control), the fact that they correlate highly with intelligence test performance, and their low interscorer reliability for certain subtests.

In an attempt to remedy some of the limitations associated with the WMS, two independent variations of the test have been developed. By evaluating the factor structure of the test, Russell (1975) selected two memory subtests that appeared to be most sensitive to brain damage and, at the same time, differentiated verbal from figural memory (Logical Memory and Visual Reproduction). The remainder of the subtests were discarded for either their inability to

discriminate mild brain-damaged patients from a mixed group of non-brain-injured subjects, or for not measuring memory *per se*. In addition, a delayed-recall condition was included for both remaining subtests to evaluate retention over a one-half hour interval.

The Boston revision of the WMS retains all the elements of the original WMS, but has added items to the scale to improve its utility for routine neuropsychological assessment (Milberg, Hebben, & Kaplan, 1986). However, the most important changes from a memory assessment standpoint relate to the Logical Memory, Visual Reproduction, and Paired Associate subtests. Not only were delayed-recall conditions added, but direct questioning and recognition assessment were also included. These changes will be discussed in greater detail later.

Logical Memory

The Logical Memory (LM) subtest has face validity, appearing to tap that aspect of memory of which patients complain when they state that they are "unable to remember things." Two paragraphs are read to the subject, and following each passage, the subject repeats as much of the story as he or she is able. No explicit instructions requesting verbatim recall are provided, however, although some examiners have modified Wechsler's instructions to include a phrase similar to "try to remember as much of the story as you can, using as many of the same words and phrases." The stories are divided respectively into 22 or 24 "memory units," and credit is received for each idea recalled.

The division of the paragraph into memory units results in two interrelated problems. The first is practical, involving both scoring and reliability, whereas the second involves the issue of construct validity. Russell (1975, 1981) states that the paragraphs are administered and scored according to Wechsler's directions, except that a sum rather than an average value is obtained. However, no scoring criteria for correct ideas are presented by Wechsler, nor are descriptions of such criteria found in published reports. Some neuropsychologists require a verbatim response to give credit, others give one-half credit for a "close" response, while many will score an item correct if the essential content of the memory unit is preserved (e.g., "kids" for "little children"). The nature of the closeness is a matter of personal judgement on the part of the neuropsychologist, since responses may be similar on a number of dimensions (e.g., associative, connotative, denotative, referential).

A telling illustration of this problem emerges when one examines the literature on LM performance. Table 1 summarizes the performance of a variety of nonneurologic and control subjects in different studies. As the table clearly shows, there is a great heterogeneity in what has been reported as nonimpaired performance. Therefore, what may be interpreted with one set of norms as impaired performance on this subtest, another set of norms would suggest is acceptable and within the limits of normal variation.

Efforts to remedy this lack of scoring consistency have recently been made. Power, Logue, McCarthy, Rosenstiel, and Ziesat (1979) suggest several rules

Table 1

Means and SDs (in parentheses) of Logical Memory performance reported as normative or control data. Some values have been recomputed based upon information provided in the reports to reflect average performance on the two LM passages.

Study	Age	*N*	Performance	Note
Bachrach & Mintz (1974)	32.7 (12.9)	42	10.6	3
Cauthen (1977)	30-39	14	7.1	1
Charter (1981)	47 (16.7)	122	8.3 (3.2)	5
Cohen (1950)	29.1 (5.7)	81	8.9 (3.6)	4
Crosson et al. (1984)	53.9	23	8.4 (2.7)	2,5
Dodrill (1978)	27.3 (8.4)	50	11.3 (3.8)	1
Hulicka (1966)	30-39	53	8.0 (3.0)	5
Ivinskis et al. (1971)	16-18	44	7.7 (2.1)	1
Kear-Colwell & Heller (1978)	<35	56	13.6 (3.0)	1
Kear-Colwell & Heller (1978)	>35	60	10.4 (3.1)	1
Kljaic (1975)	46.3 (10.9)	18	5.4 (2.7)	2
Osborne & Davis (1978)	30-39	61	9.4	6
Prigatano (1977)	36.4 (R=20-59)	26	8.6 (2.4)	3
Russell (1975)	36.5 (14.2)	30	11.6 (3.0)	2,5
Wechsler (1945)	20-25	50	9.3 (3.1)	1
Wechsler (1945)	40-49	46	8.1 (2.5)	1

1 = Normal Volunteers
2 = Negative Neurologic Patients
3 = Psychiatric Patients Without Neurologic Evidence
4 = Psychiatric Patients (Nonpsychotic) Without Neurologic Evidence
5 = Medical Patients Without Neurologic or Psychiatric Complaints
6 = Mixed Population of Neurologic and Psychiatric Patients

with which full or half-credit response can be scored, and more recently, Schear (1985) has provided scoring examples that are similar to those used with the WAIS Vocabulary subtest. As these reports indicate, explicit criteria have provided greater scoring reliability, and with sufficient testing, satisfactory normative information can ultimately be obtained.

Even with explicit scoring criteria, a concern regarding adequate construct validity remains. For example, Power et al. (1979) propose that half-credit responses be given if articles, adjectives, or adverbs are omitted since "according them a full credit of one point would equate them with exact reproductions... [and]... half-credit reflects a recognition of the essential accuracy of the response while penalizing to some extent the tendency to modify the idea" (p. 344). However, the process of learning new material is an active one, and involves transformation of the raw input and its incorporation into an existing cognitive schema or framework and is implicitly ignored when verbatim responses for prose passage are required. People abstract the essential features of a story and

retain those features rather than the surface structure of the passage in which those features are embedded (e.g., Bartlett, 1932; Brandsford & Franks, 1971; Cofer, 1965; Pompi & Lachman, 1967). Current memory theorists believe that memory is determined by selection, abstraction, interpretation, and integration (see Alba & Hasher, 1983). Therefore, the response of "kids" for "little children" not only indicates that the main idea has been retained, but since it has been transformed, may actually be a better response than a verbatim one since it indicates that the subject has not simply retained an item, but has assimilated it into a representational framework.

In addition to lack of scoring criteria for these paragraphs and the variability in the instructions used by different neuropsychologists, the paragraphs themselves are dated, and consequently, not appropriate for today's subjects. The phrase "made up a purse" is no longer an idiom in American speech, whereas the phrase "liner" is frequently misinterpreted by many younger subjects to refer to an airplane or "airliner". According to Schear's (1985) scoring examples, luxury liner would receive full credit, passenger liner would be scored a half-credit response, while airliner would earn no credit. In each of these examples, it is clear that the information encoded has been guided by different schemata, yet, some responses merit memory credit while others do not. Thus, some subjects are penalized for possessing somewhat different knowledge frameworks with which to interpret information.

Delayed Recall. The inclusion of a delayed-recall component is a genuine improvement since this approach examines retention of material over time, shifting the test from one of comprehension or encoding to one of long-term retention. Although this approach provides valuable information, it is only a partial solution to the problems under consideration since all of the difficulties associated with scoring the paragraphs discussed above are still pertinent. For example, if the response "Annie was a cleaning lady and got mugged. Her kids were hungry so the cops gave her some money" were given, by strict verbatim scoring criteria, no credit would be obtained. If this same response is given after a one-half hour delay, again no credit would be earned, yet the subject would have certainly demonstrated constancy of information retention over time.

Problems specific to the delayed-recall procedure involve the issue of prompting. Russell (1975) states that, if the subject cannot remember the stories at all, a prompt is provided by asking "do you remember a story about a washer woman?" (p. 803). The paragraph is then scored "according to Wechsler's instructions, except that the item used to remind the subject is not counted in the second score" (p. 803). However, if the information of interest is the *retention* of material over time, and not the ability to *retrieve* information (free recall), then the scoring procedure should not penalize those individuals who initially have difficulty remembering the main theme of the story, but who, once prompted (cued), can retrieve information about the story. By attempting to examine both retrieval and retention, which are two distinct operations, neither is genuinely assessed. The nature of the prompts themselves should also be carefully con-

sidered, since prompts may be either semantic cues that summarize the story, or associative cues that refer to specific details from it. Accordingly, one may bias the subjects' retrieval depending on the type of cue used and thus affect the patient's score. It is amusing to note that although one full point is subtracted from a patient's score if a prompt is required, if the prompts were themselves scored according to verbatim criteria, no credit would be received, and if scored according to half-credit criteria, would only receive partial credit.

The Boston Revision of the WMS implicitly requests a verbatim response for the LM paragraphs by asking "I want you to tell me everything I said to you." Again, no scoring criteria are presented. However, the Boston version contains an immediate-recall probe asking specific questions regarding story elements (e.g., What kind of work did this person do? Did this person have a family?), attempting to assess retention rather than the ability to retrieve information spontaneously. In the delayed-recall condition, however, it is not indicated whether or not prompts are to be given if the subject fails at free recall. A delayed probe for only the Anna Thompson story is administered, although the format for this probe is multiple choice. A multiple choice format, presumably, was not chosen for the immediate-recall probe in order that additional information provided to the subject with multiple choice would not confound the delayed-recall performance. However, changing the format makes direct comparisons across time more difficult. Moreover, it is likely that delayed-recall performance might be aided with the immediate-recall probe since it provides the opportunity for rehearsal. Therefore, available normative information for delayed recall is not necessarily comparable for the Russell and the Boston revisions. That is, the immediate probe, if it in fact provides rehearsal opportunity, should lead to higher scores for delayed performance and would suggest better retention if compared to norms obtained without probe questioning.

Practical Utility. The above criticisms aside, the LM subtest can still be an effective measure of verbal memory. A clinical neuropsychologist, routinely administering this scale, develops a "feel" for what type of responses are suggestive of either normal or impaired memory abilities. This paper is not intended to discuss the strengths and limitations of the psychometric vs. clinical-intuitive approach to evaluation. It is sufficient to note that, when a patient's memory functioning is determined by clinical acumen and experience, then the particular choice of prose passages used is irrelevant. That is, the LM passages do not provide unique information that could not be obtained by a similarly experienced clinician using different stories that had been routinely used by him or her. There is nothing wrong with this approach; however, it does not fulfill the requirements for objective psychological testing and should not be presented as though it does.

We are not attempting to argue that this test has no value. However, for performance at questionably impaired levels, we must first be sure that low performance is not purely an artifact of our testing procedure. Although experienced neuropsychologists will not examine LM scores out of context, perhaps

making some compensation in describing the patient's memory in the clinical report, it is reasonable to expect that these compensations be articulated so that scoring modifications can be developed and made explicit in future revisions of this or other tests or prose passage recall (e.g., Power et al., 1979). In the interim, we suggest that all published reports of LM performance make explicit the scoring procedures used and whether or not verbatim recall is explicitly requested.

Visual Reproduction

The Visual Reproduction (VR) subtest is the only attempt by Wechsler to include a measure of nonverbal memory. In this subtest, subjects are presented with geometric designs for 10 s, and immediately following their presentation, they are requested to draw the designs from memory. In contrast to LM, explicit scoring criteria have been provided which have resulted in greater scoring reliability and greater consistency in reports of normative performance (see Crosson, Hughes, Roth, & Monkowski, 1984).

Unfortunately, factor-analytic studies of this subtest have indicated a principal factor loading in a visual-perceptual-motor ability factor, and only secondarily on memory (e.g., Ivinskis, Allen, & Shaw, 1971; Larrabee, Kane, & Schuck, 1983; Larrabee, Kane, Schuck, & Francis, 1985). This is consistent with some earlier work performed on the Benton Visual Retention Test (Benton, 1962). For example, Silverstein (1962, 1963) found high correlations between copying and immediate-delay performance, and views this test as a measure of visual perception and visual motor ability. As Trahan and Larrabee (1984) have pointed out, however, delayed visual-reproduction performance is more closely related to memory than perceptual motor skills in normal subjects. Further, delayed VR performance has been shown to be superior to immediate recall for discrimination of right unilateral temporal-lobe seizure activity (Delaney, Rosen, Mattson, & Novelly, 1980).

Since immediate reproduction is plagued by concerns regarding construct validity, the only genuine measure of figural memory performance appears to be the delayed component and is included in both the Russell and Boston revisions. However, it must be remembered that visual motor and constructive abilities may be selectively impaired without a corresponding deficit in visual spatial memory. Therefore, recognition assessment should be routinely performed even in the absence of frank visual constructive deficits in order to insure that retention and/or retrieval is not confounded with construction difficulties. The Boston Revision of the WMS adequately addresses this important issue. Immediately following the standard VR administration, patients are presented with a recognition task in which the correct design is displayed with four similar but slightly distorted distractors. After each of the four designs is selected, the patient is asked to copy the designs to examine potential constructional difficulties.

In addition to the standard delayed condition, the Boston version also

includes multiple-choice recognition and a matching task requiring the patient to select the correct design from distractor items with the original stimulus present. As with the delayed recall for the Boston modification of LM, caution must be exercised if norms using the Russell version are used given the additional exposure with the immediate multiple choice and copy performance in the Boston procedure.

As with the delayed recall for LM, the scoring compensation required if the subject cannot immediately remember the designs is problematic. Russell states that, if prompting is required (e.g., the picture with the flags), then one point should be subtracted from the patient's score. Again, this confounds free recall with long-term retrieval.

Additional Boston Modifications

The Boston Revision of the WMS has altered the procedure in several additional ways. The Paired-Associate Learning (PAL) subtest of the WMS is a standard paired-associate learning task consisting of three trials. In each trial, 10 word pairs are read to the patient, six of which are easy associations (e.g., north-south) with the remainder being difficult ones (e.g., obey-inch). A subject's performance is measured by summing all the hard associations acquired plus one half the number of easy associations reported. In the Boston version, an additional trial is presented with the pairs reversed; that is, the subject is provided with the second word of the pair and is requested to remember the first. This has the advantage of examining the strength of the association, and tests the ability of the subject to perform an unexpected task. In addition, a delayed free-recall condition is administered in which patients are asked to provide all word-pairs spontaneously. Patients are then cued with the first word for those pairs they failed to recall.

The Cowboy Story (Talland, 1965) is presented as a story comprehension task. Although analogous to LM recall, it differs on one potentially relevant dimension. The patient is able to pace the speed at which the information is provided since it is read by the subject rather than being recited by the examiner. This is an important provision since the ability to encode information depends in part on rate of presentation and since performance of the elderly on at least some tasks is more comparable to that of young adults when the rate of presentation is slowed (e.g., Kinsbourne, 1973). Although each of the 27 memory units must be recalled verbatim in order to be credited, a provision for content ideas is made in which credit is received for synonyms or paraphrases. Probe question are administered for all information not spontaneously recalled.

BUSCHKE SELECTIVE REMINDING PROCEDURE

The selective reminding (SR) procedure is an attempt to ground verbal memory assessment within the context of contemporary information-processing theory (Buschke, 1973; Buschke & Fuld, 1974). Subjects are instructed to learn a list of

words in any order. In contrast to standard serial word tests, however, the subject is "selectively reminded *only* of those items he did not recall in the immediately preceding trial" (Buschke & Fuld, 1974, p. 1019). The rationale for this procedure is that the subject is provided the opportunity to recall words spontaneously, presumably demonstrating recall of items from long-term memory. Consequently, this technique has been widely implemented to characterize the verbal memory deficit in a number of clinical populations (e.g., Caine, Ebert, & Weingartner, 1977; Levin, Grossman, Rose, & Teasdale, 1979; Loring, Meador, Mahurin, & Largen, 1986; Muramoto, 1984; Peters & Levin, 1977, 1979; Thal & Fuld, 1983).

It must be remembered that this is a procedure and not a standardized test. Consequently, it must be borne in mind that the information about the test available in the literature is based upon tests of different word lengths with varying number of trials. Further, recent evidence suggests that, even when the composition of the test words is carefully matched on the basis of frequency and length, the difficulty of the lists is not necessary equivalent (Hannay & Levin, 1985).

Hanney and Levin (1985) report test-retest reliabilities ranging from .48 to .65 which, as the authors recognize, are lower than those which are generally acceptable for psychological tests. Further, they report that in their sample of college students, there was a significant practice effect with repeated administration of alternative forms for most of the SR-dependent measures. This practice effect suggests that the ability to learn how to perform a complex task, and not exclusively the ability to remember words, may be accounting for the group differences in research using clinical populations. It is therefore possible that these factors are partially responsible for observed deficits in patients' performances rather than an impairment of memory alone. This point is clearly of paramount importance since SR is becoming a popular technique to investigate the efficacy of pharmacologic intervention (e.g., Mohs, Tinklenberg, Roth, & Koppell, 1980; Yesavage, Lierer, Becker, & Holman, 1980). Researchers have the luxury of administering multiple baseline conditions in order to arrive at a stable level of performance, as is a requirement in single-case experimental designs (see Hersen & Barlow, 1976). However, since the earlier administrations are confounded by task novelty and complexity, SR should not be used clinically as the sole measure of verbal memory.

The SR procedure has found widespread appeal since it purports to parcel verbal memory into long-term storage (LTS), long-term retrieval (LTR), consistent long-term retrieval (CLTR), and short-term recall (STR). However, studies employing the SR procedure have typically found high correlations for memory performance measures in both clinical and control samples (i.e., total recall, LTS, LTR, CLTR), suggesting that these measures are assessing similar constructs (Kenisten, cited in Kraemer, Peabody, Tinklenberg, & Yesavage, 1983; Loring et al., 1986). Further, although the attempt to isolate different components of memory is admirable, the distinction between long-term storage and

retrieval is arbitrary. According to Buschke's definition, a word has entered LTS if it has been successfully recalled on two consecutive trials. Therefore, by definition, and definition only, failure to recall is due to retrieval failure. Just as plausible and conceptually appealing, however, is that these "memory traces" have been stored in a weak or degraded form, and that, through the process of additional repetitions from the examiner, the word is encoded more deeply and efficiently, analogous to the way in which a complex motor sequence such as playing music on an instrument is slowly learned with repetition. Therefore, operationally defined retrieval may have in fact little to do with retrieval itself. Although SR appears to offer unique information regarding short- and long-term memory, this distinction can be made using traditional word lists by analyzing both primacy and recency effects (e.g., Brooks, 1972; Glanzer & Cunitz, 1966), or by the number of words, either presented by the examiner or recalled by the subject, between presentation and recall for any particular word (Tulving & Colotla, 1970).

In conclusion, there are important limitations in the use of SR as a technique to assess verbal memory. The standardization for any particular form is limited, forms designed for equivalence differ in their level of difficulty, and there appears to be a significant practice effect with repeated administration. Finally, the operational definitions of storage and retrieval have dubious construct validity.

BRAIN DAMAGE AS A CRITERION

In addition to the psychometric problems of standardization and reliability and the theoretical consideration of construct validity, a problem with these tests is that they have been selected on the basis of their ability to detect brain damage. For example, Russell (1975) states that the subtests selected from the WMS were chosen on the basis of their ability to discriminate mild brain-damage from a mixed group of non-brain-injured subjects. This was obviously important prior to the widespread availability of cerebral imaging techniques. However, since the function of neuropsychologists has shifted, this basis of test selection now assumes a secondary status.

The first decision to be made by neuropsychology is to determine which memory abilities are necessary for independent living and ought to be assessed during the evaluation. Then, on the basis of cognitive theory, operations that mediate the abilities of interest must be defined. Finally, tests must be constructed to assess those operations independently of brain pathology. Ideally, performance on these tests will correlate with different types of cerebral pathology. However, we should not define our constructs of cognitive operations based solely on those tests sensitive to brain damage since sensitivity to brain damage depends on what type of population is included as the criterion group.

The California Verbal Learning Test (CVLT; Delis, Kramer, Ober, & Kaplan, in press), for example, is a memory scale that illustrates strong theoretical

grounding in cognitive psychology. Patterned after Rey's Auditory Verbal Learning Test (see Lezak, 1983), the CVLT requires the subject to learn a list of 16 words over five trials in a standard free-recall procedure. The words are from four semantic categories (i.e., spices and herbs, fruits, tools, and clothing). After the fifth trial, a second shopping list is administered and following its free recall, the subject is asked for spontaneous recall of the initial list. For the CVLT, it was decided what information regarding memory performance would be beneficial for clinical assessment, and those techniques available in cognitive psychology were selected (e.g., proactive and retroactive interference, semantic cuing, and prototypic response distractors for recognition memory). In addition, the use of a distractor list allows for the examination of prior list intrusion errors, which may be an important aspect of a patient's neuropsychological status. For example, both alcoholic Korsakoff and Alzheimer's disease patients can be distinguished from those with Huntington's disease by the presence of a higher number of intrusion errors (Butters, 1985).

A related issue to "brain damage as a criterion" is the issue of selective hemispheric damage assessment. With the Russell modification of the WMS, for example, there is a temptation to equate these tests as equivalent on all cognitive dimensions except for the verbal vs. figural distinction. That is, they may be treated as equivalent tests assessing the memory capacity of the left and right hemisphere. However, these tests differ from each other on a number of other potentially relevant dimensions. For example, LM is presented auditorially whereas VR is, by definition, a visual test. As such, differences in basic sensory functioning may produce an apparent difference in memory functioning. Further, the response mode of the two subtests differs. Clearly different cognitive operations mediate verbal recall and still others (i.e., visual constructive) are necessary for reproduction of geometric figures. Therefore, in addition to differences in basic sensory functioning, differences in other nonmemory skills may confound the results. In addition, the differences in the scoring procedures and reliability which, as previously mentioned, exist between these two subtests, may produce an apparent difference in verbal vs. figural memory functioning.

More importantly, LM is a task that is typically experienced in everyday life, whereas VR is a novel task not normally encountered. Therefore, differences in familiarity with the two testing procedures may produce differences that are interpreted only in terms of memory abilities. VR is likely to load heavily on what Cattell (1963) has termed "fluid intelligence," that is, the ability to acquire and manipulate new or novel information. An example of the possible misinterpretation of performance on these two subtests is seen in Logue and Wyrich (1979). These authors concluded on the basis of their aging and dementia study, that the right hemisphere may be more susceptible than the left hemisphere to aging. However, as correctly pointed out by Russell (1981), this is probably a psychometric artifact since the figural test has such a strong fluid intelligence component. Fluid intelligence is highly susceptible to even mild cerebral patho-

logy, such as that observed with the aging process (e.g., Cunningham, Clayton, & Overton, 1975; Reed & Reitan, 1963).

CONCLUSIONS

The imperfections in memory assessment demonstrated in the previous sections of this paper do not constitute proof that these tests are of no value since the obtained scores are usually interpreted with respect to a neuropsychologist's personal experience. Most psychological tests, in fact, do not meet the psychometric criteria to the degree that would satisfy the purist, and compensation is frequently made.

The question becomes one of selection. Among the available methods to assess memory (tests, interviews, etc.), which method is better? To answer the question, we must consider in what contexts memory tests have been used. If the neuropsychologist is asked to evaluate a patient with possible dementia, then it does not matter if the memory tests employed measure a combination of varied abilities, including freedom from distraction, performance on novel tasks, in addition to some memory component; the test needs only be sensitive to cerebral pathology. For example, the rate of accurate differentiation of organic from psychiatric patients using WAIS scores alone has been reported to be equal or better than using the entire Halstead-Reitan or Luria-Nebraska batteries (Kane, Parsons, & Goldstein, 1985; Watson, Thomas, Anderson, & Felling, 1968). Therefore, the critical concern is the referral question. As far as this goes, the original WMS has been reported to be a poor screening test of brain damage (Prigatano, 1977).

If we wish to determine if memory tests offer the advantage of assessing specific aspects of memory over other assessment techniques, we need only consider their construct validity. As we have detailed above, the major tests do not appear to be entirely valid measures of any construct of memory. This is not to say that with low LM, no statement about impaired memory functioning is warranted. Grossly impaired memory ability will be revealed by LM regardless of how it is scored (i.e., full vs. half-credit) and the need for adequate normative information in unnecessary. However, a set of instruments is necessary to help evaluate memory performance in which the ability, or lack of it, is not immediately clear, for example, as in an individual with a grade school education with possible mild anoxia. It is in these borderline cases that the need for clearly identified constructs and adequate tests of those constructs becomes necessary.

With LM, an additional concern is whether the limited normative information in the literature is better than no normative data at all. Specifically, are clinicians in a better position to evaluate memory by relying, in part, on data collected by other laboratories and clinical services than relying solely on personal experience and personal norms? As it currently stands, it is difficult to interpret the literature since the reader knows neither the instructional set

provided to the subject nor which scoring approach has been employed for any particular study. That is, one does not know if an average score of 12 responses refers to 12 verbatim responses, 12 "close" responses, or some combination of 12 or more full and half-credit responses, or even whether or not a verbatim response was explicitly requested.

As noted by Crosson et al. (1984), approximately half of both Hulicka's (1966) control subjects in the 15-17 year-old age group and Wechsler's (1945) subjects would be classified as impaired using Russell's (1975) norms. Therefore, we believe that, given the differences in administration and scoring, one must rely either on data that has been personally obtained, or alternatively, on norms that have had the instructions and scoring criteria personally verified. We realize, however, that many neuropsychologists will continue to rely on normative information without the above precautions. Therefore, even the pooled data of Crosson et al. (1984), which is based on several reports, should be used with extreme caution since it is neither clear if explicit instructions requesting verbatim recall were given to the subject nor the precise method of scoring used in the different studies.

Future Directions

What might be the proper direction for test development for clinical memory assessment? First, we suggest that for prose passage recall, the passage should be broken down using a propositional representation network to approximate the story grammar and to arrive at some convention of identifying a memory unit (e.g., Kintsch, 1974, Kintsch & van Dijk, 1975; Mandler & Johnson, 1977). Further, some hierarchy is necessary since all information contained in a story is neither equally important nor equally well remembered (e.g., deVilliers, 1974; Kintsch & Keenan, 1973; Thorndyke, 1975). Verbatim scoring or number of words correctly recalled can never be a sound method given that people transform and abstract the essence of what is presented.

Once the story (or stories) is/are analyzed in terms of propositional representation, allowing for sufficient quantification of the subjects' responses, the basic structure of the test is complete. A title should be provided before reading the story to the patient, to facilitate its interpretation as it is being read, and also to serve as a cue for delayed recall. The title should be provided to all patients when the delayed-recall condition is requested.

The clearest advantage to using paired-associates or the CVLT is the objectivity of the scoring criteria. As illustrated in the literature, there is general agreement in the normative values available for PA. Unfortunately, for the CVLT, there are not yet published norms that would allow the scale to be used immediately for clinical purposes, and differences may exist in patient's ability to appreciate, and thereby benefit from, semantic clustering. For PA performance, we suggest that easy and hard associates be considered separately (e.g., Ivinskis et al., 1971) since the decision to treat the hard associations as twice as valuable as the easy associations was not empirically derived.

Given the various models of verbal memory that appear in the cognitive psychology literature, one is not obliged to assess memory as conceptualized in every memory model nor should one necessarily attempt to assess verbal memory using all the various techniques. However, it must be remembered that these different tests assess different aspects of verbal memory and may be selectively impaired. As previously mentioned, LM passages are more appropriate for the assessment of propositional memory whereas list learning is a test of rote memory. In a patient whose planning and organizational skills are impaired (e.g., frontal-lobe pathology), performance on SR may be impaired. However, on a task that is more highly structured with repeated trials and instructions such as a word list, performance may be at more normal levels. Again, we must carefully define the constructs that we are attempting to assess. We have not discussed nonverbal memory as much since there has been limited clinical research in this area, most of which typically consists of VR type tasks. We would hope to see how performance on VR compares to memory for faces (e.g., Warrington & James, 1967) since the latter is ecologically valid and is more comparable to LM in terms of fluid intelligence.

The basic problem associated with the selection of an instrument to assess memory is the choice of an appropriate criterion variable. In the field of rehabilitation neuropsychology, one approach has been to use either real or simulated job performance or even whether or not a patient is employed (e.g., Ben-Yishay et al., 1980). Of course, many factors in addition to memory functioning are essential for vocational activity. An alternative technique has been to examine the relationship between memory test performance and questionnaires of everyday memory, which are completed by either the patient or their family (see Hermann, 1982, for a review). Unfortunately, the relationship between questionnaires and memory test performances has been poor with the highest correlations observed for test performance and ratings completed by family members (Baddeley, Sunderland, & Harris, 1982; Sunderland, Harris, & Baddeley, 1984).

Given the limitations of directly observing patients' performances in the home, it is clear that we must find some more practical method of assessing or observing everyday memory performance in the laboratory and perhaps use this as our criterion. A recent approach developed by the Rivermead Rehabilitation Centre in Oxford, hopefully, will provide one such ecologically valid criterion (Wilson, in press). The Rivermead Behavioral Memory Test (RBMT) consists of a series of items sampling memory of those behaviors that have been frequently reported to be impaired following brain injury (Sunderland et al., 1983). In addition to conventional assessment of memory, the RBMT examines memory performance using a wide variety of everyday tasks. Some tasks are familiar and consist of associating a name with a photograph, and assessing memory for faces, memory for pictures, and orientation. Several novel memory tasks are administered: (1) A personal possession of the subject is borrowed and hidden from view. The subject is told to ask for its return at the end of the testing

session; (2) Patients are informed that they are to ask the examiner a particular question, given to them by the examiner, when an alarm sounds; (3) A short route through five areas of the room is taken, and the patient retraces the route during an immediate and delayed condition; (4) While the route is being shown to the subject, an envelope is left along the route with the subject required to leave the envelope at the same location.

Preliminary results of the RBMT on neurologic patients with and without memory problems, as determined by judgments of occupational therapists based upon patient performance during therapy, suggest that the RBMT does not correlate with intelligence, and correlates most highly with delayed logical memory of the WMS (absolute and percentage of immediate recall) and paired-associate learning. That is, based upon this preliminary report, paired-associate performance and LM, as administered and scored in Oxford, are assessing some abilities that are related to everyday memory performance. This information is encouraging, relating cognitive components to the more pragmatic behavioral components, and we anxiously await replication and extension using other memory tests and procedures.

REFERENCES

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203-231.

Bachrach, J., & Mintz, J. (1974). The Wechsler Memory Scale as a tool for the detection of mild cerebral dysfunction. *Journal of Clinical Psychology, 30,* 58-60.

Baddeley, A., Sunderland, A., and Harris, J. (1982). How well do laboratory-based psychological tests predict patients' performance outside the laboratory? In S. Corkin, K. L. Davis, J. H. Growder, E. Usdin, & R. J. Wurtman (Eds.), *Alzheimer's disease: A report of progress in research* (pp. 141-148). New York: Raven Press.

Bartlett, F. C. (1932). *Remembering: An experimental and social study.* Cambridge: Cambridge University Press.

Benton, A. L. (1962). The visual retention tests as a construction praxis test. *Confinia Neurologica, 29,* 1-16.

Ben-Yishay, U., Diller, L., Rattok, J., Ross, B., Lakin, P., & Cohen, J. (1980). Relationships between aspects of anterograde amnesia and vocational aptitude in TBD patients: Preliminary findings. Paper presented at the 8th annual meeting of the International Neuropsychological Society, San Francisco, CA.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology, 2,* 331-350.

Brooks, D. N. (1972). Memory and head injury. *Journal of Nervous and Mental Diseases, 155,* 350-355.

Buschke, H. (1973). Selective reminding for analysis of memory and behavior. *Journal of Verbal Learning and Verbal Behavior, 12,* 543-550.

Buschke, J., & Fuld, P. A. (1974). Evaluating storage, retention, and retrieval in disordered memory and learning. *Neurology, 24,* 1019-1025.

Butters, N. (1985). Alcoholic Korsakoff's syndrome. *Journal of Experimental and Clinical Neuropsychology, 7,* 181-210.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54,* 1-22.

Caine, E. D., Ebert, M. H., & Weingartner, H. (1977). An outline for the analysis of dementia: The memory disorder of Huntington's disease. *Neurology, 27,* 1087-1092.

Cauthen, N. R. (1977). Extension of the Wechsler Memory Scale norms to older age groups. *Journal of Clinical Psychology, 33,* 200-211.

Charter, R. A. (1981). Prorating the Wechsler Memory Scale. *Journal of Clinical Psychology, 37,* 183-186.

Cofer, C. N. (1965). On some factors in the organizational characteristics in free recall. *American Psychologist, 20,* 261-272.

Cohen, J. (1950). Wechsler Memory Scale performance on psychoneurotic, organic, and schizophrenic groups. *Journal of Consulting Psychology, 14,* 371-375.

Crosson, B., Hughes, C. W., Roth, D. L., & Monkowski, P. G. (1984). Review of Russell's (1975) norms for the logical memory and visual reproduction subtests of the Wechsler Memory Scale. *Journal of Consulting and Clinical Psychology, 52,* 635-641.

Cunningham, W. R., Clayton, V., & Overton, W. (1975). Fluid and crystallized intelligence in young adulthood and old age. *Journal of Gerontology, 30,* 53-55.

Delaney, R. C., Rosen, A. J., Mattson, R. H., & Novelly, R. A. (1980). Memory function in focal epilepsy: A comparison of nonsurgical unilateral temporal lobe and frontal lobe samples. *Cortex, 16,* 103-117.

Delis, D., Kramer, J., Ober, B. A., & Kaplan, E. (in press). The California Verbal Learning Test: Administration and Interpretation. New York: The Psychological Corp.

deVilliers, P. A. (1974). Imagery and theme in recall of connected discourse. *Journal of Experimental Psychology, 103,* 263-268.

Dodrill, C. B. (1978). A neuropsychological battery for epilepsy. *Epilepsia, 19,* 611-623.

Erickson, R. C. & Scott, M. L. (1977). Clinical memory testing: A review. *Psychological Bulletin, 84,* 1130-1149.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351-360.

Hannay, H. J., & Levin, H. S. (1985). Selective reminding test: An examination of the equivalence of four forms. *Journal of Clinical and Experimental Neuropsychology, 7,* 251-263.

Hart, T., & Hayden, M. E. (1986). The ecological validity of neuropsychological assessment and remediation. In B. P. Uzzell & Y. Gross (Eds.), *Clinical neuropsychology of intervention* (pp. 21-50). Boston: Kluwer.

Hermann, D. J. (1982). Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin, 92,* 434-452.

Hersen, M., & Barlow, D. H. (1976). *Single case experimental designs: Strategies for studying behavior change.* New York: Pergamon Press.

Hulicka, I. R. (1966). Age difference in Wechsler Memory Scale scores. *Journal of Genetic Psychology, 109,* 135-145.

Ivinskis, A., Allen, S., & Shaw, E. (1971). An extension of Wechsler Memory Scale norms to lower age groups. *Journal of Clinical Psychology, 27,* 354-357.

Kane, R. L., Parsons, O. A., & Goldstein, G. (1985). Statistical relationships and discriminative accuracy of the Halstead-Reitan, Luria-Nebraska, and Wechsler IQ

scores in the identification of brain damage. *Journal of Clinical and Experimental Neuropsychology, 7,* 211-223.

Kear-Colwell, J. J., & Heller, M. (1978). A normative study of the Wechsler Memory Scale. *Journal of Clinical Psychology, 34,* 437-442.

Kinsbourne, M. (1973). Age effects on letter span related to rate and sequential dependency. *Journal of Gerontology, 28,* 317-319.

Kintsch, W. (1974). *The representation of meaning in memory.* Hillsdale, N.J.: Lawrence Erlbaum Associates.

Kintsch, W., & Kennan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology, 5,* 257-274.

Kintsch, W., and van Dijk, T. (1975). Recalling and summarizing stories. *Language, 40,* 89-116.

Kljajic, I. (1975). Wechsler Memory Scale Indices of Brain Pathology. *Journal of Clinical Psychology, 31,* 698-701.

Kraemer, H. C., Peabody, C. A., Tinklenberg, J. R., & Yesavage, J. A. (1983). Mathematical and empirical development of a test of memory for clinical and research use. *Psychological Bulletin,* 367-380.

Larrabee, G. J., Kane, R. L., & Schuck, J. R. (1983). Factor analysis of the WAIS and Wechsler Memory Scale: An analysis of the construct validity of the Wechsler Memory Scale. *Journal of Clinical Neuropsychology, 5,* 159-168.

Larrabee, G. J., Kane, R. L., Schuck, J. R., & Francis, D. J. (1985). Construct validity of various memory testing procedures. *Journal of Clinical and Experimental Neuropsychology, 7,* 239-250.

Levin, H. S., Grossman, R. G., Rose, J. E., & Teasdale, G. (1979). Long-term neuropsychological outcome of closed head injury. *Journal of Neurosurgery, 50,* 412-422.

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.

Loring, D. W., Meador, K. J., Mahurin, R. K., & Largen, J. W. (1986). Neuropsychological performance in dementia of the Alzheimer type and multi-infarct dementia. Manuscript submitted for publication.

Logue, P., & Wyrich, L. (1979). Initial validation of Russell's revised Wechsler Memory Scale: A comparison of normal aging versus dementia. *Journal of Consulting and Clinical Psychology, 47,* 176-179.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111-151.

Milberg, W. P., Hebben, N., & Kaplan, E. (1986). The Boston process approach to neuropsychological assessment. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 65-86). New York: Oxford University Press.

Mohs, R. C., Tinklenberg, T. R., Roth, W. T., & Koppell, B. S. (1980). Sensitivity of some human cognitive functions to effects of methamphetamine and secobarbital. *Drug and Alcohol Dependence, 5,* 145-150.

Muramoto, O. (1984). Selective reminding in normal and demented old people: Auditory verbal versus visual spatial task. *Cortex, 20,* 461-478.

Osborne, D., & Davis, L. J., Jr. (1978). Standard scores for Wechsler Memory Scale subtests. *Journal of Clinical Psychology, 38,* 115-116.

Peters, B. H., & Levin, H. S. (1977). Memory enhancement after physostigmine treatment in the amnesic syndrome. *Archives of Neurology, 34,* 215-219.

Peters, B. H., & Levin, H. S. (1979). Effects of Physostigmine and lecithin on memory in Alzheimer's disease. *Annals of Neurology, 6,* 219-221.

Pompi, K. F., & Lachman, R. (1967). Surrogate processes in the short-term retention of connected discourse. *Journal of Experimental Psychology, 75,* 143-150.

Power, D. G., Logue, P. E., McCarthy, S. M., Rosenstiel, A. K., & Ziesat, H. A. (1979). Inter-rater reliability of the Russell revision of the Wechsler Memory Scale: An attempt to clarify some ambiguities in scoring. *Journal of Clinical Neuropsychology, 1,* 343-345.

Prigatano, G. P. (1977). The Wechsler Memory Scale is a poor screening test for brain dysfunction. *Journal of Clinical Psychology, 33,* 772-777.

Prigatano, G. P. (1978). Wechsler Memory Scale: A selective review of the literature. *Journal of Clinical Psychology, 34,* 816-832.

Reed, H. B. C., & Reitan, R. M. (1963). A comparison of the effect of the normal aging process with the effect of organic brain-damage on adaptive abilities. *Journal of Gerontology, 18,* 177-179.

Russell, E. W. (1975). A multiple scoring method for the assessment of complex memory functions. *Journal of Consulting and Clinical Psychology, 43,* 800-809.

Russell, E. W. (1981). The pathology and clinical examination of memory. In S. B. Filskov & T. J. Boll (Eds.), *Handbook of clinical neuropsychology* (pp. 287-319). New York: Wiley.

Schear, J. M. (1985, February). *Utility of half credit scoring of Russell's revision of the Wechsler Memory Scale.* Paper presented at the meeting of the International Neuropsychological Society, San Diego, CA.

Silverstein, A. B. (1962). Perceptual, motor and memory functions in the Visual Retention Test. *American Journal of Mental Deficiency, 66,* 613-617.

Silverstein, A. B. (1963). Qualitative analysis of performance on the Visual Retention Test. *American Journal of Mental Deficiency, 68,* 109-113.

Sunderland, A., Harris, J. E., & Baddeley, A. (1983). Do laboratory tests predict everyday memory? A neuropsychological study. *Journal of Verbal Learning and Verbal Behavior, 22,* 341-357.

Sunderland, A., Harris, J. E., & Baddeley, A. D. (1984). Assessing everyday memory after severe head injury. In J. E. Harris & P. E. Morris (Eds.), *Everyday memory, actions and absent mindedness* (pp. 191-206). London: Academic Press.

Talland, G. A. (1965). *Deranged memory.* New York: Academic Press.

Thal, L. J., & Fuld, P. A. (1983). Memory enhancement with oral physostigmine in Alzheimer's disease. *New England Journal of Medicine, 308,* 720-721.

Thorndyke, P. W. (1975). Conceptual complexity and imagery in comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 14,* 359-369.

Trahan, D. E., & Larrabee, G. J. (1984, February). Construct validity and normative data for some recently developed measures of visual and verbal memory. Paper presented at the 12th Annual Meeting of the International Neuropsychological Society, Houston, TX.

Tulving, E., & Colotla, V. A. (1970). Free recall of trilingual lists. *Cognitive Psychology, 1,* 86-98.

Warrington, E. & James, M. (1967). An experimental investigation of facial recognition in patients with unilateral cerebral lesions. *Cortex, 3,* 317-326.

Watson, C. G., Thomas, R. W., Anderson, D., & Felling, J. (1968). Differentiation of organics from schizophrenics at two chronicity levels by use of the Halstead-Reitan organic brain test. *Journal of Consulting and Clinical Psychology, 32,* 679-684.

Wechsler, D. (1945). A standardized memory scale for clinical use. *Journal of Psychology, 19,* 87-95.

Wilson, B. (in press). *Rehabilitation of memory.* New York: Guilford Press.

Yesavage, J. A., Lierer, V., Becker, L., & Holman, C. (1980). Effect of an ergot compound on cognitive ability and depression. *Journal of Psychiatric Treatment and Evaluation, 3,* 14 9-152.