7.07 Neuropsychological Assessment of the Elderly

JOHN R. CRAWFORD and ANNALENA VENNERI
University of Aberdeen, UK

and

RONAN E. O'CARROLL
University of Stirling, UK

7.07.1 THE AIMS OF NEUROPSYCHOLOGICAL ASSESSMENT 134

7.07.2 SOME SPECIFIC CONSIDERATIONS IN NEUROPSYCHOLOGICAL ASSESSMENT OF THE ELDERLY 135

7.07.3 THE NEUROPSYCHOLOGICAL INTERVIEW 136

7.07.4 BASIC PSYCHOMETRIC ISSUES 137

7.07.4.1 The Rationale of Deficit Measurement in Clinical Practice 137
7.07.4.2 Converting Scores to a Common Measurement Scale 137
7.07.4.3 Reliability 138
7.07.4.4 Reliable vs. Abnormal Differences 138
7.07.4.5 Monitoring Change 139

7.07.5 APPROACHES TO ASSESSMENT IN CLINICAL NEUROPSYCHOLOGY 140

7.07.6 SCREENING TESTS FOR MENTAL STATUS 141

7.07.7 MEASURES OF CURRENT INTELLECTUAL ABILITY 141

7.07.7.1 Use of the WAIS-R with the Elderly 141
7.07.7.2 Analysis of Subtest Profiles 142
7.07.7.3 WAIS-R Factor Scores 144

7.07.8 SPECIFIC METHODS FOR THE ESTIMATION OF PREMORBID ABILITY 145

7.07.9 AGING AND MEMORY 146

7.07.9.1 Introduction to Aging and Memory 146
7.07.9.2 Age-associated Memory Impairment 147
7.07.9.3 Does Memory Functioning in Dementia Represent Accelerated Aging? 147
7.07.9.4 Commonly Used Memory Tests: Implications for use with Elderly Subjects 147
7.07.9.4.1 The Auditory Verbal Learning Test 147
7.07.9.4.2 The California Verbal Learning Test 148
7.07.9.4.3 The Wechsler Memory Scale-Revised 148
7.07.9.4.4 The Rivermead Behavioural Memory Test 149
7.07.9.4.5 The Rey Complex Figure Test 149
7.07.9.4.6 The Recognition Memory Test 149
7.07.9.4.7 The Doors and People Test 149

7.07.10 ASSESSING LANGUAGE DYSFUNCTION 150

7.07.10.1 Naming 150
7.07.10.2 Vocabulary 151
7.07.10.3 Discourse 151
7.07.10.4 Verbal Comprehension 152
7.07.11 ASSESSING VISUOPERCEPTUAL DYSFUNCTION 152
7.07.11.1 Visual Recognition 152
7.07.11.2 Visual Organization 153
7.07.11.3 Visuoconstructional Ability 153
7.07.11.4 Visual Interference 153
7.07.12 ASSESSMENT OF ATTENTION 153
7.07.12.1 Unilateral Neglect 153
7.07.12.2 The Anterior Attention System 154
7.07.13 ASSESSMENT OF EXECUTIVE FUNCTIONS 156
7.07.13.1 Frontal Hypotheses 156
7.07.13.2 Verbal Fluency 156
7.07.13.3 Nonverbal Fluency 157
7.07.13.4 Behavioral Assessment of the Dysexecutive Syndrome 158
7.07.14 THE DIFFERENTIAL DIAGNOSIS OF DEMENTIA VS. DEPRESSION 159
7.07.14.1 Delayed Verbal Recall Impairment as an Indicator of Early DAT? 159
7.07.14.2 Kendrick Cognitive Tests for the Elderly 160
7.07.14.3 Effortful vs. Automatic Processing and Memory 161
7.07.14.4 Is Poor Memory Test Performance in Elderly Depressed Patients an Artifact of Poor Motivation? 161
7.07.14.5 Future Directions 161
7.07.14.6 Recommendations and Conclusions for the Differential Diagnosis of Dementia vs. Depression 162
7.07.15 REFERENCES 162

7.07.1 THE AIMS OF NEUROPSYCHOLOGICAL ASSESSMENT

In the early days of clinical neuropsychology, most neuropsychological assessment was aimed at answering questions concerning differential diagnosis and lesion location. With the rapid advances in brain imaging technology the importance of this aim has diminished; the focus of assessment has moved to identifying the cognitive and behavioral consequences of cerebral dysfunction whether this be for rehabilitative or medico-legal purposes. However, the neuropsychologist continues to play a valuable role in these former areas, most notably in the diagnosis of dementing illnesses. In addition, although most neuropsychologists work in acute neurological/ neurosurgical services or rehabilitation settings, most also take direct referrals from general medical and psychiatric services. As a result of this they are often the first to establish evidence for the presence of a neurological disorder.

Medico-legal assessments constitute a significant part of many neuropsychologists' workloads. Areas of particular relevance to the elderly in which a neuropsychological opinion is sought include issues of guardianship (e.g., is a client competent enough to manage their own financial affairs?) and fitness to hold a driving license. Other areas include personal injury litigation, fitness to plead in criminal cases (i.e., could a client with known or suspected neuropsychological deficits follow a legal argument against her/him and have an appreciation of its import), and evaluation of pleas of diminished responsibility (i.e., did pre-existing neuropsychological deficits impair a client's responsibility for her/ his actions?).

In broad terms, neuropsychological assessment is aimed at answering the following questions:

(i) Which components of the cognitive system are dysfunctional, how severe is this dysfunction, and also, just as importantly, which components have been spared?

(ii) To what extent has there been change in mood, comportment, and personality; to what extent are these likely to be a direct effect of neurological damage as opposed to a psychological reaction to any injury or illness?

(iii) What are the implications of any changes in cognition, mood, and comportment for a client's every day functioning now and in the foreseeable future?

(iv) Based on the pattern of cognitive strengths and weaknesses and changes in mood and comportment, what practical advice can be given regarding the design of any formal rehabilitation and what advice can be given to the client and significant others to help them adjust to any deficits?

These questions are answered by integrating information gained from the medical notes, an interview with the client (and whenever possible a relative or close acquaintance), and the qualitative and quantitative findings from formal neuropsychological testing.

7.07.2 SOME SPECIFIC CONSIDERATIONS IN NEUROPSYCHOLOGICAL ASSESSMENT OF THE ELDERLY

There is a strong relationship between sensory deficits and cognitive function in elderly populations. For example, Lindenberger and Baltes (1994) conducted structural equation modeling of the relationship between age, cognitive test performance, and sensory functioning in an elderly sample (age range 70–103) and found that visual and auditory acuity accounted for 49% of the total and 93% of the age-related variance in cognitive performance. They evaluated a number of causal hypotheses to explain this relationship and concluded that it arose because both cognitive performance and acuity were indicators of the underlying physiological integrity of the aging brain. However, an additional possibility not considered by Lindenberger and Baltes is that sensory deficits impose an additional central processing load, resulting in poorer performance on cognitive tasks. Consistent with this possibility, Rabbit (1990) has reported that elderly subjects with 35–50 db hearing loss could repeat aloud flawlessly words read out to them but showed poor recall relative to age and IQ-matched controls with good hearing. In contrast, they had no such problem when the words were presented visually. This effect can be modeled in young subjects with good hearing when they have to process words heard through white noise. Rabbitt (1968) found that, under these conditions, the words were repeated successfully but recall was poor. It would appear, then, that the extra effort expended on making out the material places demands on processing resources which would otherwise be free for rehearsal or elaborative encoding, etc. This has far-reaching implications for the interpretation of neuropsychological test results in the elderly as evidence that clients are able to successfully to make out the stimulus materials does not rule out the possibility that their sensory deficits, rather than a higher level problem, has impaired their performance.

The presence of sensory loss should also be borne in mind when attempting to evaluate if there is dispositional change resulting from a central organic process. For example, hearing loss can lead to withdrawal from previous social activities because of frustration and the impatience of others.

Test fatigue can be an issue with the elderly and requires avoiding long test sessions. The neuropsychologist should also be sensitive to the possibility of effects arising from the order of test administration when examining test profiles. Clinicians experienced in working with individuals with dementia, particularly those in the early stages, will also be aware of their ability to "lift their game" for short periods sometimes to the frustration of their relatives who have sought an appointment because of concerns over cognitive deterioration.

As the result of more elderly people taking an increased interest in their health and the risks associated with aging, the neuropsychologist can play an important role in establishing objective evidence of adequate performance, thereby providing reassurance to clients.

Prior to 1990, there was a lack of adequate normative data on neuropsychological tests for the elderly. However, as governments have digested the implications of the change in the age structure of their populations, resources have been allocated to large-scale studies such as the Australian Longitudinal Study of Ageing (e.g., Bryan, Luszcz, & Crawford, 1997), the Cambridge Ageing Project (see Huppert, 1994), the Mayo Older Americans Normative Study (e.g., Ivnik et al., 1992), and the Canadian Study of Health and Aging (e.g., Tuokko & Woodward, 1996) among others. As a result, it will soon be possible to have better norms for the elderly than for other age groups.

The conventional approach to norming of tests is to obtain a sample which has been screened carefully to exclude potential participants with any medical conditions which may have a detrimental effect on cognitive performance. The case for an alternative approach has been well made by Huppert and colleagues. For example, Huppert (1994) has argued that "Rather than excluding individuals with risk factors which might affect performance, normalisation samples should be based on unselected population samples and the effects of these risk factors should be examined" (pp. 315–316). These two approaches are best seen as complimentary rather than competitors. Comparison of an individual's performance with conventional norms, or norms based on regression scores derived from demographic variables (i.e., age, education, current or former occupation, etc.), can address the issue of whether there is evidence of acquired impairment in a particular functional domain. The latter approach, presumably using multiple regression equations, can then potentially inform the clinician as to whether such impairment is likely to be unusual for someone with this particular combination of factors in their medical history. It should be noted, however, that this latter approach would be complex and would require very large samples.

7.07.3 THE NEUROPSYCHOLOGICAL INTERVIEW

Conducting a neuropsychological interview demands considerable empathy, tact, and intellectual effort. In many cases, the client will have multiple cognitive and physical disabilities and organically induced changes in mood and personality; they will also be attempting to cope with the often massive interpersonal and economic upheaval arising from their illness or injury. Added to this is the fact that a neuropsychological assessment can be anxiety provoking because of potential threats to self-esteem or because of its medico-legal ramifications; accordingly some clients can exhibit more concern at the prospect of such an assessment than over the lengthy and sometimes painful and bewildering physical investigations they may have to undergo.

As in all clinical interviewing, time must be spent establishing rapport and clarifying the nature and purpose of the investigation. General, open-ended, and nonthreatening questions should be employed, followed by a request for the client to describe any problems they have been experiencing in their everyday life. The client's description of these everyday problems constitutes a principal source for the generation and refinement of clinical hypotheses which will be tested in the course of the examination. Such descriptions and follow-up questions are also crucial in establishing the degree of insight the client has into the nature and severity of their deficits. Lack of insight is a major barrier to effective intervention and a client's successful adjustment. A client may exhibit a cheery disregard in the face of severe deficits coupled with grossly unrealistic expectations concerning a return to a former occupation or lifestyle; in other cases there may be a disproportionate concern with a relatively minor cognitive or physical problem when other deficits have much more serious implications.

Often, the most useful clinical information from the interview is gained by asking clients about their short-, medium-, and long-term goals and how they intend to achieve them. Answers to such questions are relevant to assessing drive, mood, planning ability, and level of insight. Inquiries about how they spend a typical day can also be employed to address these issues which have implications for the approach that should be adopted in any rehabilitation/remediation attempts. As Brooks (1989) states in this latter context, "the examiner is attempting to estimate whether the patient is a passive receptor of disability or an active fighter struggling to achieve a better outcome" (p. 66).

The neuropsychologist has a large agenda to cover in the interview and often will be under time pressure; this necessitates a structured, systematic inquiry. However, most formal neuropsychological tests also provide a structure for the client; the examiner sets out the concrete aims and rules, and largely controls the pace and order of the items administered (see the section on assessment of executive deficits for further discussion and exceptions to this). It is, therefore, important that enough of the interview is minimally structured to determine the extent to which the client initiates and organizes the discussion of topics.

A detailed educational history is important in building a picture of a client's premorbid abilities. The history should include years of schooling (with a check on whether any of these were repeat years), the nature of any tertiary education and any further study (evening classes, day release, etc.), and formal qualifications and grades achieved. The possibility that health, economic, social, or attitudinal factors may have prevented a client achieving their full educational and occupational potential should be carefully explored.

The client should be asked to describe the nature of their previous and/or current occupation with a detailed breakdown of the duties and functions performed. The nature of their work and responsibilities provides additional clues to general premorbid level of functioning but also an indication of specific premorbid cognitive skills that are likely to have been strongly developed/highly practiced. The clinician should inquire if there have been any recent changes in living circumstances, and if relevant, the work environment (i.e., introduction of new technology, reorganization of duties) as such changes can often be the catalyst which exposes problems in an individual who had previously been coping despite diminished cognitive resources.

Lezak (1995) sets out a number of fundamental questions which must be addressed prior to formal testing; these include: does the client understand the reason for referral and do they have specific questions of their own they want answered? Do they understand the uses to which the information gained from the examination may be put and who will and will not have access to it? (of particular importance in medico-legal contexts); do they know how and when feedback will be provided? Finally, do they appreciate that the assessment will largely be concerned with cognitive functioning (misperceptions abound) and are they aware of the general nature of the tests to be employed? In individuals with severe deficits the answers to these questions cannot be fully in the affirmative and it is therefore important that their relatives or guardians are briefed.

7.07.4 BASIC PSYCHOMETRIC ISSUES

7.07.4.1 The Rationale of Deficit Measurement in Clinical Practice

Attempting to detect and quantify cognitive deficits in the individual case is problematic because of the wide variability in cognitive abilities within the general population. Scores on neuropsychological measures which are average or even above average can still represent a significant impairment for an individual of high premorbid ability (and have serious implications for return to a previous lifestyle or occupation). Similarly, test scores which fall well below the mean do not necessarily reflect an acquired impairment for an individual whose premorbid resources were modest. Because of this, normative comparison standards are of limited utility in neuropsychological assessment and must be supplemented by individual comparison standards when assessing acquired deficits (Crawford, 1992; Lezak, 1983; Walsh, 1991). Ideally, this individual comparison standard can be obtained from psychological test scores obtained in the premorbid period. However, this is rarely a viable option; the amount of routine psychological testing conducted varies greatly between countries so that many individuals may have had no prior formal testing. Even where such test results exist they are often difficult or impossible to obtain, the content of the tests may have limited relevance, or they may have been administered so long ago that they are of questionable value. Because of these difficulties clinicians normally have to settle for an individual comparison standard which is based on a client's current performance. The explicit rationale here is that (i) cognitive ability measures are almost invariably positively correlated, therefore performance on one measure allows some level of prediction of performance on another, and (ii) some abilities will be preserved, or relatively so, following most neurological injuries or illnesses. Thus, the areas in which a client has performed best are used as the standards (i.e., estimates of premorbid ability) against which to compare performance on other measures. Large discrepancies between measures are taken as indicators of the presence and severity of acquired impairment (Lezak, 1991).

7.07.4.2 Converting Scores to a Common Measurement Scale

In constructing a neuropsychological profile of a client's strengths and weaknesses, most clinicians use instruments drawn from diverse sources. These instruments will differ from each other in the measurement scale used to express test scores; for some instruments no formal scaling will have been developed so that clinicians will be working from the means and standard deviations (SDs) of the raw score from normative samples. The process of assimilating the information from these tests is eased greatly if the scores are all converted to a common scale of measurement.

Converting all scores to percentiles has the important advantage that percentiles are comprehended easily by other health workers. However, because such a conversion involves an area transformation (i.e., the difference between a percentile score of 10 and 20 does not reflect the same underlying raw score difference as that between 40 and 50), they are not ideally suited to the rapid and accurate assimilation of information from a client's profile; percentiles are also inappropriate for use with many inferential statistical methods. Z scores do not suffer from these problems but have the disadvantage of including negative values and decimal places which can cause problems in communication.

One option is to convert all scores to have a mean of 100 and SD of 15 (McKinlay, 1992) as tests commonly forming a part of the neuropsychologist's armamentarium are already expressed on this scale, that is, IQs and factor score indices on the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981), memory indices from the Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987), and estimates of general premorbid ability such as the National Adult Reading Test (NART; Nelson & Willison, 1991). The most common alternative is to use *T* scores (mean 50; SD = 10); the meaning of such scores are easy to communicate and are free of the conceptual baggage associated with deviation IQs.

Regardless of which method is used, the clinician must be constantly aware that the validity of any inferences regarding relative strengths and weaknesses is heavily dependent on the degree of equivalence of the normative samples for the test results compared. Although the quality of normative data for neuropsychological tests has improved markedly since the 1980s, there are still tests used in practice which are standardized on small samples of convenience, the representativeness of which are largely unknown. Thus, discrepancies in an individual's profile may in some cases be more a reflection of differences between normative samples than differences in the individual's relative level of functioning in the domains covered by the tests.

7.07.4.3 Reliability

Adequate reliability is a fundamental requirement for any instrument used in neuropsychology regardless of purpose (Franzen, 1989). However, when the concern is with assessing the cognitive status of an individual its importance is magnified, particularly as the demands of clinical practice are such that decisions, must be commonly made based on information from single administrations of each instrument.

Information on test reliability is used to quantify the degree of confidence that can be placed in test scores, for example, when comparing an individual's scores with appropriate normative data or assessing whether discrepancies between scores on different tests represent genuine differences in the functioning of the underlying components of the cognitive system as opposed to simply reflecting measurement error in the tests employed. In the latter case, that is, where evidence for a differential deficit is being evaluated, it is important to consider the extent to which the tests are matched for reliability; an apparent deficit in function A with relative sparing of function B may simply reflect the fact that the measure of function B is less reliable. This point was well made in a classic paper by Chapman and Chapman (1973) in which the performance of a schizophrenic sample on two parallel reasoning tests was examined. By manipulating the number of test items the schizophrenic sample was made to appear to have a large differential deficit on one or other of the tests.

Particular care should be taken in comparing test scores when one of the measures is not a simple score but a difference or ratio score. Such scores are used quite commonly in neuropsychological assessment; for example, in the assessment of implicit memory functioning a priming score may be derived from the difference between completion of word fragments (e.g., c–pe– / carpet) from a list that has been primed in an earlier study phase and an unprimed list (performance on the unprimed list essentially controls for individual differences in verbal ability). This priming score will have poor reliability because the measurement error associated with the two lists is additive. For example, if both lists had reliabilities of 0.75 and an intercorrelation of 0.6, the formula for the reliability of a difference score (e.g., see Crocker & Algina, 1986) reveals that the reliability of this priming score would be only 0.37. This compares unfavorably with the reliability of most explicit memory tests and raises the danger that the clinician may conclude that an individual has impaired explicit memory coupled with preserved implicit memory when the pattern may simply be an artifact of differences in reliability.

7.07.4.4 Reliable vs. Abnormal Differences

The distinction between the reliability and the abnormality of differences between test scores is an important one in clinical neuropsychology. A difference between scores normally would be considered reliable if it exceeded the 95% confidence interval for the difference, that is, a difference of this magnitude is unlikely to have arisen from measurement error in the instruments. Establishing if a difference is reliable is only the first step in neuropsychological profile analysis. There is considerable intraindividual variability in cognitive abilities in the general elderly population such that reliable (i.e., statistically significant) differences between tests of these different abilities are common. In evaluating the probability that a discrepancy reflects acquired impairment, it is important to consider the abnormality or rarity of the difference, that is, what percentage of the general (i.e., unimpaired) elderly population would be expected to exhibit a difference of this magnitude?

Base rate data, which permits the clinician to assess the abnormality of test score discrepancies, is available for some neuropsychological measures. When such data are not available, a simple formula can be used to estimate the abnormality of any discrepancy from the correlation between the two tests of interest (see Payne & Jones, 1957). It is also possible to assess the abnormality of the discrepancy between a single test and a client's mean scores on a series of measures (Silverstein, 1984).

To highlight the distinction between the reliability and abnormality of a difference take the example of a discrepancy between the General Memory and Attention/Concentration indices of the WMS-R. A discrepancy of 17 points would be necessary for a reliable difference ($p < 0.05$). Based on the average correlation between these two indices in the WMS-R standardization sample ($r = 0.51$), approximately 25% of the general population would be expected to exhibit a discrepancy of this magnitude; to be abnormal (operationally defined for the present purpose as discrepancy which would occur in less than 5% of the general population), a 29-point discrepancy would be required.

The importance of evaluating the abnormality of discrepancies through base rate data or correlational techniques cannot be overstressed. Most clinical neuropsychologists have not had the opportunity to administer neuropsychological measures to significant numbers of individuals drawn from the general elderly population. It is possible, therefore, to form a distorted impression of the degree of intrasubject variability found in unimpaired individuals. The impression is that the degree of normal variability may be underestimated leading to a danger of overinterference when working with clinical populations. For example, Matarazzo, Daniel, Prifitera, and Herman (1988) have reported that in the WAIS-R standardization sample, a sample presumed to be free of neurological or psychiatric disorder, the mean difference between an individual's highest and lowest WAIS-R subtest score was 6.7. For this difference (i.e., the subtest range to be abnormal it would have to exceed 11 Scaled score points. As subtests have an SD of 3, it can be seen that a subtest range of around 3 SDs is not unusual in individuals without acquired impairments. These considerations are just as relevant when profile analysis is carried out with more specific neuropsychological instruments. The WMS-R and WAIS-R are used here as examples simply because of the availability and quality of the relevant data.

7.07.4.5 Monitoring Change

There are many situations in which the neuropsychologist needs to measure potential changes in cognitive functioning in the elderly. Common examples would be to determine whether cognitive decline is occurring in an individual in whom a degenerative neurological process is suspected, or to determine the extent of recovery of function following a stroke or traumatic brain injury. In both these cases, neuropsychological assessment will provide useful information to assist clients, relatives, and other health professionals to plan for the future. Monitoring the cognitive effects of surgical, pharmacological, or cognitive interventions in the individual case is also an important role for the neuropsychologist. Although the aim here is most commonly to determine if there has been any improvement, the possibility of detrimental effects can also be an issue. For example, many drugs can potentially impair cognitive functioning, particularly in the elderly. Anticholinergic agents provide a good example as they are widely used in the treatment of various conditions and yet can have serious negative effects on cognition, particularly memory (Crawford, Besson, & Ebmeier, 1990).

In assessing the effectiveness of a rehabilitation effort it is often possible to obtain multiple repeated measures of an individual's performance before, during, and after intervention. A number of inferential statistical techniques can be used in this situation because there are multiple data points for the different phases (see Barlow & Hersen, 1984 for a general treatment of single-case designs). However, in general clinical practice the neuropsychologist often must come to conclusions about change from only a single retesting. This situation will also arise in rehabilitation settings; although multiple measures may have been obtained on the training task(s), the issue of the generalizability of any improvement is often addressed by comparing single before-and-after scores on related, but separate tests.

Monitoring change on the basis of a single retesting is a formidable task except in cases where the level of change has been dramatic. It is rendered more formidable by the fact that many of the standard instruments used in clinical neuropsychology (e.g., the WMS-R) do not have parallel versions. The clinician must therefore differentiate changes resulting from systematic practice effects and random measurement error from change reflecting genuine improvement or deterioration. Among other complications are the fact that (i) the magnitude of practice effects vary with the nature of the task (e.g., the Performance subtests of the WAIS-R show larger practice effects than their Verbal counterparts), (ii) the length of time that has elapsed between test and retest will influence the magnitude of effects, and (iii) a diminution of practice effects is to be expected in neurological populations given the high prevalence of memory and learning deficits, but the expected diminution is difficult to estimate for individuals.

One approach to dealing with many of these considerable interpretive problems is to gather information from test–retest studies on the relevant neuropsychological tests. Provided that such studies report the test–retest correlations, means, and SDs, regression equations can be constructed to predict performance on retest from scores at initial testing. The predicted scores are then compared with the retest scores actually obtained by the client to assess whether the observed gain or decline significantly exceeds that expected (see Knight & Shelton, 1983; McSweeny, Naugle, Chelune, & Lüders, 1993). The utility of this approach is determined by the extent and nature of retest studies available for a particular test. For example, the test–retest scores on memory tasks for an elderly client with suspected dementia could be compared with estimated retest scores derived from a healthy, elderly sample retested after a similar period to gauge how atypical any decline may be. For some questions test–retest data from a clinical sample may be used. For example, a stroke patient's scores on measures of attention or speed of processing could be compared with estimated retest scores from a stroke sample if the clinician suspects that the extent of recovery is atypical.

7.07.5 APPROACHES TO ASSESSMENT IN CLINICAL NEUROPSYCHOLOGY

Approaches to the assessment process can be characterized as falling on a continuum between what has been referred to as the "fixed, big battery" approach and a flexible, hypothesis-testing approach (Brooks, 1989; Walsh, 1991). In the former, a large comprehensive battery is routinely administered to all clients. This approach is most common in the USA and is exemplified by the Halstead–Reitan battery which consists of a large number of specific neuropsychological measures, for example, tests of tactual performance, sensory extinction, finger tapping, categorization, language functioning, etc., which are often supplemented with a full-length WAIS or WAIS-R and personality inventories (see Reitan, 1986).

In the latter approach, measures are selected to test clinical hypotheses derived from the neuropsychological literature on a client's known or suspected disorder and from information gained from the interview. The process is a dynamic one; the results of preliminary testing are used to test or modify existing hypotheses and generate new ones. As a simple example, screening measures may detect problems with the organization of visual material which are consistent with spatial and/or planning deficits; follow-up measures can then be administered to evaluate these three competing possibilities.

Proponents of the flexible approach acknowledge that there is a potential danger of failing to detect a client's difficulties in some cognitive domains if the clinician focuses too rapidly on testing inappropriate hypotheses (Miller, 1992). However, they argue that the fixed, big battery approach may represent an inefficient use of resources and is simply not an option in some hard-pressed services (Brooks, 1989; Walsh, 1991). Further, because of the time taken to complete an assessment, many of the measures may be administered by psychology assistants rather than fully accredited clinicians. It has been argued that important clinical information may therefore be missed or misinterpreted (Brooks, 1989). In the fixed, battery approach all clients and normative groups are administered the same tests, therefore services employing this approach accumulate substantial data on their measures. Such databases are tremendous clinical assets. However, the very existence of this database, and the resources expended to obtain it, may promote a reluctance to substitute existing tests with newer measures designed to reflect developments in neuropsychological knowledge.

Approaches to assessment can also be characterized by the emphasis placed on qualitative vs. quantitative evidence. The extreme quantitative pole of this dimension is characterized by a strict statistical/actuarial methodology in which comparison of a client's test scores against various cut-offs is the principal focus. At the other extreme, tests are used to reveal qualitative features of a client's behavior, the emphasis being on how a task is approached rather than on any empirical analysis of the test scores obtained. In the service of this end there may be major deviations from standardized test instructions, thus precluding statistical analysis in any case. Although there is no fundamental reason why this dimension should not be orthogonal to the fixed/flexible dimension, in practice they have tended to be correlated. The fixed, battery approach arose from researchers working in the actuarial tradition and the sheer number of measures administered almost demands the employment of actuarial indices to assist in the assimilation of the information obtained.

A strict quantitative methodology could be employed with the flexible hypothesis-testing approach. Indeed, the term hypothesis-testing might be seen as implying this and indeed influential figures in the development of this approach devoted considerable attention to quantification (e.g., Shapiro, 1973). However, many clinical proponents of this approach have tended to emphasize qualitative analysis that is, hypotheses may be expressed and tested in qualitative as well as quantitative terms (Walsh, 1991). It should be stressed that very few clinical neuropsychologists are at the extreme ends of this qualitative/quantitative dimension; qualitative observations are verified with quantitative instruments whenever possible whilst recognizing that much important information is not easily amenable to such verification.

7.07.6 SCREENING TESTS FOR MENTAL STATUS

This section will briefly review screening tests for mental status. All these Scales include questions covering basic orientation for time and place but differ in the extent to which they address other competencies. Among the most common Scales are the Mini Mental State Examination (MMSE; Folstein, Folstein, & McHugh, 1975), the Mattis Dementia Rating Scale (MDRS; Mattis, 1976), and the cognitive section (CAMCOG) of the Cambridge Mental Disorders of the Elderly Examination (CAMDEX; Roth et al., 1986). These instruments produce a global score which summarizes a patient's cognitive competence. They are brief and easy to administer and provide sufficient information for epidemiological studies, or as an index of severity of decline. They cannot be viewed as exhaustive diagnostic instruments.

The MMSE (Folstein et al., 1975) is probably the most frequently used screening Scale and is often used as part of a larger battery for a comprehensive assessment of dementia. It is a very brief and easily administered instrument and takes about 5–10 minutes in total. It tests orientation, information, and visuoconstructive abilities; total scores can range between 0 and 30. In their original study, Folstein et al. reported a high test–retest reliability of 0.83 when retesting was conducted by a different examiner, and 0.89 when the same examiner was used. High test–retest reliability, ranging from 0.8 to 0.95, has been confirmed by later studies (Anthony, Le Resche, Niaz, Van Korff, & Folstein, 1982; Dick et al., 1984). An effect of age on MMSE scores has been reported by several studies (Magni, Binetti, Bianchetti, Rozzini, & Trabucchi, 1996; Pi, Olive, & Esteban, 1994; Tombaugh & McIntyre, 1992). Furthermore, caution is required when interpreting scores obtained from poorly educated individuals. Level of education has been shown consistently to have an effect on MMSE score (Kittner et al., 1986; Kukull et al., 1994; Uhlmann & Larson, 1991).

A useful instrument for the evaluation of the mental status is the Mattis Dementia Rating Scale (MDRS) (Mattis, 1976). This Scale can take about 30–45 minutes to administer, but since items are ordered in descending order of difficulty passes on early items can shorten the time required because of discontinuation rules. The MDRS tests orientation, attention, initiation and perseveration, visuoconstructive abilities, conceptualization, and memory. Split-half reliability for this Scale is high (0.90) (Gardner, Oliver-Munoz, Fisher, & Empting, 1981). Norms are available (Montgomery, 1982); however, according to Spreen and Strauss (1991) the normative sample cannot be considered representative of the elderly population as it contained a disproportionate number of high socioeconomic status (SES), highly educated participants.

As noted, the CAMDEX (Roth et al., 1986) includes a cognitive section (CAMCOG). This Scale contains the items of the MMSE with the addition of some additional coverage of perception, memory, and abstract thinking. A study (Huppert et al., submitted) has examined the psychometric properties of the CAMCOG; the test–retest reliability of the Scale as a whole was very high (0.86), as was its internal consistency (coefficients ranged from 0.82 to 0.89). A study carried out in a group of 222 elderly people reported that CAMCOG scores are affected by age, sociocultural factors, and hearing and visual deficits (Blessed, Black, Butler, & Kay, 1991). The cognitive subScale of the Alzheimer's Disease Assessment Scale (Rosen, Mohs, & Davis, 1984) is used increasingly widely in clinical trials in DAT because of its proven sensitivity to the effects of cholinesterase inhibitors and other drugs (Schneider, 1996). It takes around an hour to administer and includes tests of word recall, orientation, naming, spoken language and comprehension, and constructional praxis.

7.07.7 MEASURES OF CURRENT INTELLECTUAL ABILITY

7.07.7.1 Use of the WAIS-R With the Elderly

The WAIS-R (Wechsler, 1981) continued to be used widely in neuropsychological assessment of all individuals of all ages. Lezak (1988a), for example, describes the Wechsler as, "the workhorse of neuropsychological assessment" and identifies it as "the single most utilized component of the neuropsychological repertory" (p. 38). A number of points should be made about the use of the WAIS-R with the elderly. First, as will be discussed later, many clinicians base their analysis of WAIS-R performance on a client's subtest profile. It is important that age-graded Scaled scores are used in this process rather than the standard Scaled scores which are not corrected for age (Lezak, 1995). Failure to use these former scores can produce a totally misleading picture of individuals' strengths and weaknesses relative to their peers; (see Crawford, 1992 p. 27). Second, as aging is associated with reduced visual acuity, it is regrettable that the stimulus materials for some of the subtests (most notably Picture Completion and Picture Arrangement) are unnecessarily small and the artwork poor.

The most obvious problem in using the WAIS-R with the elderly is that the original norms only cover those up to age 74 years. For clients over this age two sets of independently collected norms are available (Ivnik et al., 1992; Ryan, Paolo, & Brungardt, 1990). The Mayo norms were established on a sample of 512 healthy individuals from Rochester and surrounding areas in the US and covers up to age 97 years. However, many of the total subject pool were under age 74 years and were used to examine the comparability of the MAYO norms with the existing WAIS-R norms. Thus, the norms for those over age 74 years were based on a subsample of 222. The sample was relatively highly educated and had an underrepresentation of ethnic minorities. Two features of these norms are potentially appealing. First the norms were established using the method of overlapping midpoint age ranges described by Pauker (1988); this takes maximum advantage of the data available and, because it permits more age bands to be formed, lessens the sometimes substantial apparent improvement in scores that can occur when an individual moves from one age band to the next. As an aside, it is worth noting that, using the existing WAIS-R norms and any spreadsheet package, it is possible to go one step further and derive continuous norms for those aged under 74 years (see Zachary & Gorsuch, 1985). Second, age corrections are carried out at the subtest level rather than when converting the sum of Scaled scores to IQs. This is the same approach used for the Wechsler children's Scales; why it was not adopted for the adult Scales remains a mystery.

Ryan et al's. (1990) norms were established for two age groups, 75–79 ($n=60$) and 80 and above ($n=70$). Although this sample was somewhat smaller than the Mayo sample, it was recruited to be broadly representative of the US elderly population in terms of those demographic variables that were considered to be most pertinent to cognitive tests. Paolo and Ryan (1995) discuss the pros and cons of both samples in detail. They also provided an example subtest profile for a case aged 75 years (i.e., just outside the standard WAIS-R norms) and compared the results obtained by both sets of elderly norms and by scoring the case as if he were 74 years using the original norms. The original and Ryan norms showed much greater evidence of convergence; the Mayo norms indicated significant deficits on some Performance subtests which were not present in the other two profiles. An additional advantage of the Ryan norms is that base rate data on subtest scatter and other aspects of performance are available (e.g., Ryan & Paolo, 1992).

7.07.7.2 Analysis of Subtest Profiles

Analysis of WAIS-R performance can be based on the IQs, the subtests or scores derived from factor analysis. Authorities on the Wechsler have divided opinions on the usefulness of the summary IQs. Matarazzo and Herman (1985), for example, have suggested that the discrepancy between Verbal and Performance IQs is the best validated Wechsler index of cerebral dysfunction, whereas Lezak (1995) is dismissive of its importance. Many clinicians primarily base their interpretation of the WAIS-R on the pattern of strengths and weaknesses in a client's subtest profile (e.g., Lezak 1995; Walsh, 1991). Lezak (1988b), for example, notes that IQs can obscure clinically important strengths and weaknesses and suggests that "although the traditional scoring scheme for the Wechsler exemplifies the IQ problem, these tests also show us a way out of the problem by providing a profile of (subtest) scores" (p. 359).

A starting position in conducting such a profile analysis is to assess whether any observed subtest differences are reliable (i.e., statistically significant) as opposed to simply reflecting measurement error. The only assistance for this process provided in the WAIS-R manual appears in Table 13 (Wechsler, 1981). This table provides the critical values required for significance when comparing any subtest with any other. It is appropriate to use these critical values if and only if the clinician has made an *a priori* decision to examine the difference between only one pair of subtests out of the 55 potential comparisons that could be made. However, in practice, clinicians will want to compare more than one pair; indeed, when the client's history and behavior at interview provide insufficient grounds for forming clinical hypotheses, such comparisons will be *post hoc*. In effect, all possible comparisons are made in this situation even though the clinician will focus attention on the subtest comparisons which yielded the largest discrepancies.

The principal problem with Wechsler's approach is that it does not introduce any correction for the inflation of the Type I error rate that occurs when multiple comparisons are conducted (Crawford, 1992; Silverstein, 1982). Any attempt to overcome this problem inevitably involves striking a balance between controlling the Type I error rate while retaining reasonable power to reject the null hypothesis of no subtest differences. For example, an unsatisfactory solution would be to produce a modified table in which a Bonferroni correction was applied to derive new critical values reflecting the fact that 55 comparisons were involved. Although such an approach would control for inflation of the Type I error rate, it would result in very low power to detect subtest differences, that is, the Type II error rate (i.e., the rate of false negatives) would be unacceptably high.

Knight and Godfrey (1984) and Silverstein (1982) independently suggested a method which achieves a useful compromise between the need to minimize both Type I and Type II errors; instead of potentially comparing all subtests with each other, each subtest is compared with a client's mean subtest score. In the case of a full-length WAIS-R this reduces the number of comparisons to 11 so that, when the Bonferroni correction is applied to maintain the Type I error rate at the specified level, the loss of power to detect subtest differences is much less acute; 11 appears in the denominator rather than 55. Both authors provided a table which recorded the size of discrepancy between each subtest and an individual's mean subtest scores required for varying levels of statistical significance. This approach, when combined with equivalent tables on the abnormality of any subtest differences (see below), provides a rapid and straightforward means of analyzing a client's profile of strengths and weaknesses and therefore provides information which can be integrated usefully with other quantitative and qualitative information available to the clinician.

Crawford (1997) presented a modified table which corrects a minor error common to both of the original tables. Further tables are provided for use with any of nine short-forms of the WAIS-R, including those specifically developed for neuropsychological purposes (e.g., Reynolds, Willson, & Clark, 1983; Walsh, 1991) and those in which the selection of subtests was guided by their suitability for use with the elderly and other groups in which reduced visual acuity can be a problem (e.g., Britton & Savage, 1966; Crawford, Allan, & Jack, 1992a).

If subtest differences are found to be reliable, it is often appropriate to determine if they are also abnormal, that is, to establish how rarely differences of the magnitude observed in an individual's profile occur in the general population (see Section 7.07.4.4). Many healthy individuals have reliable strengths and weaknesses in their subtest, thus reliable differences do not imply the presence of acquired impairment. Silverstein (1984) provided a table to assess the abnormality of subtest differences. As was the case for the reliability of subtest differences, this table is used to compare each subtest with an individual's mean subtest score. Crawford, Allan, McGeorge, and Kelly (1997) adopted the same method to prepare a set of abnormality tables for use with any of the nine WAIS-R short-forms referred to earlier.

The above methods allow clinicians to test hypotheses based on clinical experience or the research literature on a client's known or suspected neurological disorder. However, the methods themselves make no *a priori* assumptions about the nature of an individual's subtest pattern. In contrast, there have also been attempts in the past to identify fixed, specific subtest patterns for differential diagnostic purposes. Most of these attempts naively aimed to identify "organicity" and hence were based on a unitary concept of brain damage which is totally discredited. Not surprisingly, empirical studies of these profiles, such as Wechsler's "Hold–Don't Hold" pattern, have been consistently disappointing (e.g., Christensen & MacKinnon, 1992; Savage, Britton, Bolton, & Hall, 1973).

A more rational and interesting attempt to identify specific subtest patterns was conducted by Fuld (1984). This approach arose from evidence implicating cholinergic dysfunction in the cognitive deficits seen in dementia of the Alzheimer type (DAT) (e.g., Perry et al., 1978; see Crawford et al., 1990 for a review). Fuld (1984) examined Drachman and Leavitt's (1974) data on the effects of scopolamine on the WAIS performance of healthy adults and reported that the age-graded subtest profile could be typified by the formula $A > B$; $B > C$, $C \leqslant D$, $A > D$ where A is the mean of Information and Vocabulary, B of Similarities and Digit Span, C of Digit Symbol and Block Design, and D is Object Assembly. Fuld reported that around a half of patients with DAT exhibited this profile, whereas it was observed in less than 7% of controls and patients with multi-infarct dementia (MID). These results provoked much interest (see Massman & Bigler, 1993 for a review). Subsequent studies have reported rates of occurrence in DAT ranging from a high of 57% (Brinkman & Braun, 1984) to a low of 7% (Logsdon, Teri, Williams, Vitiello, & Prinz, 1989). Most subsequent data on the specificity of the profile indicated that it occurred with low frequency in the healthy elderly and in neurological disorders other than DAT. However, reasonably high rates of occurrence have been reported in depression (16%) and schizophrenia (15%). A particularly disappointing set of results was reported in the previously mentioned study by Logsdon et al. (1989) in which the 7% rate of occurrence in DAT was exactly equivalent to the rates in the healthy elderly and depressed patients. Gfeller & Rankin (1991) have also reported that the profile was exhibited by 12% of cases of multi-infarct dementia, a figure which is little different from the rate found in their DAT sample (15%).

With knowledge on the extent of intraindividual variability in subtest scores (e.g., Matarazzo et al., 1988), it should not be surprising that the profile has low sensitivity for DAT. For example, it would not be uncommon for individuals to have premorbid Block Design or Digit Symbol scores which were well over one SD higher than their average subtest score; similarly, it would not be uncommon for individuals to have Information or Vocabulary scores which were substantially below their average subtest score. Thus, although these sets of subtests may be differentially affected by the onset of DAT, a massive effect would be required to reverse the premorbid pattern; accordingly, these differential effects often would be obscured in the individual case. In conclusion, the Fuld profile has little to recommend its use in clinical practice. However, it is still of some theoretical interest as scopolamine-treated healthy subjects and DAT cases are apparently the only groups studied to date in which the profile is exhibited in their respective means. This provides some support for the cholinergic hypothesis in DAT. However, it should be noted that there has been no attempt to replicate the original report on the effects of scopolamine. It would also be important to demonstrate that the profile has pharmacological specificity, that is, that a similar profile would not be observed following administration of a benzodiazepine or some other centrally active agent.

7.07.7.3 WAIS-R Factor Scores

An alternative to the analysis of both IQs and subtest discrepancies is to examine scores obtained from factor analysis of the WAIS-R (Atkinson, 1991; Canavan, Dunn, & McMillan, 1986; Crawford, Allan, Stephen, Parker, & Besson, 1989a; Kaufman, 1990). Although this suggestion has a long history (Maxwell, 1960), factor scores have not been employed widely in clinical practice or research. Factor analyses of the WAIS and WAIS-R commonly have extracted three factors: a verbal (V) factor on which Information, Vocabulary, Comprehension, and Similarities have high loadings; a perceptual–organization (PO) factor defined by high loadings from Block Design and Object Assembly (other Performance subtests have more modest loadings); and a third factor termed attention/concentration (A/C) or freedom-from-distractibility consisting of Arithmetic, Digit Span, and often Digit Symbol (Crawford et al., 1989a; Leckliter, Matarazzo, & Silverstein; 1986). The arguments in favor of using factor scores when interpreting WAIS-R performance are many and compelling. First, the IQs do not have optimal construct validity, the subtests were allocated to a Verbal and Performance Scale on intuitive grounds. Second, the factor structure of the WAIS-R appears to be very robust in that it is consistently extracted from samples from countries outwith the USA. Most crucially, it emerges when the WAIS-R scores of clinical samples are factor analyzed (e.g., Atkinson, Cyr, Doxey, & Vigna, 1989). Third, the reliabilities of factor scores are much higher than those of the individual subtests from which they are formed. Fourth, although authorities on the Wechsler have laid stress on the analysis of subtest scatter, the WAIS-R could be viewed as primarily providing broad indicators of current functioning against which more specific neuropsychological measures are compared. Again, the superior construct validity of the factorially derived composites suggest they are better suited to provide these broad intraindividual comparison standards than the IQs.

Crawford, Johnson, Mychalkiw, and Moore (in press-a) have examined the clinical utility of IQs, indices of subtest scatter, and factor scores in discriminating between healthy ($n = 356$) and head-injured samples ($n = 233$). The scatter indices employed were the Profile Variability Index (McLean, Reynolds, & Kaufman, 1990), which simply records the variance of an individual's subtest scores, and The Mahalanobis Distance Index (Burgess, 1991); this latter index factors in the correlation between subtests (therefore, a discrepancy involving a subtest which has a relatively low average correlation with the other subtests is given less weight than a discrepancy involving a subtest with a high correlation). Factor scores were significantly better than both the IQ Scales and the scatter indices at discriminating between healthy and impaired performance; the scatter indices essentially performed at chance levels. The extent to which these findings will generalize to neurological conditions more directly associated with the elderly (e.g., DAT, stroke, etc.) remains to be determined.

The lack of a rapid method of factor scoring the WAIS-R, combined with the absence of data with which to assess the reliability and abnormality of discrepancies between factors, acted as barriers to the effective use of factor scores in clinical practice. Both these limitations have simultaneously been rectified in a very useful paper by Atkinson (1991).

7.07.8 SPECIFIC METHODS FOR THE ESTIMATION OF PREMORBID ABILITY

The National Adult Reading Test (NART) (Nelson & Willison, 1991) is the test most widely used to estimate premorbid ability. The NART is a single word, oral reading test consisting of 50 items. All the words are irregular, that is, they violate grapheme–phoneme correspondence rules (e.g., chord). Because the words are irregular, intelligent guess work should not provide the correct pronunciation, therefore the test taps previous word knowledge; as the test only requires the reading of single words, clients do not have to analyze a complex stimulus and it is argued that the test therefore makes minimal demands on current cognitive capacity (Nelson & O'Connell, 1978). Development of the NART arose from the clinical observation that oral reading is commonly preserved in dementia (whereas reading for meaning is commonly impaired). However, the test is used to estimate premorbid ability in a wide range of conditions.

To qualify for use as a measure of premorbid ability, a test must fulfill three criteria (Crawford; 1992; O'Carroll, 1995). First, as with any psychological test, it must possess adequate reliability. The NART has high split-half reliability/internal consistency, test–retest reliability, and inter-rater reliability (Crawford, 1992). Second, it must have high criterion validity. The NART is normally used to provide an estimate of general premorbid IQ against which current performance on the WAIS-R is compared. Thus, to meet the second requirement, the NART must be capable of predicting a substantial proportion of IQ variance. There has been some confusion in the literature on how to examine this aspect of putative measures of premorbid ability (e.g., Klesges, Wilkening, & Golden, 1981), so it is worth noting that it must necessarily be studied using unimpaired rather than clinical samples. In a clinical sample, NART performance and performance on the criterion (e.g., the WAIS-R) will commonly become dissociated, indeed it is the presence of such a dissociation (i.e., a large discrepancy between estimated premorbid ability and current ability in favor of the former) that is used to infer impairment. In most studies using the WAIS or WAIS-R as the criterion variable the NART predicted well over 50% of IQ variance. For example, in one study the NART predicted 66% of WAIS IQ variance in a sample of 151 healthy subjects (Crawford, 1992).

The final criterion for a putative measure of premorbid ability is that test performance be resistant to neurological or psychiatric disorder. NART performance appears to be largely resistant to the effects of many neurological and psychiatric disorder, for example, depression, acute schizophrenia alcoholic dementia, closed head injury, and Parkinson's disease (Crawford, 1992; O'Caroll, 1995) and compares favorably with previously suggested alternative measures of premorbid ability such as the Vocabulary subtest of the WAIS or WAIS-R (e.g., Crawford et al., 1988). Mixed findings have been found in samples with probable DAT. O'Carroll, Baike, and Whittick (1987) readministered the NART to a sample of dementia cases after one year and found no change in performance despite a significant decline on measures of dementia severity. Crawford et al. (1988) found that the NART performance of DAT cases did not differ significantly from matched controls despite the presence of severe deficits on other cognitive measures and the presence of marked morphological and blood flow abnormalities on brain imaging. Stebbins, Wilson, Gilley, Bernard, and Fox (1988), in contrast, found that NART performance was impaired significantly in DAT cases classified as moderate or severe, although impairment was not in evidence in mild cases. Subsequent studies have continued to produce conflicting results in DAT (e.g., Patterson, Graham, & Hodges, 1994; Sharpe & O'Carroll, 1991) but it is becoming increasingly clear that NART performance can be impaired substantially in many cases of moderate to severe dementia. Any findings of impaired NART performance poses a threat to the validity of this approach. However, the practical implications of impaired performance in cases of severe neurological disorder are not as serious as they may appear; in such cases the presence of deficits is unfortunately only too obvious, thereby largely obviating the need for the NART or a similar instrument to assist in its detection and quantification.

Most research on the NART's ability to estimate premorbid ability has used scores on IQ tests as the criterion variable. Although this is in keeping with the notion of obtaining an estimate of general level of premorbid functioning, the NART also has the potential to provide estimates of premorbid functioning for other WAIS-R indices and for more specific neuropsychological tests. For example, Crawford, Moore, and Cameron (1992c) built a regression equation which can be used to estimate premorbid performance on the FAS verbal fluency test (see Section 7.07.13). Similarly Crawford, Obonsawin, and Allan (in press-b) have provided an equation which uses NART and age to estimate premorbid scores on the Paced Auditory Serial Addition Test (PASAT; Gronwall, 1977).

Schlosser and Ivison (1989) derived a regression equation which incorporated age and NART scores to estimate performance on the Wechsler Memory Scale (WMS). They reported that the discrepancy between current WMS performance and estimated premorbid performance was a highly successful index for the detection and quantification of memory dysfunction in DAT. O'Carroll et al. (1994) extended this work by comparing a group of DAT patients, a group of depressed patients, and a group of healthy controls on the discrepancy between NART-estimated premorbid intellectual ability and current mnemonic ability as assessed using the various indices of the WMS-R. While large mean differences were demonstrated between groups, the degree of overlap between the DAT and depressed group was large, leading them to conclude that none of the NART/WMS-R discrepancy quotients were significantly discriminating between depression and DAT to be useful in clinical practice.

A proposed modification to the NART is the Cambridge Contextual Reading Test (CCRT; Beardsall & Huppert, 1994). As its name suggests, the NART words are embedded in sentences to provide context for the examinee. Baddeley, Emslie, and Nimmo-Smith (1993) proposed another alternative reading test which they termed the Spot The Word Test (STW). STW is a lexical decision task in which the examinee has to identify the legitimate words from a series word/pseudoword pairs (e.g., stamen/floxid). Both of these new tests have considerable potential as alternatives to the NART. However, their use is mainly limited to research applications as regression equations have yet to be developed to provide estimates of premorbid ability for individuals and there is insufficient evidence on their resistance to impairment.

An entirely different approach to the estimation of premorbid ability uses demographic variables (e.g., years of education, occupational classification) as predictors. This has the advantage that the estimates it provides are entirely independent of current level of functioning (care should, however, be taken to assess whether prodromal difficulties could have led to a client's failing to achieve their educational or occupational potential). As evidence of impaired NART performance in various clinical conditions accumulates, the demographic approach may be a useful alternative method in cases where use of the NART would be inappropriate. However, the obvious disadvantage of the demographic approach is its modest criterion validity. Where WAIS or WAIS-R scores have been used as the criterion variable, regression equations based solely on demographic variables account for only 30–54% of the variance (Barona, Reynolds, & Chastain, 1984; Crawford & Allan, 1997; Crawford et al., 1989b; Wilson et al., 1978). This compares unfavorably with the NART and its variants (Crawford, 1992; O'Carroll, 1995).

7.07.9 AGING AND MEMORY

7.07.9.1 Introduction to Aging and Memory

It is a well-established phenomenon that as humans age, memory complaints increase. These subjective complaints are reflected in objective measurements of memory test performance. For example, Davis and Bernstein (1992) examined verbal memory performance, as assessed by the Auditory Verbal Learning Test (AVLT) in a sample of 474 healthy individuals ranging from those in their twenties to those in their eighties; a steady decrease in performance was observed with increasing age in those over 30 years. Free recall performance often appears to be more detrimentally affected by the aging process than recognition, although other memory processes do appear to be sensitive to age effects, for example, priming skill learning and classical conditioning (see Binks, 1989 and Schonfield, 1980 for reviews).

Shimamura (1994) proposed that age-related effects on memory may be associated with regional brain changes, that is, that problems of attention, working memory, word finding, and source memory may be related to aging effects in frontal lobe functioning, whereas problems with new learning may be related to medial temporal lobe dysfunction. Superimposed on these changes, Shimamura (1994) argued that general changes in neuronal efficiency may pervade all neural systems as a consequence of age (e.g., reduced blood flow, inability to prevent damage from oxidation, and poor metabolic uptake). It is proposed that these metabolic changes account for the pervasive finding of cognitive slowing in older individuals. It can be imagined easily how these specific age-related memory effects may act synergistically with nonspecific slowing to result in impaired performance. Evidence has shown that differing brain regions undergo differing degrees of neuronal loss with age. Many areas, such as striate-cortex and parietal cortex, do not lose an appreciable number of neurons during life. However, 15–20% of the neurons in the neo-striatum and frontal cortex are lost between young adulthood and old age (Squire, 1987). Similarly, approximately 5% of hippocampal neurons are lost per decade, so that at the age of 80 years; approximately 20–30% of hippocampal pyramidal cells will have been lost (Squire (1987). It is tempting, therefore, to link the neuronal cell loss in frontal and hippocampal regions to frontal and temporal memory systems which Shimamura (1994) has implicated in the aging process.

7.07.9.2 Age-associated Memory Impairment

Crook et al. (1986) proposed the concept of age-associated memory impairment (AAMI). The following criteria were developed:

(i) at least 50 years of age;

(ii) gradual onset of complaints of memory loss in everyday activities;

(iii) memory test performance one standard deviation below the mean of young adults on standardized tests of recent memory with adequate normative data;

(iv) adequate intellectual function (e.g., Scaled score of at least nine on WAIS-R vocabulary);

(v) absence of dementia (i.e., Mini-Mental Status Examination score of 24 or above); and

(vi) exclusion criteria, delirium, neurological disorder, cerebrovascular pathology, head injury, etc.

However, it is clear that a vast number of elderly people would meet these criteria, particularly the reference to scoring one standard deviation below memory test scores for young adults; indeed by definition over 16% of young adults would meet this criterion. This led to the proposed diagnostic category being severely criticized (e.g., O'Brien & Levy, 1992). Partly as a consequence of this type of criticism, AAMI was not included in the *Diagnostic and Statistical Manual of Mental Disorders* (*4th ed., DSM-IV*) of the American Psychiatric Association. Some skeptics felt that the condition may have been invented in order to legitimize pharmaceutical treatment of this new disorder. However, it is clear that aging is associated with significant impairment of memory functioning and more needs to be known about this naturally occurring phenomenon rather than labeling it as a categorical disorder. Caine (1992) has proposed that instead of the term AAMI, age-related cognitive decline (ARCD) should be discussed and researched and this has been included in *DSM-IV* (code 780.9).

7.07.9.3 Does Memory Functioning in Dementia Represent Accelerated Aging?

In many studies, the finding that memory after a delay is poorer in the elderly than in young subjects has usually been attributed to an acquisition deficit rather than to faster rate of forgetting. For many years, a controversy has raged over whether dementia (particularly of the Alzheimer type) represents accelerated aging effects or whether it represents a quite separate clinical entity (Huppert, 1994). In an important study, Huppert and Kopelman (1989) found that normal aging produces a mild acquisition deficit as well as a significant increase in forgetting rate, and the rapid rate of forgetting in the elderly cannot be attributed simply to differences in the initial level of acquisition. In a comparison of normal aging with dementia, normal aging produced a mild acquisition deficit and a slight increase in the rate of forgetting, whereas dementia superimposes a severe acquisition defect but no further effect on forgetting rate (Huppert & Kopelman, 1989). In their study, Greene, Baddely, and Hodges (1996) employed the Doors and People test (Baddeley, Emslie, & Nimmo-Smith, 1994; see below) in a comparison of early DAT patients vs. elderly controls. They concluded that the episodic memory deficit in DAT is general in nature and primarily reflects impaired learning rather than accelerated forgetting or disrupted retrieval. Thus, memory functioning in dementia appears to be qualitatively different from that observed in "normal" aging.

7.07.9.4 Commonly Used Memory Tests: Implications for use with Elderly Subjects

In the following sections several commonly used clinical memory measures will be outlined briefly and evaluated, with a particular emphasis on the appropriateness of their use with the elderly. For further details, the interested reader is referred to Lezak (1995) and Spreen and Strauss (1991). Unfortunately, pressure of space precludes coverage of all but a few of the most widely used instruments.

7.07.9.4.1 The Auditory Verbal Learning Test

The auditory verbal learning test (AVLT; Rey, 1964) is one of the most widely used word learning tests in clinical research and practice. Five presentations of a 15-word list are given, each followed by attempted recall. This is followed by a second 15-word list (list B), followed by recall of list A, and delayed recall and recognition are also tested. A key feature of the AVLT (and its successor, the California Verbal Learning Test) is that it affords the opportunity to measure rate of learning, as opposed to recall of a single stimulus, or series of stimuli. An equivalent form of AVLT has been provided by Crawford, Stewart, and Moore (1989c). Because of the rather abstract nature of the test, that is, 15 unconnected words and repetition over five trials, this can be a rather stressful measure for an elderly individual of failing cognitive abilities. The initial normative data that was referred to widely was rather limited (Lezak, 1983). However, the situation has improved, with excellent norms for the elderly provided by Geffen, Moar, O'Hanlon, Clark, and Geffen (1990), Ivnik et al. (1992), and Tuokko and Woodward (1996). The normative data on this test has also been reviewed by D'Elia, Boone, and Mitrushimo (1995). For a detailed review of the AVLT, see Peaker and Stewart (1989).

7.07.9.4.2 The California Verbal Learning Test

In an attempt to expand upon the assessment of learning and retrieval strategies, the California Verbal Learning Test (CVLT) evolved from the AVLT and was developed to provide an instrument "reflecting the multifactorial ways in which examinees learn, or fail to learn, verbal material. The CVLT's assessment of learning strategies, processes, and errors is designed for use in both clinical and research practice" (Delis, Kramer, Kaplan, & Ober, 1987). A variety of memory measures are obtainable from this measure including short- and long-term free recall and recognition, serial learning curves, learning strategy (i.e., semantic vs. serial clustering), etc. The CVLT also has the added advantage of presenting the stimuli in an everyday, relatively nonthreatening manner, that is, learning two shopping lists. The reliability data for the principal measures appear to be adequate although some of the scores which are derived from them have poor reliability. The normative database consists of 273 neurologically intact individuals, 104 males, 169 females, with a mean age of 58.9 (15.4) years. This mean age for controls is considerably higher than that normally employed for psychometric test development. However, the control subjects were recruited from several research projects and the authors state "in order to make the best use of the data while giving due regard to the varying number of cases at each age and sex, smoothed age curves were fitted to the raw data using multiple regression" (Delis et al., 1987, p. 35). This curve-fitting procedure for normative data has been criticized (see Elwood, 1991).

7.07.9.4.3 The Wechsler Memory Scale-Revised

The WMS-R (Wechsler, 1987) is generally considered to be a considerable improvement on its predecessor, the WMS. The WMS-R includes new subtests such as figural memory, visual paired associates, and visual memory span. As Mayes and Warburg (1990) comment, "the revised WMS therefore appears (potentially) to be considerably more satisfactory in a number of respects than its older brother—particularly in respect of normative data, information of standardization, validity and reliability, scoring criteria (especially logical memory), range of memory functions sampled and provision for clinical interpretation of test rests. This has been achieved at the cost of increased administration time, particularly if the Delayed Recall is tested, abandonment of provision of an equivalent alternate form, and increased cost of the test." For elderly subjects, however, the Delayed Recall Test requires an attempt to reach a criterion on both verbal and visual paired associates. This can lead to repeated experience of failure over six trials which can provoke distress in elderly individuals. In addition, the scoring of the tapping forwards and backwards subtest leads to considerable anxiety on the part of many psychologists as they find it an extremely difficult subtest to score. However, preliminary data on the inter-rater reliability of this subtest appears to be satisfactory (O'Carroll & Badenoch, 1994).

The normative sample consists of 316 subjects in six age groups, selected to be representative of the US population in terms of race, geographic region, education, and IQ. Elwood (1991) points out that the normative sample of 316 "seems hardly adequate for the standardization sample, given the broad age range of its target population. In relative terms, comparison between the WMS-R sample and the 2000 subjects used for the WAIS-R is striking" (p. 185). It is important to note that the WMS-R standardization sample systematically excluded subjects in the age range of 45–54 years using interpolated values in their place.

Given that the WMS-R is one of the most widely used memory tests, it would be expected that the normative sample would be representative of the general population. However, Huppert (1994) has pointed out that examiners were instructed to screen potential examinees for a list of several medical factors which may affect cognitive performance and has suggested that, not only were these criteria poorly specified, they almost certainly excluded a high percentage of normal nondemented older individuals. As she states, "this has serious consequences for the use of these norms in assessing whether or not an individual's memory performance is below that expected for their age. The effect will be to over-estimate the number of elderly people with clinically significant memory impairment" (p. 315). A further brief discussion of this issue can be found in Section 7.07.2.

7.07.9.4.4 The Rivermead Behavioural Memory Test

The Rivermead Behavioural Memory Test (RBMT) was designed specifically to try to detect impairment of everyday memory function by providing test items that resembled activities in everyday life, for example, remembering to deliver a message, remembering to retrieve a personal belonging after an interval, etc. The RBMT has the advantage of having four matched parallel versions, thus allowing for repeated assessment to determine the effects of disease progression and/or clinical intervention. An additional important advantage of the RBMT is the careful work the authors have undertaken in order to ensure the measure's validity. They assessed validity in three ways: (i) by demonstrating a high correlation between the RBMT and other standard memory tests; (ii) by demonstrating a high correlation between RBMT scores and subjective ratings of memory impairment; and (iii) most importantly, by demonstrating a high correlation between RBMT score and observer ratings of memory lapses (Wilson, Cockburn, Baddeley, & Hiorns 1989). Such a rigorous approach to demonstrating the ecological validity of a memory measure is both unusual and welcome.

In general, the RBMT is well tolerated by elderly individuals on account of its "everyday feel" and the fact that it consists of a number of brief, relatively nonthreatening subtests. The initial normative data were provided from 118 subjects aged 16–69 years. However, a supplement to the manual provides details of elderly healthy subject's performance and is, therefore, a welcome addition to the clinical neuropsychologist's armamentarium for the assessment of memory in the elderly adult. Cockburn and Smith (1989) recruited 119 people aged between 70 and 94 years, with a mean age of 80.5 years. In order to provide as random a sample as possible, every fourth name of people born in or before 1917 was taken separately from anyone from the age/sex register of a five-doctor general practice that covered both urban and rural areas. Such a detailed attempt to recruit a truly "normative" elderly sample is welcome. Furthermore, the authors determined that age and general intellectual ability were significant correlates of Rivermead performance. Accordingly, they developed useful tables for determining abnormal memory scores in relation to age and estimated premorbid intellectual level. This additional work renders the RBMT one of the best memory measures for use in an elderly sample. Furthermore, the nonthreatening nature of the stimulus materials makes it more user-friendly for an elderly sample than many of its competitors. A limitation of the RBMT is that it may be rather insensitive to mild impairments of memory. Following on from this, ceiling effects may be encountered in mildly impaired patients, making any positive effects of intervention difficult to demonstrate.

7.07.9.4.5 The Rey Complex Figure Test

This test has been used widely (Rey, 1964), particularly because of the dearth of adequate visuospatial memory tests. A complicated figure is presented and the subject is requested to copy it. The original and copy are then removed and the subject is asked to draw the figure again from memory, after varying delay intervals. As Mayes and Warburg (1990) point out there have been many variants on administration and scoring criteria; and the available norms, particularly for elderly samples, have been rather limited. Spreen and Strauss (1991), however, provide normative data from the age ranges 60–85 years. While this is a welcome addition, the numbers in each respective age band are again limited. For example, in the 70+ age range, the sample size is $n = 10$. This obviously makes clinical interpretation somewhat difficult.

7.07.9.4.6 The Recognition Memory Test

The recognition memory test (RMT) is a widely used instrument, particularly in amnesia research (Warrington, 1984). It consists of two subtests, recognition memory for 50 words and recognition memories for 50 faces. Following presentation of the words and faces, the testee is required to perform an immediate two-choice recognition task and is generally considered to be less stressful and anxiety provoking than many other memory measures. The normative data consists of 310 control in-patients without cerebral disease, aged 18–70 years. As Nelson (1990) points out in a review of the RMT, "This test can only be used with subjects up to the age of 70 because of the unknown effects of normal aging on the test scores past this age."

7.07.9.4.7 The Doors and People Test

The Doors and People Test was devised in order to provide comparable measures of visual and verbal memory that test both recall and recognition, that does not provide floor or ceiling effects and includes both learning and forgetting measures (Baddeley et al., 1994). The test comprises four sections: a doors test, where single doors have to be recognized against three competing distracters; a shapes test, where diagrams have to be reproduced from memory; a names test, which is a verbal recognition test; and finally, the peoples test, where the subject is required to learn an association between a series of photographs and names. Normative data from 238 subjects is provided, in five age bands (16–31, 32–47, 48–63, 64–79, 80–97). In fact, one of the stated aims was that doors and people should "provide an unstressful test that is acceptable to a wide range of subjects, extending from patients suffering from dementia or dense amnesia to healthy young normal subjects" (Baddeley et al., p. 4). However, some elderly, cognitively impaired subjects may find this test quite demanding.

7.07.10 ASSESSING LANGUAGE DYSFUNCTION

In normal aging, deterioration of language has been considered to be either very marginal or a consequence of deterioration of other cognitive functions. Although the majority of elderly people do not develop any gross language disturbance, close scrutiny suggests that some subtle changes are indeed present (Kemper, 1992). In contrast, language deficits are a very pronounced aspect of cognitive deterioration in DAT. Aspects of language may also be disrupted by other forms of cortical and subcortical dementias (e.g., fronto-temporal dementias, Huntington disease, Parkinson's disease, vascular dementia, etc.).

Given the frequency of occurrence of language disorders after degenerative neurological diseases associated with old age, language should be investigated in depth. In the following sections, the most frequent language disturbances encountered either in normal aging or as a consequence of degenerative disorders will be reviewed briefly and suggestions made about appropriate assessment procedures. The examination of language competence should include assessment of verbal expression (naming, vocabulary, discourse and verbal fluency, repetition, prosody), verbal academic skills (reading, writing, spelling), and verbal comprehension (syntax, idioms, and proverb interpretation). As yet there is no comprehensive battery specifically designed for assessing language abilities in the elderly, therefore clinicians often fall back on general aphasia batteries. This can be problematic as the norms for the elderly are limited. In addition there is a need to derive individual comparison standards when assessing language function because of the large degree of variability in premorbid competence (see Section 7.07.4.1). Therefore, examiners should devote time in their initial interview to inquire about factors such as educational level, previous occupation, and reading habits.

7.07.10.1 Naming

Everybody has personal experience of looking for the name of an object, having a very strong feeling of exactly knowing what one wants to say but being unable to recall the correct word. This phenomenon is often referred to as "tip of the tongue." Word finding difficulty is the most common language symptom experienced by healthy people over 60 (Burke, Mackay, Worthley, & Wade, 1991). It is experienced occasionally by young people, but aging seems to exacerbate the problem. The standard task for assessing word finding abilities is confrontation naming, in which subjects are required to name a given object, or more commonly, to name objects depicted in line drawings. One of the most popular tests is the Boston Naming Test (BNT) (Kaplan, Goodglass, & Weintraub, 1983). The original version of this test included 85 items (Kaplan, Goodglass, & Weintraub, 1978); however, the shortened version produced in 1983 is used more frequently. This test consists of 60 line drawings of objects of graded difficulty, ranging from very common objects (e.g., a tree) to less familiar objects such as an abacus. Although the original normative data are scanty, several studies have provided supplementary norms; these studies have been reviewed comprehensively by D'Elia et al. (1995). Data collected from a large sample of 750 older adults (over the age of 55 years) showed that performance on the BNT was influenced by age which accounted for more than 10% of the raw score variance (Ivnik, Malec, Smith, Tangalos, & Petersen, 1996).

A study carried out with patients showed that correlation between the two (long and short) forms of the test and two 42-item forms (obtained by dividing the original test into two nonoverlapping forms) ranged from 0.96 to 0.92 (Huff, Collins, Corkin, & Rosen, 1986). A study (Mitrushina & Satz, 1995) carried out with healthy elderly individuals has reported that repeated testing with the BNT reveals a high stability of scores over time, suggesting a lack of practice effects.

Naming may also be assessed using the Object Naming Test (Newcombe, Oldfield, Ratcliff, & Wingfield, 1971) or the Graded Naming Test (Warrington, 1984). A detailed examination of naming should also include naming of actions (verbs), body parts, and colors. Naming of actions may be tested by means of formal action naming tests (McCarthy & Warrington; 1986; Obler & Albert, 1978). As an alternative, the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1987) provides a comprehensive examination of naming abilities. However, this test is time-consuming and is best suited to highly experienced examiners; furthermore, there is limited normative data for the healthy elderly.

Confrontation naming studies based on the BNT indicate that there is an increase in the variability of scores among apparently healthy individuals in their 60s, thereby suggesting impairment in a subset of this age group (Van Gorp, Satz, Kiersch, & Henry, 1986). There is consistent evidence of a steep decline in the late 70s (Goodglass, 1980; Mitrushina & Satz, 1995). This decline is evident with both low-frequency nouns and verbs but is more marked in the former (Obler, Albert, & Goodglass, 1981). Poor scores on confrontation naming can arise because of a failure to give any response or because of incorrect responses. Increases in the number of errors on confrontation naming tasks were observed in a couple of studies that compared the performance of over 80-year-old subjects with younger subjects (60- and 70-year-olds) (Albert, Heller, & Milberg, 1988; Borod, Goodglass, & Kaplan, 1980). Obler and Albert (1985) analyzed the qualitative features of errors made by the elderly and noted that they tend to be characterized by the use of circumlocutions.

Naming can also be studied using verbal fluency tasks which require the generation of words from semantic or phonemic cues. These tasks are reviewed in Section 7.07.13.2. Two hypotheses have been put forward to explain the naming problems that occur with normal aging. One argues for a degradation in the structure of lexical knowledge, whilst the other argues for an impairment in retrieval mechanisms (see Light, 1992 for a review). Experimental evidence shows that semantic priming effects are virtually identical for young and healthy older adults (Balota & Duchek, 1988; Burke, White, & Diaz, 1987; Monti et al., 1996); further, in naming tasks, performance both of young and old subjects is enhanced more by phonological cues than by semantic cues (Nicholas, Obler, Albert, & Goodglass, 1985). These findings suggest that structural semantic knowledge of words is intact and that a deficit of retrieval or access to this knowledge may be responsible for word-finding difficulties.

The situation is more complicated when trying to understand the severe word-finding difficulties that occur in DAT. Conflicting explanations have been provided for this, and evidence in favor of an access or retrieval deficit (Kempler, 1988; Neils, Brennan, Cole, Boler, & Gerdeman, 1988), or evidence of a structural disruption (Hodges, Patterson, Graham, & Dawson, 1996; Kempler, Andersen, & Henderson, 1995; Nebes, Brady, & Huff, 1989) has been reported. Further, in demented people, failure on confrontation naming could be due to visuoperceptual deficits that result in failure to identify and recognize an object, as suggested either by the fact that performance improves when additional sensory cues are provided (Barker & Lawson, 1968) or by the fact that errors consist of naming objects which are visually rather than semantically related to the target.

7.07.10.2 Vocabulary

As discussed earlier, structural knowledge about words does not change greatly in the normal elderly population, and vocabulary scores do not typically decrease in the normal aged (Salthouse, 1988), rather they may increase, albeit modestly. This evidence supports the notion of a preserved structural knowledge of words in the normal aged population. The most widely used instrument for the assessment of vocabulary ability is the Vocabulary subtest of the WAIS-R (Wechsler, 1981) which consists of 35 items organized in order of increasing difficulty. Performance on this test is greatly influenced by education (Malec, Ivnik, Smith, & Tangalos, 1992) and socio-cultural factors (Huttenlocher, Haight, & Bryk, 1991) rather than age. A test–retest reliability coefficient of 0.71 has been reported in a study of the healthy elderly (Snow, Tierney, Zorzitto, Fisher, & Reid, 1989).

7.07.10.3 Discourse

The assembling of a meaningful and well-structured discourse requires lexical and syntactic competence, including a knowledge of pragmatics (i.e., discourse structure and rules), adequate sequencing skills, including the ability to integrate concepts within and between sentences, and adequate working memory capacity. Material from the clinical interview can be used to assess the quality of discourse and this can be supplemented by asking a patient to describe a standard picture or relate a well-known story (e.g., Little Red Riding Hood). As a standard picture, the Cookie Theft Picture of the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983) may be used. The examiner should assess whether the production is assembled coherently, if elements are connected and sequenced in the appropriate order, whether an individual makes correct use of reference pronouns, and whether appropriate elements are produced. Performance should also be evaluated in terms of the complexity of sentences, grammar and syntax, and richness of vocabulary.

Aging does not have macroscopic effects on discourse; some subtle deficits, however, may be present in the use of pronouns in discourse, with older subjects using pronouns without specifying their referent in spontaneous speech, a situation that especially involves memory (Pratt, Boyes, Robins, & Manchester, 1989). It has also been reported that deficits on discourse production reflect changes in those macrolinguistic skills that require integration of linguistic and nonlinguistic cognitive processes rather than language-specific cognitive processes (Glosser & Deser, 1992). It was observed that elderly subjects did not differ from young subjects on microlinguistic aspects such as syntactic complexity, lexical production, and use of lexical cohesive ties; however, older individuals failed to refer coherently to the general topic. Discourse production is impaired significantly in patients with DAT, even in the very early stages, and is characterized by incoherence and emptiness of speech; spontaneous speech therefore remains fluent but carries less information (Giles, Patterson, & Hodges, 1996; Ripich & Terrell, 1988).

7.07.10.4 Verbal Comprehension

Comprehension is usually well preserved in healthy elderly individuals, and apparent deficits are often simply a reflection of poor auditory acuity (see Section 7.07.2). The most common test of verbal comprehension is the Token Test (De Renzi & Vignolo, 1962). The original version of this test consists of 62 items but a 36-item short-form (De Renzi & Faglioni, 1978) is more commonly used in clinical practice. The materials consist of tokens which differ in color, shape (squares and circles), and size (large and small). The examinee has to follow verbal instructions which increase in complexity from simple commands (e.g., "Touch a circle," "Touch the red circle") to commands such as "Before touching the yellow circle, pick up the red square." The test authors recommend that scores be adjusted for years of education. Adjusted scores of 25–28 are regarded as indicating mild comprehension problems and 17–27 moderate problems; scores below this are classified as severe or very severe. This test is extremely sensitive to impairment of linguistic processes. De Renzi and Faglioni (1978) reported a true positive classification rate of 93% when assessed against results from an extensive aphasia examination; the true negative rate was 95%.

Test–retest reliability with aphasic patients is very high, ranging from 0.92 to 0.96 (Gallagher, 1979; Orgass, 1976). There are inconsistent reports about the effects of age and education on performance, some suggesting a moderate increase in number of errors with aging (see Lezak, 1995).

Memory and attention play an important role in performance on comprehension tests. For example, working memory load is high on the later items of the Token Test. Therefore, results obtained from measures of attention and memory must be integrated with the results from tests of comprehension when reaching a formulation regarding the nature of an individual's difficulties.

7.07.11 ASSESSING VISUOPERCEPTUAL DYSFUNCTION

Visuoperceptual deficits represent a very common finding in sufferers of degenerative cortical illnesses. A comprehensive evaluation of visuoperceptual abilities should include assessment of color perception, visual recognition, visual organization, and visual interference. A first step in assessing visuoperceptual functions should be to test for the presence of unilateral neglect; this topic is covered in Section 7.07.12.1.

7.07.11.1 Visual Recognition

Evaluation of visual recognition abilities includes assessment of several aspects. It includes, first of all, assessment of perception of angulation. The most widely used test in this case is the Judgment of Line Orientation Test (Benton, Hannay, & Varney, 1975). This test includes 30 pairs of angled lines to be matched with the corresponding lines in a radially arranged 11-line display. An advantage of this test over many clinical tests used to assess visuospatial functioning is that it imposes minimal demands on praxic, mnestic, or executive functions. The norms include correction factors for age and education but are based on only 144 subjects. Age effects on this test have been confirmed by Eslinger and Benton (1983); test–retest reliability over a one-year delay in an elderly sample was modest ($r = 0.59$) (Levin et al., 1991).

The recognition of objects depicted under conventional and unusual views should be examined. Warrington and Taylor's (1973) Unusual Views Test can be used for the latter. Assessment of visual recognition also includes testing of face recognition abilities. For this purpose the Benton Facial Recognition Test (BFRT) (Benton, Hamser, Varney, & Spreen, 1983; Benton & Van Allen, 1968) may be administered. The original version of this test includes 22 items; it is a matching task rather than a true recognition task since it requires matching a front-view photograph of a target face to a matrix of faces photographed either in front-view, three-quarter-view, or under different lighting conditions. Norms are available up to the age of 74 years. Some degree of age-related decline in performance on this test has been observed (Benton, Eslinger, & Damasio, 1981) and the norms have correction factors for age and education. Further evidence of age-associated impairment at several stages of the face recognition process has been reported (Maylor, 1990). It should be noted that prosopagnosics can perform within normal limits on the BFRT. However, they take an inordinate amount of time per trial as their performance is characterized by a piecemeal strategy in which each feature of the target face is compared with each feature of the faces in the matrix (Ellis, 1992).

7.07.11.2 Visual Organization

Making sense out of fragmented, incomplete, or ambiguous figures requires intact visual organization abilities. Those abilities may be tested using tasks such as the Street Completion Test (Street, 1931) or the Hooper Visual Organization Test (HVOT) (Hooper, 1983). The HVOT requires examinees to name 30 fragmented objects. Test–retest reliability in an elderly sample was 0.68 (Levin et al., 1991) and the original norms cover up to age 68 years. Studies which have investigated this task in elderly subjects (55–92 years of age) indicate that performance drops off significantly with age (Farver & Farver, 1982; Montgomery & Costa, 1973; Villardita, Cultrera, Cupone, & Mejia, 1985).

7.07.11.3 Visuoconstructional Ability

Visuoconstructional abilities are usually assessed by means of copying or free drawing. The Rey-Osterreith Complex Figure (Osterreith, 1944) is the most widely used copying test (for further details see Section 7.07.9.4.5). The Block Design subtest of the WAIS-R may be used for assessing two-dimensional constructional abilities.

7.07.11.4 Visual Interference

Visual interference may be evaluated by means of the Hidden Figures Test (Gottschaldt, 1928) which requires a subject to recognize a simpler figure, hidden in a more complex one, by marking its outline. Normative data up to the age of 79 years are provided in Spreen and Strauss (1991). Another useful and well-known task for assessing visual interference is the Overlapping Figure Test (Poppelreuter, 1917). An effect of normal aging on performance on this test has been reported in a recent study (Della Sala, Laiacona, Trivelli, & Spinnler, 1995) in which severely impaired performance was observed in a sample of patients with DAT.

7.07.12 ASSESSMENT OF ATTENTION

7.07.12.1 Unilateral Neglect

Unilateral cerebral lesions (especially right parietal lesions) often results in unilateral neglect. Neglect labels a phenomenon that involves lack of awareness of stimuli appearing on the side contralateral to the cerebral lesion. It is accepted widely that neglect is not a unitary phenomenon and evidence of dissociations across modalities within the same patient and between patients have been reported (Halligan & Robertson, 1992).

Assessment of neglect should explore representational and motor abilities as well as visuospatial abilities. Personal (relating to the subjects body), peripersonal (within arms length), and extrapersonal space should be investigated. A dissociation between personal peripersonal space and extrapersonal space has been illustrated by Halligan and Marshall (1991) who described a case with severe neglect in the former spheres but who nevertheless failed to exhibit neglect for distant experimental stimuli and could aim successfully at areas on a dart board which were to the left of the midline.

The most frequently administered test for assessing visuospatial exploration is the Albert Test (Albert, 1973) which consists of 40 lines drawn out at various angles on a sheet of paper; the subject's task is to cross out all of them. Normal subjects usually perform flawlessly on this test. Neglect may also be assessed using the Bells Test (Gauthier, Dehaut, & Joanette, 1989) that includes 315 silhouettes of common objects, of which 35 are bells; the subject's task is to search for the bells. Normative data (including data from subjects aged 50–89 years) are available for this test, and no more than three bells are usually missed by normal elderly subjects. Test–retest reliability over a two-week period was reported as 0.69.

Line Bisection is also a widely used task for assessing neglect; it requires a subject to divide horizontal lines at the center point. Performance on this task can be highly variable, and it is advisable to acquire data from multiple trials. Patients with right hemisphere lesions show a right-deviation error of the bisection point, that increases as a function of the length of the line (Halligan & Marshall, 1989).

Representational neglect, that involves mental imagery, may be assessed by asking a patient to describe a well-known scene from a particular perspective. A patient may fail to recall objects on the left but recall the same objects when asked to imagine the scene from the opposite perspective (Bisiach & Luzzatti, 1978).

A comprehensive battery for testing visual neglect has been assembled by Wilson, Cockburn, and Halligan (1987). This battery is called the Behavioural Inattention Test (BIT) and includes conventional tests (star cancellation, letter cancellation, figure copying, line crossing, line bisection, and representational drawing) and behavioral subtests (picture scanning, telephone dialing, menu reading, article reading, telling and setting the time, coin sorting, address and sentence copying, map navigation, and card sorting). The test manual reports impressively high coefficients for inter-rater reliability (0.99), test–retest reliability (0.99) and parallel form reliability (0.91). Among the six conventional tests included in the battery, star cancellation had substantially greater sensitivity than the other tests (Halligan & Marshall, 1988) (see Figure 1).

Although many studies have explored the performance of normal subjects on some of the tests commonly used to assess neglect, relatively little is known about the effect of normal aging and dementia on these tasks. Neglect is seldom found in DAT, despite the frequency of visuoperceptual disorders in this condition; Huff et al. (1987) reported an incidence of 3% based on 95 cases. A high incidence of neglect has been reported by one study of visuospatial ability in DAT (Freedman & Dexter, 1991). These authors concluded that neglect in DAT may not be an unusual phenomenon but could often be missed because of a failure to conduct a detailed assessment. However, it should be noted that neglect was considered to be present if the discrepancies between errors on left- vs. right-sided stimuli on visuospatial tasks exceeded control values (these control values were simple difference scores). As there typically will be a high number of errors on visuospatial tasks in DAT cases, and often they will exhibit considerable intrasubject variability, these results may be an artifact of Freedman and Dexter's (1991) scoring methods. This alternative explanation is strengthened by the fact that the number of cases classified as exhibiting right-sided neglect was exactly equivalent to the number classified as exhibiting left-sided neglect (right-sided neglect is relatively rare). When looking at individual reports of cases of DAT displaying neglect, only three cases have been reported (Ishiai, Okiyama, Koyama, & Seki, 1996; Pentore & Venneri, (1994). Pentore and Venneri (1994) in their study reported two patients with DAT of moderate severity who manifested neglect in the later stages; one of these patients manifested motor neglect in addition to visuospatial neglect. Such limited evidence may lead to the conclusion that neglect in DAT is actually rather rare. This would be in accord with the fact that neglect is generally observed after unilateral lesions, while DAT, although it may present some degree of asymmetry at onset, involves bilateral degeneration (see Kirk & Kertesz, 1991, for further discussion).

7.07.12.2 The Anterior Attention System

The anterior attention system is considered to underlie performance on tasks of sustained, divided, and selective attention (Posner & Petersen, 1990). Such attentional problems are common and debilitating features of many neurological and psychiatric disorders associated with aging (Crawford, McKinlay, & Parker, 1992b; Lezak, 1995; Walsh, 1991). However, the assessment of attentional dysfunction is a problematic area in clinical neuropsychology as many potentially useful tasks used in experimental studies have not been translated into practical clinical instruments. Many existing tests do not reflect theoretical knowledge on attention and its fractionation (Posner & Petersen, 1990) and, in many cases, are poorly standardized (Lezak, 1995). An additional consideration is that tests such as the PASAT, which have proved valuable in work with younger clinical populations because of their demonstrated sensitivity and predictive validity, are too demanding to be suitable for use with the majority of elderly clients. Because of space limitations, we will limit coverage to a single battery, the Test of Everyday Attention

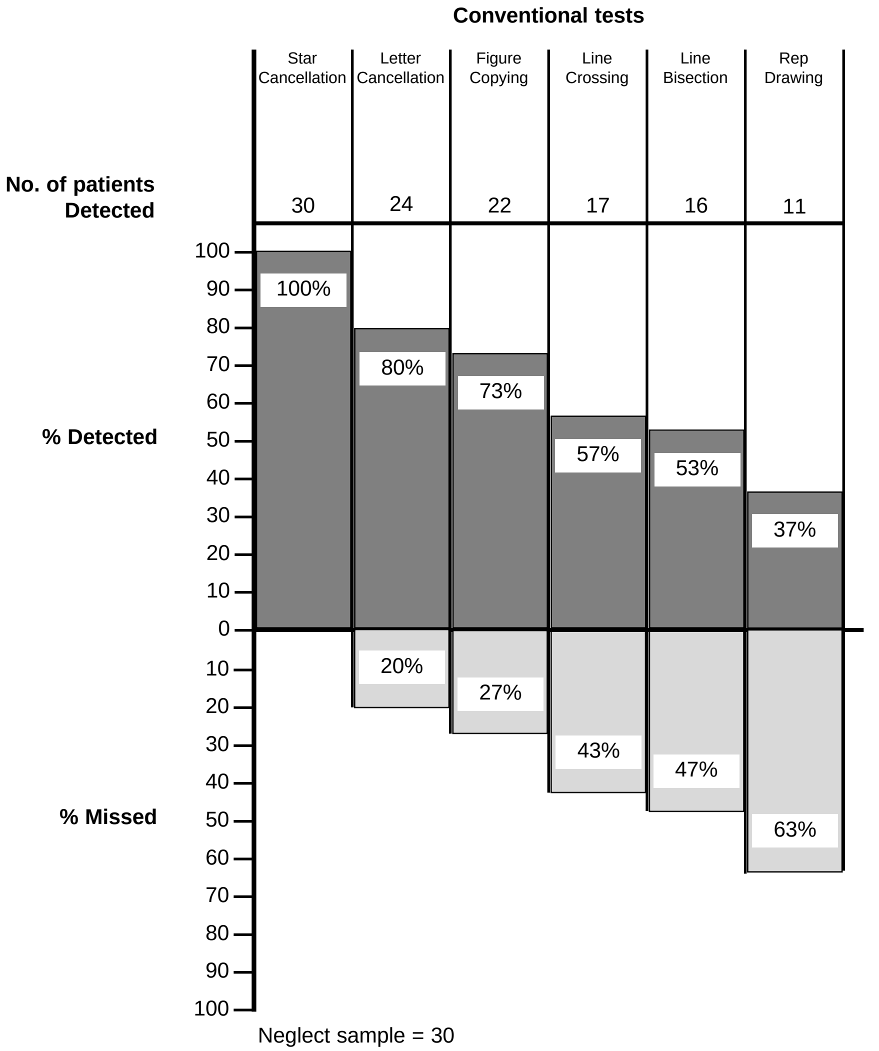

Figure 1 The relative sensitivities of the six conventional screening tests of the Behavioral Inattention Test (reproduced from Halligan & Robertson, 1992).

(TEA) (Robertson, Ward, Ridgeway, & Nimmo-Smith, 1994; Robertson, Ward, Ridgeway, & Nimmo-Smith, 1996).

TEA is a welcome development as it attempts to overcome some of the aforementioned problems in the assessment of attentional dysfunction. It consists of eight subtests, standardized to have means of 10 and SDs of 3, which measure sustained, selective, and divided attention. For example, in the Map Search subtest, which was designed as a measure of visual selective attention, the examinee has to search for symbols (e.g., a knife and fork representing eating facilities) on a tourist map of a city. The Elevator Counting subtest, which was designed to measure sustained attention, is a tone counting task in which the examinee is asked to imagine they are in an elevator in which the floor indicator has failed (thus counting the tones is the only way to establish which floor the elevator is on). A related subtest, designed to measure auditory selective attention, requires the examinee to ignore distracter tones which are of a higher pitch.

The TEA was normed on 154 healthy participants aged 18–80 years. In addition to reflecting current thinking on the fractionation of attention, it is designed to be ecologically valid, that is, many of the subtests are designed to mimic everyday activities. Another advantage is that it has three parallel versions; this avoids the interpretative problems encountered when attempting to measure change using the same test materials.

Crawford, Somerville, and Robertson (1997) have prepared tables which allow clinicians to examine a client's relative strengths and weaknesses on the TEA (in terms of the reliability and abnormality of subtests deviations from the subtest mean), thereby supplementing the conventional normative comparison standards presented in the test manual with individual comparison standards.

7.07.13 ASSESSMENT OF EXECUTIVE FUNCTIONS

Executive deficits arising from damage to the prefrontal cortex and related structures can have a much more profound effect on a client's prospects for successful adjustment and independent living than the more circumscribed deficits arising from posterior cortical lesions. However, these problems have proven difficult to quantify and as such can be regarded as constituting the most problematic area in neuropsychological assessment. For example, reviews of tests such as the Wisconsin Card Sorting Test (Grant & Berg, 1948; Heaton, 1981), which are aimed specifically at capturing the core cognitive problems (e.g., planning deficits, underutilization of feedback, etc.), indicate that existing indices are of limited utility when used with individuals (Mountain & Snow, 1993; Reitan & Wolfson, 1994). Thus, considerable reliance must be placed on careful scrutiny of a client's approach to a range of tasks to identify these cognitive difficulties.

7.07.13.1 Frontal Hypotheses

Executive deficits are thought to underlie many of the behavioral disturbances seen in neurological disorders associated with increasing age. One of the more obvious examples is Parkinson's disease (PD). Executive deficits would be expected in PD on pathophysiological grounds as the basal ganglia, which are deprived of dopaminergic innervation because of degeneration in the substantia nigra, have strong reciprocal connections with the frontal cortex. In addition, the ventral tegmental area, which provides a direct dopaminergic innervation of the prefrontal cortex, also deteriorates in PD (see Crawford, Besson & Ebmeier, 1990 for brief review). There is neuropsychological evidence which suggests that executive deficits are an early feature in PD and that these deficits qualify as differential deficits in executive functioning (Lees & Smith, 1983). Indeed, it has been argued that fronto-executive dysfunction provides a parsimonious explanation for all of the cognitive deficits observed in nondemented Parkinson's cases (Della Sala, 1988). This is an extreme position but it is interesting to note that Bondi, Kaszniak, Bayles, and Vance (1993) have demonstrated that deficits in other cognitive domains were no longer significant after partialing out performance on tests of executive function.

It has also been proposed that normal aging is associated with a selective decline in fronto-executive function (Dempster, 1992). Daigneault, Braun, and Whitaker (1992), for example, note that "The currently dominant neuropsychological model of normal brain aging postulates that cognitive functions dependent on the integrity of the prefrontal brain regions are among the first to deteriorate" (p. 99). This frontal hypothesis stems from indications that age-related changes in the neuroanatomy and neurochemistry of the brain are more pronounced in the frontal lobes than in other cortical areas (Fuster, 1989) and from demonstrations of age-related deficits on neuropsychological tasks sensitive to frontal dysfunction (Daigneault & Braun; 1993; Daigneault et al., 1992; Mittenberg, Seidenberg, O'Leary, & DiGiulio, 1989). However, there is only limited evidence that such deficits qualify as differential deficits (i.e., that they significantly exceed the deficits found in other cognitive domains).

7.07.13.2 Verbal Fluency

Verbal fluency tests, and in particular initial letter fluency, are widely used as tests of executive dysfunction (Parker & Crawford; 1992; Walsh, 1991). Initial letter fluency, also referred to as the Controlled Oral Word Association Test (COWAT) (Benton; 1969) or phonemic fluency, requires the generation of words from initial letters (normally F, A, and S or C, F, and L) under time constraints (normally 60 seconds per letter, although 90 seconds is also used). Fluency tests are reliable, quick to administer, and even patients with quite severe deficits can understand the task requirements. In an area replete with failures to replicate, studies of initial letter fluency have been remarkably consistent in demonstrating impaired fluency following left or bilateral damage (Benton, 1968; Bornstein, 1986; Miceli, Caltagirone, Gainotti, Masullo, & Silveri, 1981; Miller, 1984; Milner, 1964). In view of the above, fluency tests are among the most useful tests in the clinician's armamentarium.

Performance on verbal fluency tasks reflects the integrity of the semantic system (i.e., the system responsible for storage of words and their associations), the efficiency of retrieval strategies, and the efficiency of self-monitoring and inhibition of inappropriate responses. In an early influential paper, Perret (1974) reported significant deficits on initial letter fluency in patients with focal frontal lesions whilst noting that semantic or category fluency (i.e., generating exemplars from categories such as animals, fruits, etc.) often failed to expose deficits (Newcombe, 1969). Perret (1974) and others have argued that the differential sensitivity of initial letter fluency to frontal dysfunction stems from the fact that in everyday life words are retrieved predominantly on the basis of their meaning. Thus, initial letter fluency is relatively novel and places greater demands on strategy generation and inhibition of semantic associates. In contrast, semantic fluency places greater demands on the integrity of the semantic system. Hodges, Salmon, and Butters, (1990), for example, is one of many studies in which there was evidence of a differential deficit in semantic fluency vs. initial letter fluency in DAT. This is consistent with other evidence indicating that deterioration of the semantic system is a marked feature of DAT (e.g., Grober, Buschke, Kawas, & Fuld, 1985).

Further support for this distinction comes from dual task studies of initial letter and semantic fluency in healthy participants. Martin, Wiggs, Lalonde, and Mack (1994) demonstrated that a secondary task involving motor sequencing (designed to activate frontal cortex) produced a differential deficit in initial letter fluency, whereas a secondary task involving semantic decisions regarding visual stimuli (designed to activate posterior cortex) produced a differential deficit in semantic fluency.

Some very useful normative data for the elderly have been published. Ivnik et al., (1996) present age- and education-based norms based on 743 elderly participants using the COWAT (i.e., the letters C, F, and L with 60 seconds per trial). Particularly positive features of these data are the age range covered (55–97 years) and the fact that most participants were also administered other commonly used neuropsychological instruments. This latter feature makes the data useful when attempting to construct a profile of a client's relative strengths and weaknesses as raw scores are converted to a common Scale and there is no worry that observed discrepancies are artifacts of differences between normative samples (see Section 7.07.4.2). Kozora and Cullum (1995) have published elderly norms for initial letter fluency (F, A, and S) and semantic fluency covering the age range 50–90 years. The semantic fluency test used was the supermarket item list from the Mattis Dementia Rating Scale (Mattis, 1988).

Errors on verbal fluency tasks have been afforded great significance in both the research literature and in advice offered to clinicians (Walsh, 1991). Perseverative errors (repeating a word produced previously) and rule-break errors (e.g., producing a proper noun or a word beginning with the wrong letter on initial-letter fluency trials) have been regarded as evidence of disinhibition or a failure of self-monitoring (Crowe, 1992; Walsh, 1991). Crowe (1992) has produced data to suggest that such errors tend to be associated with orbito-frontal damage, whereas impoverished but legitimate output is more indicative of damage to the lateral convexity.

Given the importance attached to such errors, there is surprisingly little normative data on errors to assist clinicians interpret an individual client's performance. Kozora and Cullum's (1995) study provided means and ranges for perseverative and rule-break errors. Although the mean and modal rates were low in all age bands, a finding consistent with a COWAT normative study conducted by Ruff, Light, Parker, and Levin (1996), the ranges are quite substantial (e.g., 0–15 in the 50–59 age group). It would be useful to know how many errors are required to exceed the 85th or 95th percentiles of the healthy population but there is no relevant published data.

There is a strong relationship between verbal fluency and general verbal ability in the general population. For example, Miller (1984) reported a correlation of 0.87 between initial letter fluency and Verbal IQ in a healthy sample ($N = 30$). Crawford, Obonsawin, and Bremner (1993) reported a lower but still substantial correlation of 0.64 between Wechsler Verbal IQ in a sample of 144 healthy participants. This indicates that an individual's premorbid verbal ability must be considered as an important factor when interpreting fluency performance in clinical populations. This was graphically illustrated by Borkowski, Benton, and Spreen (1967) who reported that the fluency performance of a "brain-damaged" sample of above average verbal IQ was significantly higher than the performance of healthy subjects of below average verbal IQ. Crawford et al. (1992c) have provided a NART-based regression equation which provides an estimate of a client's premorbid fluency performance (see Section 7.07.8 for details of the NART). This equation was derived from a healthy sample ($N = 142$) with an age range of 16–88 years.

7.07.13.3 Nonverbal Fluency

A number of attempts have been made to develop methods of assessing nonverbal fluency. Jones-Gotman and Milner (1977) developed a task they termed Design Fluency which requires examinees to draw abstract (nonsense) designs under a time constraint. This test can capture the lack of creativity and spontaneity and the perseverative tendencies which are often striking clinical features of patients with frontal damage. The disadvantages of this test include the lack of adequate normative data, the highly subjective nature of scoring, and the fact that it can be difficult to maximize cooperation, either because clients fail to grasp the nature of the task or view it as odd and lacking in relevance.

Ruff and colleagues (Ruff, 1988; Ruff, Light, & Evans, 1987) have developed a figural fluency test (hereafter RFFT) which requires the examinee to join up a repeating sequence of dots such that each design formed is unique. The procedure removes subjectivity from scoring as each design is unequivocally unique, a perseveration, or (more rarely) an error. The RFFT was normed on a sample of 358 healthy individuals with an age range of 16–70 years. Group studies of clinical populations indicate that the RFFT is sensitive to anterior cerebral damage, particularly in the right hemisphere (Lee et al., 1997; Ruff, Allen, Farrow, Nieemann, & Wylie, 1994). For example, Lee et al. found that 35% of designs produced by a group with right frontal lesions were perseverations, a figure which substantially exceeded the rate in right temporal (14%) and left frontal (17%) cases.

Motor slowing is a common consequence of cerebral damage and should be considered as a factor when interpreting the number of legitimate designs produced on the RFFT. Ruff, Evans, and Marshall (1986) have reported that significant RFFT deficits are observed in neurological samples after controlling for motor speed using a finger-tapping test. However, the appropriateness of finger-tapping as a control measure could be questioned. Crawford, Wright, and Bate (1995) used a simple motor task which required subjects to join repeated pairs of dots; the motor demands of this task are very similar to the RFFT but it does not require the generation of designs. They reported that the RFFT deficit in a head-injured sample did not significantly exceed the deficit on the motor task.

7.07.13.4 Behavioral Assessment of the Dysexecutive Syndrome

A very useful battery for assessing executive problems has been developed (Wilson, Alderman, Burgess, Emslie, & Evans, 1996; Wilson, Evans, Emslie, Alderman, & Burgess, in press). The Behavioral Assessment of the Dysexecutive Syndrome (BADS) battery consists of six subtests and was normed on a sample of 216 healthy participants with an age range of 16–87 years. As was noted previously, many of the formal neuropsychological tests can fail to detect executive problems because they are highly structured. Shallice and Burgess (1991) note "The patient typically has a single explicit problem to tackle at any one time . . . the trials tend to be very short . . . task initiation is strongly prompted by the examiner and what constitutes successful trial completion is clearly characterized" (pp. 727–728). Although some of the tests reviewed above clearly attempt to deviate from this formula, the BADS represents the most ambitious and systematic attempt to date to capture the core elements of dysexecutive syndromes.

The BADS subtests include the Rule Shifts Cards Test in which a previously established response set (responding "yes" to red cards, "no" to black) has to be inhibited in favor of responding in terms of whether or not a card matches the color of the card immediately preceding it. The Action Program Test is a planning task in which the solution requires the client to utilize various everyday materials (e.g., plastic, cork, and wire). The Modified Six Elements Test (see Shallice & Burgess, 1991) assesses scheduling and time management by requiring clients to tackle three different tasks within the time limit; there are two versions of each task and the rules prohibit tackling these contiguously.

The inter-rater reliability of the BADS is excellent; correlations between raters ranged from a low of 0.88 on one index from the Modified Six Elements Test, to unity, or near unity, on most of the other tasks. Test–retest reliability coefficients obtained from a small subset of the standardization sample (retested over a 6–12 month interval) ranged from modest to poor but this, in part, reflects the effects of range restriction.

A useful supplement to the formal BADS subtests is the Dysexecutive Questionnaire (DEX). The DEX covers dispositional and cognitive changes and comes in two forms, one for completion by the client, the other by a relative or carer. A provisional factor analysis indicates that the DEX measures three broad factors labeled as "Behavior," "Cognition," and "Emotion." Large discrepancies between client and carer reports of change are common when working with clients who have executive problems. The DEX provides one means of quantifying this lack of insight; DEX results can also form a basis for discussion when counseling clients and their families.

The DEX has been used to examine the relationship between carers' ratings of executive problems and formal test scores. Multiple regression analyses found that BADS scores were the only significant predictors of rated problems from among a set of measures which included traditional "frontal" tasks and measures of current and premorbid general intellectual ability.

7.07.14 THE DIFFERENTIAL DIAGNOSIS OF DEMENTIA VS. DEPRESSION

The differential diagnosis of dementia vs. depression has been described as "probably the knottiest problem of differential diagnosis" (Lezak, 1995), yet is a common reason for elderly patient referral to clinical psychologists. Depression is the functional disorder which can most closely resemble the early stages of a dementing condition. It is vitally important for an accurate differential diagnosis to be made because as Woods and Britton (1985) point out, "Depression can be treated; dementia typically leads to therapeutic despair." Des Rosiers (1992) reviewed 18 studies and reported that, on average, 10% of cases initially diagnosed as organic dementia were later rediagnosed as depression. In practice, the diagnosis of dementia vs. depression is made by drawing on information from a wide variety of sources, that is, previous history, informant's account, phenomenology, brain scan measures, blood biochemistry, etc. Problems emerge, however, when an individual presents with no well-documented history and no useful informant information.

Huppert (1994) has pointed out that in most studies where an early DAT group has been compared with a control group, it is likely that the controls vary from the demented subjects in ways which are rarely specified. For example, as the normal controls are usually community volunteers, they are less likely to have mobility problems than sensory deficits and are more likely to be in good health. The patients usually have been referred to specialists which means that dementia is sufficiently severe to have come to specialist attention, and second, that the patients who see specialists may not be representative of all demented subjects. As Huppert (1994) states "one implication of these considerations is that observed differences are likely to be an exaggeration (if not a distortion) of the performance which would be seen in a genuinely representative sample" (p. 295).

7.07.14.1 Delayed Verbal Recall Impairment as an Indicator of Early DAT?

Over the years there have been many attempts at using memory tests to help distinguish between dementia and depression. In order to highlight the problems associated with this approach, in the following section the Delayed Word Recall test will be used as an example.

In two studies by the consortium to establish a registry for DAT (CERAD), the investigators reported that delayed verbal recall was a highly sensitive indicator of early DAT (Morris et al., 1989; Welsh, Butters, Hughes, Mohs, & Heyman, 1992). In DAT, there appears to be a rapid rate of forgetting within the first 5–10 minutes following acquisition (Butters et al., 1988; Hart, Smith, & Swash, 1988) but a relatively normal rate of forgetting thereafter (Corkin et al., 1984; Kopelman, 1985). Knopman and Ryberg (1989) suggested that elaborative encoding may provide a substantial benefit in normal elderly individuals but not to patients with DAT. They proposed that the optimum memory test procedure for the early diagnosis of DAT should involve: (i) a study phase in which subjects were required to engage in elaborative encoding; (ii) a filled delay interval; and (iii) a test phase which involved free recall. Accordingly, Knopman and Ryberg (1989) developed the Delayed Word Recall test (DWR). Ten nouns are presented on cards, one at a time (chimney, salt, harp, button, meadow, train, flower, finger, rug, book). For each word, the subject is asked to read out the word and try to remember it and then to make up a sentence using the word. When all 10 words have been presented, the complete procedure is repeated, with subjects instructed to produce an alternative sentence. Following a five minute filled delay, free recall for the 10 words in the list is tested. Based on the delayed free recall cut-off score of less than or equal to 2 out of 10 = DAT, Knopman and Ryberg (1989) reported an overall predictive accuracy of 95% in a group of DAT patients vs. elderly control subjects. This is an extremely impressive degree of between-group separation.

In a follow-up study, Coen et al. (1996) compared a group of mildly demented DAT patients with individually matched healthy controls and found that the DWR recall measure achieved 98% overall accuracy (i.e., almost perfect between-group separation). In an extension to the original DWR, Coen et al. included a four-choice recognition memory test and, using a cut-off score of less than or equal to 9 out of 10 = DAT, greater than nine = controls, they reported that the recognition measure yielded a sensitivity of 98%, specificity of 95%, and an overall accuracy of 96%. Taken together, the results of the Knopman and Ryberg (1989) and Coen et al. studies suggests that the DWR is an extremely promising measure to aid in the early detection of DAT.

However, for any neuropsychological test index to have clinical utility, the method must be shown clearly to identify patients suffering from DAT without producing false-positives from an elderly depressed comparison group. Unfortunately, too many studies have appeared which have focused on the simple comparison between DAT and very healthy controls, and this criticism can be leveled at the Knopman and Ryberg (1989) and Coen et al. studies.

O'Carroll, Conway, Ryman, and Prentice (1997) attempted to test the utility of the DWR in separating patients with DAT from patients with major depression. Fifty patients with DAT were compared with 50 patients who fulfilled *DSM-IV* criteria for a major depressive episode. While there were highly significant differences in mean free recall and recognition scores, the between-group overlap was large (e.g., see Figure 2 for free recall results—very similar results were obtained for recognition). Put simply, if the recommended cut-off point for free recall of 2 or less out of 10 = DAT is adopted, 44% depressed patients would be misclassified as suffering from DAT. Using the less than 10 recommended cut-off score for recognition score, 48% of depressed patients would be misclassified as suffering from DAT. This study highlights the degree of cognitive impairment which often presents in elderly depressed patients which can masquerade as "pseudo-dementia" (Robbins, Joyce, & Sahakian, 1992).

7.07.14.2 Kendrick Cognitive Tests for the Elderly

The Kendrick Cognitive Tests for the Elderly (KCTE) comprise two components, a pictorial memory task (the Object Learning Test or OLT) and the Digit Copying Test (DCT). The principle underlying the use of the KCTE in differential diagnosis lies in the expected pattern of visual memory and psychomotor speed in depression vs. dementia. According to Kendrick, Gibson, and Moyes (1979), both DCT and OLT should be affected in primary dementia and not in depression. An innovation of the Kendrick test is the requirement that the measures be readministered six weeks following the initial assessment. desRosiers (1992) describes this requirement as somewhat less than practical. However, in our opinion, repeat neuropsychological assessment in cases where diagnosis is in doubt represents good clinical practice.

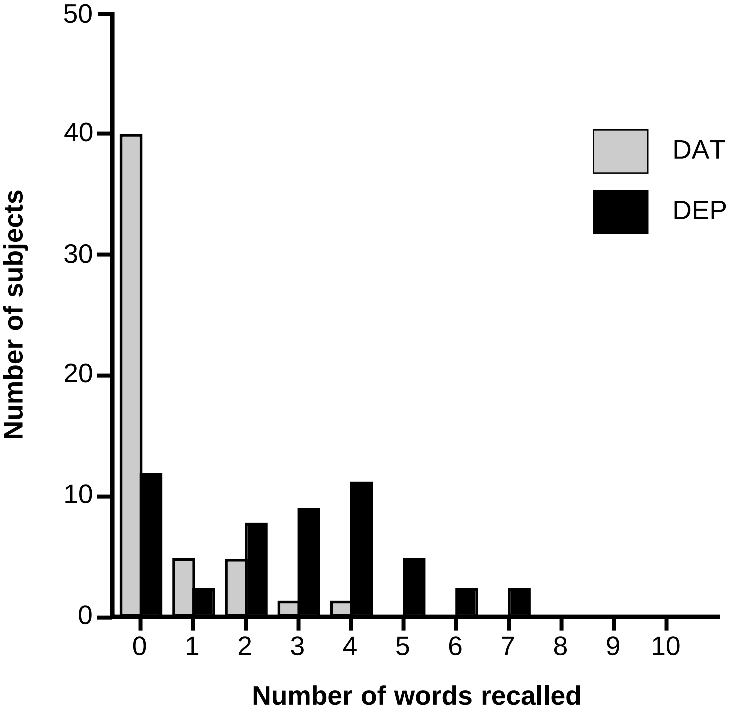

Figure 2 DWR Recall results. Using < 3 recommended cut-offs, only 2 (4%) DAT (Alzheimer) sujects would be misclassified, but 22 (44%) DEP (Depressed) would be classed as DAT (reproduced from O'Carroll et al., 1997).

One of the strongest claims regarding the KCTE is that "it has been found possible to discriminate absolutely between dementias and non-dementing subjects" (Kendrick et al., 1979, p. 329). The literature on the Kendrick test is well reviewed by desRosiers (1992). Some reports have confirmed the utility of the KCTE in differentiating a depressed sample from a demented sample. However, as is usually the case in studies of this sort, the samples consist of very well-defined groups where differential diagnosis is not a problem (Knight & Moroney, 1985). As desRosiers (1992) cautions, the validity of the KCTE should best be examined in patients for whom diagnosis is still ambiguous. A further point to consider when considering the theoretical foundation of the KCTE is the assumption that in depression, psychomotor speed, and mnemonic function are not as detrimentally affected as in early DAT. However, in depressive pseudodementia, it is precisely because both domains are seriously impaired that diagnostic difficulties arise.

7.07.14.3 Effortful vs. Automatic Processing and Memory

Weingartner, Cohen, Bunney, Ebert, and Kaye (1982) proposed that tasks that are normally accomplished automatically and require little cognitive capacity are particularly useful in distinguishing progressive dementia patients from depressed patients. Weingartner (1984, 1986) reported that depressed patients experienced difficulty on memory tasks that require elaborate or effortful organization and processing of material to be remembered, but that they do not show problems in memory tasks on which "automatic" memory encoding is presumed to be involved. In contrast, it is postulated that demented patients have difficulty with both automatic and effortful processing. In line with this proposal, Weingartner, Cohen, Murphy, Martello, and Gerdt (1981) reported that depressed patients demonstrate impairments on delayed free recall (effortful) with no impairment on recognition memory (relatively automatic), whereas demented patients perform poorly in both conditions. However, Lachner, Satzger and Engel (1994) have provided evidence which challenges the Weingartner model. They compared three groups (demented, depressed, and healthy subjects) who were in their 70s on five verbal recall and two recognition memory tasks. They found that both delayed recall and recognition after long delay were the measures which best discriminated between the demented and the depressed patient groups. It is interesting to note that the recognition measure was more discriminating than free recall, selective reminding, or serial learning, all presumably more "effortful" than a recognition task.

Lachner et al. (1994) point out that the labeling of some cognitive tasks as effortful vs. automatic is often ambiguous, and propose sensibly that "In future research, the tasks cognitive capacity requirements should be examined empirically to avoid inconsistencies by subjectively estimating capacity demands" (Lachner et al., p. 10). A particular advantage of the Lachner et al. study is their comparison of severely depressed psychiatric inpatients with a demented patient sample where 60% were still capable of independently running their household, as it is this type of comparison of mildly demented and severely depressed patients that is of particular clinical interest with regard to the problem of differential diagnosis.

7.07.14.4 Is Poor Memory Test Performance in Elderly Depressed Patients an Artifact of Poor Motivation?

In commenting on the relatively poor performance of many depressed patients with standard neuropsychological measures, many writers have attributed this impairment to the nonspecific effects of motivational deficits. Few studies have tested this hypothesis experimentally. One notable exception is the study of Richards and Ruff (1989), who randomly assigned two groups of subjects, depressed and nondepressed, to either motivation or nonmotivation conditions (motivation involved encouragement, a monitory incentive and performance feedback). Richards and Ruff (1989) reported that motivation (as measured by improvement on a simple card-sorting task) was enhanced for the subjects in the motivation condition. Crucially, motivation did not significantly affect neuropsychological test performance and the authors concluded that although depressed patients may be less motivated, this reduced motivation does not fully account for the observed cognitive impairments in depression.

7.07.14.5 Future Directions

The literature indicates that demented subjects are impaired at any task which requires the conscious recollection of newly learned material. However, much work has been carried out attempting to evaluate whether specific aspects of memory functioning are disproportionately impaired in early DAT. Divided attention tasks are particularly sensitive to DAT, that is, tasks where subjects are required to divide their attention during learning (e.g., performing a digit span test while keeping a stylus on a target on a pursuit rotor task). This would appear to be a particular deficit in "working memory," that is, the ability to hold information in a short-term store or carrying out the other processing operations (Baddeley & Wilson, 1986). Additionally, prospective memory tasks appear to be particularly impaired in early DAT. Huppert and Beardsall (1993) administered a variety of memory tasks and found that even patients with minimal dementia were disproportionately impaired on prospective memory tasks (e.g., remembering to ask for a belonging back at the end of the Rivermead Behavioural Memory Test). In the search for particularly dementia-sensitive screening tasks, a combination of divided attention and prospective memory tasks may well be worthy of further investigation.

7.07.14.6 Recommendations and Conclusions for the Differential Diagnosis of Dementia vs. Depression

Sweet, Newman, and Bell (1992) made several useful recommendations to facilitate the differential diagnosis of depressed patients from demented patients. First, they recommend administering neuropsychological measures that assess domains which the literature indicate are particularly discriminating, notably psychomotor retardation, recognition vs. recall, and easy vs. difficult paired associated learning. Second, they recommend utilization of measures of affective status (e.g., the Beck Depression Inventory or the Geriatric Depression Rating Scale). They also propose seeking collateral information through interviewing the spouse or significant other. Third, when evidence of depression is found, customary cognitive "cut-off" scores may be misleading and they propose using more stringent criteria (e.g., an additional standard deviation away from normal performance). Fourth, with depressed patients, they caution against diagnosing brain dysfunction based only on findings such as mild attentional or memory problems. For example, as psychomotor retardation in depression can easily result in a markedly impaired performance on the Stroop task (Trenerry, Crosson, DeBoe, & Leber, 1989), many depressed patients fall within the recommended cut-off score used to indicate "brain damage."

Sweet et al. (1992) also urge against differentiating depression from brain damage after only a single session with the patient and propose seeing the patient on more than one day and, if in doubt, readministering at least a portion of the measures on a second testing session (this would be in accord with the protocol of the Kendrick Cognitive Tests for the Elderly). Sweet et al. conclude "In the event that a diagnostic conclusion regarding brain dysfunction in the depressed patient appears unclear, despite all your best efforts, recommend a re-evaluation of the depressed patient after a period of pharmacological or psychotherapeutic intervention" (p. 41). This is a recommendation that is fully endorsed. Neuropsychological measures should be used in conjunction with information derived from a variety of sources (e.g., relatives and clinical history). Developments in neuroimaging precision suggest that an *in vivo* maker for DAT may become available (Jobst et al., 1994).

Given the potentially catastrophic outcome of misclassifying a treatable, depressed patient as having DAT, it has to be concluded that the available evidence indicates that no single neuropsychological measure can be recommended for use in the individual case to confidently separate early DAT from major depression in an elderly population. However, neuropsychological measures may offer useful information that can be integrated with data gathered from a variety of sources. In particular, accurate objective quantification of change over time is one of the key areas where detailed assessment of cognitive functioning can provide important diagnostic information. Furthermore, changes in neuropsychological status can offer a useful measure of treatment response in both dementia and depression.

7.07.15 REFERENCES

Albert, M. L. (1973). A simple test of visual neglect. *Neurology*, *23*, 658–664.

Albert, M. S., Heller, H. S., & Milberg, W. (1988). Changes in naming ability with age. *Psychology and Aging*, *3*, 173–178.

Anthony, J. C., Le Resche, L., Niaz, U., Von Korff, M. R., & Folstein, M. F. (1982). Limits of the Mini-Mental-State as a screening test for dementia and delirium among hospital patients. *Psychological Medicine*, *12*, 397–408.

Atkinson, L. (1991). Some tables for statistically based interpretation of WAIS-R factor scores. *Psychological Assessment*, *3*, 288–291.

Atkinson, L., Cyr, J. J., Doxey, N. C. S., & Vigna, C. M. (1989). Generalizability of WAIS-R factor structure within and between populations. *Journal of Clinical Psychology*, *45*, 124–129.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word Test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology*, *32*, 55–65.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1994). *Doors* *and people: A test of visual and verbal recall and recognition*. Bury St. Edmunds, UK: Thames Valley Test Company.

Baddeley, A., & Wilson, B. (1986). Dementia and working memory. *Quarterly Journal of Experimental Psychology*, *38*(A), 603–618.

Balota, D. A., & Duchek, J. M. (1988). Age related differences in lexical access, spreading activation, and simple pronunciation. *Psychology and Aging*, *3*, 84–93.

Barker, M. G., & Lawson, J. S. (1968). Nominal aphasia in dementia. *British Journal of Psychiatry*, *114*, 1351–1356.

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs: strategies for studying behavior change* (2nd ed.). New York: Pergamon.

Barona, A., Reynolds, C. R., & Chastain, R. (1984). A demographically based index of premorbid intelligence for the WAIS-R. *Journal of Consulting and Clinical Psychology*, *52*, 885–887.

Beardsall, L., & Huppert, F. A. (1994). Improvement in NART word reading in demented and normal older persons using the Cambridge Contextual Reading Test. *Journal of Clinical and Experimental Neuropsychology*, *16*, 232–242.

Benton, A. L. (1968). Differential behavioral effects in frontal lobe disease. *Neuropsychologia*, *6*, 53–60.

Benton, A. L. (1969). Development of a multilingual aphasia battery: Progress and problems. *Journal of the Neurological Sciences*, *9*, 39–48.

Benton, A. L., Eslinger, P. J., & Damasio, A. R. (1981). Normative observations of neuropsychological test performances in old age. *Clinical Neuropsychology, 3*, 33–42.

Benton, A. L., Hamser, K. D. S., Varney, N. R., & Spreen, O. (1983). *Contributions to neuropsychological assessment*. New York: Oxford University Press.

Benton, A. L., Hannay, H. J., & Varney, N. R. (1975). Visual perception of line direction in patients with unilateral brain disease. *Neurology*, *25*, 907–910.

Benton, A. L., & Van Allen, M. W. (1968). Impairment in facial recognition in patients with cerebral disease. *Cortex, 4*, 344–358.

Binks, M. (1989). Changes in mental functioning associated with normal ageing. In R. Tallis (Ed.), *The clinical neurology of old age* (pp. 27–39). Chichester, UK: Wiley.

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex, 14*, 129–133.

Blessed, G., Black, S. E., Butler, T., & Kay, D. W. K. (1991). The diagnosis of dementia in the elderly a comparison of CAMCOG (the cognitive section of CAMDEX), the AGECAT program, DSM-III, The Mini-Mental State Examination and some short rating Scales. *British Journal of Psychiatry, 159*, 193–198.

Bondi, M. W., Kaszniak, A. W., Bayles, K. A., & Vance, K. T. (1993). Contributions of frontal system dysfunction to memory and perceptual abilities in Parkinson's disease. *Neuropsychology, 7*, 89–102.

Borkowski, J. G., Benton, A. L., & Spreen, O. (1967). Word fluency and brain damage. *Neuropsychologia, 5*, 135–140.

Bornstein, R. A. (1986). Contribution of various neuropsychological measures to detection of frontal lobe impairment. *International Journal of Clinical Neuropsychology, 8*, 18–22.

Borod, J. C., Goodglass, H., & Kaplan, E. (1980). Normative data on the Boston Diagnostic Aphasia Examination, Parietal Lobe Battery, and the Boston Naming Test. *Journal of Clinical Neuropsychology, 2*, 209–215.

Brinkman, S. D., & Braun, P. (1984). Classification of dementia patients by a WAIS profile related to central cholinergic deficiencies. *Journal of Clinical Neuropsychology, 6*, 393–400.

Britton, P. G., & Savage, R. D. (1966). A short form of the WAIS for use with the aged. *British Journal of Psychiatry, 112*, 417–418.

Brooks, D. N. (1989). Closed head trauma: assessing the common cognitive problems. In M. D. Lezak (Ed.), *Assessment of the behavioral consequences of head trauma. Frontiers of Clinical Neuroscience* (Vol. 7, pp. 61–85). New York: Liss.

Bryan, J., Luszcz, M. A., & Crawford, J. R. (1997). Verbal knowledge and speed of information processing as mediators of age differences in verbal fluency performance among older adults. *Psychology and Aging, 12*, 473–478.

Burgess, A. (1991). Profile analysis of the Wechsler Intelligence Scales: a new index of subtest scatter. *British Journal of Clinical Psychology, 30*, 257–263.

Burke, D. M., Mackay, D. G., Worthley, J. A., & Wade, E. (1991). On the tip of the tongue: what causes word finding failures in young and older adults? *Journal of Memory and Language, 30*, 542–579.

Burke, D. M., White, H., & Diaz, D. L. (1987). Semantic priming in young and older adults: evidence for age-constancy in automatic and attentional processes. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 79–88.

Butters, B., Salmon, D. P., Cullum, C. M., Cairns, P., Troster, A. I., & Jacobs, D. (1988). Differentiation of amnesic and demented patients with the Wechsler Memory Scale-Revised. *Clinical Neuropsychologist, 2*, 133–148.

Caine, E. D. (1992). *Nomenclature and diagnosis of cognitive disorders: a US perspective*. Paper presented at the Abstract from Age-related Cognitive Disorders Conference, Nice, France.

Canavan, A. G. M., Dunn, G., & McMillan, T. M. (1986). Principal components of the WAIS-R. *British Journal of Clinical Psychology, 25*, 81–35.

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79*, 380–385.

Christensen, H., & MacKinnon, A. (1992). Wechsler intelligence scale profiles in Alzheimer type dementia and healthy ageing. *International Journal of Geriatric Psychiatry, 7*, 241–246.

Cockburn, J., & Smith, P. T. (1989). *The Rivermead Behavioural Memory Test (Suppl. 3: Elderly people)*. Bury St. Edmunds, England: Thames Valley Test Company.

Coen, R. F., Swanwick, G. R. J., Maguire, C., Kirby, M., Lawlor, B. A., Walsh, J. B., & Coakley, D. (1996). Memory impairment in Alzheimer's disease: replication and extension of the delayed word recall (DWR) test. *Irish Journal of Psychological Medicine, 13*, 55–58.

Corkin, S., Growdon, J., Nissen, M. J., Huff, F. J., Freed, D. M., & Sagar, H. J. (1984). Recent advances in the neuropsychological study of Alzheimer's disease. In R. J. Wurtman, S. Corkin, & J. H. Growdon, (Eds.), *Advances in basic research and therapies* (pp. 75–94). New York: Centre for Brain Sciences and Metabolism Trust.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, D. M. Parker & W. McKinlay, (Eds.), *A handbook of neuropsychological assessment* (pp. 21–49). Hove, UK: Erlbaum.

Crawford, J. R. (1997). WAIS-R short-forms: Assessing the statistical significance of subtest differences. *British Journal of Clinical Psychology, 36*, 601–608.

Crawford, J. R., & Allan, K. M. (1997). Estimating premorbid IQ with demographic variables: regression equations derived from a U.K. sample. *The Clinical Neuropsychologist, 11*, 192–197.

Crawford, J. R., Allan, K. M., & Jack, A. M. (1992a). Short-forms of the UK WAIS-R: regression equations and their predictive accuracy in a general population sample. *British Journal of Clinical Psychology, 31*, 191–202.

Crawford, J. R., Allan, K. M., McGeorge, P., & Kelly, S. M. (1997). Base rate data on the abnormality of subtest scatter for WAIS-R short-forms. *British Journal of Clinical Psychology, 36,* 433–444.

Crawford, J. R., Allan, K. M., Stephen, D. W., Parker, D. M., & Besson, J. A. O. (1989a). The Wechsler Adult Intelligence Scale-Revised (WAIS-R): Factor structure in a UK sample. *Personality and Individual Differences, 10*, 1209–1212.

Crawford, J. R., Besson, J. A. O., & Ebmeier, K. P. (1990). Dementia: The role of neurotransmitter deficits. In T. G. Dinan (Ed.), *Principles and practice of biological psychiatry* (pp. 105–133). London: Clinical Neuroscience Publishers.

Crawford, J. R., Besson, J. A. O., & Parker, D. M., (1988). Estimation of permorbid intelligence in organic conditions. *British Journal of Psychiatry, 153*, 178–181.

Crawford, J. R., Johnson, D. A., Mychalkiw, B., & Moore, J. W. (in press-a). WAIS-R performance following closed head injury: A comparison of the clinical utility of summary IQs, factor scores and subtest scatter indices. *The Clinical Neuropsychologist, 11*.

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992c). Verbal fluency: a NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology, 31*, 327–329.

Crawford, J. R., Obonsawin, M. C., & Allan, K. M. (in press-b). PASAT and components of WAIS-R performance: Convergent and discriminant validity. *Neuropsychological Rehabilitation*.

Crawford, J. R., Obonsawin, M. C., & Bremner, M. (1993). Frontal lobe impairment in schizophrenia: Relationship to intellectual functioning. *Psychological Medicine, 23*, 787–790.

Crawford, J. R., Parker, D. M., & McKinlay, W., & (Eds.). (1992b). *A handbook of neuropsychological assessment*. Hove, UK: Erlbaum.

Crawford, J. R., Sommerville, J., & Robertson, I. H. (1997). Assessing the reliability and abnormality of subtest differences on the Test of Everyday Attention. *British Journal of Clinical Psychology, 36,* 609–617.

Crawford, J. R., Stewart, L. E., Cochrane, R., Foulds, J., Besson, J. A. O., & Parker, D. M. (1989b). Estimating premorbid IQ from demographic variables: a regression equation derived from a UK sample. *British Journal of Clinical Psychology, 28*, 275–278.

Crawford, J. R., Stewart, L. E., & Moore, J. W. (1989c). Demonstration of savings on the AVLT and development of a parallel form. *Journal of Clinical and Experimental Neuropsychology, 11*, 975–981.

Crawford, J. R., Wright, R., & Bate, A. (1995). Verbal, figural and ideational fluency in CHI. *Journal of the International Neuropsychological Society, 1*, 321.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rhinehart & Winston.

Crook, T., Bartus, R. T., Ferris, S. H., Whitehouse, P., Cohen, G. D., & Gershon, S. (1986). Age-associated memory impairment: proposed diagnostic criteria and measures of clinical change: Report of a National Institute of Mental Health Work Group. *Developmental Neuropsychology, 2*, 261–276.

Crowe, S. F. (1992). Dissociation of two frontal lobe syndromes by a test of verbal fluency. *Journal of Clinical and Experimental Neuropsychology, 14*, 327–339.

Daigneault, S., & Braun, C. M. J. (1993). Working memory and the self-ordered pointing task: Further evidence of early prefrontal decline in normal aging. *Journal of Clinical and Experimental Neuropsychology, 15*, 881–895.

Daigneault, S., Braun, C. M. J., & Whitaker, H. A. (1992). Early effects of normal aging on perserverative and non-perseverative prefrontal measures. *Developmental Neuropsychology, 8*, 99–114.

Davis, H. P., & Bernstein, P. A. (1992). Age-related changes in explicit and implicit memory. In L. R. Squire & N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 249–261). New York: Guilford Press.

De Renzi, E., & Faglioni, P. (1978). Normative data and screening power of a shortened version of the Token Test. *Cortex, 14*, 41–49.

De Renzi, E., & Vignolo, L. A. (1962). The Token Test: A sensitive test to detect disturbances in aphasics. *Brain, 85*, 665–678.

D'Elia, L. F., Boone, K. B., & Mitrushima, A. M. (1995). Handbook of normative data for neuropsychological assessment. New York: Oxford University Press.

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (1987). *California Verbal Learning Test—Adult Version: Research Edition Manual Version 1*. San Antonio, CA: The Psychological Corporation.

Della Sala, S. (1988). Cognitive deficits of Parkinsonians and Occam's Razor. *Europa Medicophysica, 24*, 1–22.

Della Sala, S., Laiacona, M., Trivelli, C., & Spinnler, H. (1995). Poppelreuter-Ghent overlapping figures test. Its sensitivity to age, and its clinical use. *Archives of Clinical Neuropsychology, 10*(6), 511–534.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Development Review, 12*, 45–75.

desRosiers, G. (1992). Primary or depressive dementia: Psychometric assessment. *Clinical Psychology Review, 12*, 307–343.

Dick, J. P. R., Guiloff, R. J., Stewart, A., Blackstock, J., Bielawska, C., Paul, E. A., & Marsden, C. D. (1984). Mini Mental State Examination in neurological patients. *Journal of Neurology, Neurosurgery, and Psychiatry, 47*, 496–499.

Drachman, D. A., & Leavitt, J. (1974). Human memory and the cholinergic system: A relationship to aging? *Archives of Neurology, 30*, 113–121.

Ellis, H. D. (1992). Assessment of deficits in facial processing. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 141–150). Hove, UK: Lawrence Erlbaum.

Elwood, R. W. (1991). The Wechsler Memory Scale-Revised: Psychometric characteristics and clinical application. *Neuropsychology Review, 2*, 179–201.

Eslinger, P. J., & Benton, A. L. (1983). Visuoperceptual performance in aging and dementia. Clinical and theoretical implications. *Journal of Clinical Neuropsychology, 5*, 213–220.

Farver, P. F., & Farver, T. B. (1982). Performance of normal older adults on tests designed to measure parietal lobe functions. *American Journal of Occupational Therapy, 36*, 444–449.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-Mental-State: A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatry Research, 12*, 189–198.

Franzen, M. D. (1989). *Reliability and validity in neuropsychological assessment*. New York: Plenum.

Freedman, L., & Dexter, L. E. (1991). Visuospatial ability in cortical dementia. *Journal of Clinical and Experimental Neuropsychology, 13*(5), 677–690.

Fuld, P. A., (1984). Test profile of cholinergic dysfunction and of Alzheimer-type dementia. *Journal of Clinical Neuropsychology, 6*, 380–392.

Fuster, J. M. (1989). *The prefrontal cortex* (2nd ed.). New York: Raven Press.

Gallagher, A. J. (1979). Temporal reliability of aphasic performance on the Token Test. *Brain and Language, 7*, 34–41.

Gardner, R., Oliver-Munoz, S., Fisher, L., & Empting, L.

(1981). Mattis Dementia Rating Scale: Internal reliability study using a diffusely impaired population. *Journal of Clinical Neuropsychology, 3*, 271–275.

Gauthier, L., Dehaut, F., & Joanette, Y. (1989). The Bells Test: A quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology, 11*, 49–54.

Geffen, G., Moar, K. J., O'Hanlon, A. P., Clark, C. R., & Geffen, L. B. (1990). Performance measures of 16- to 86-year-old males and females on the Auditory Verbal Learning Test. *Clinical Neuropsychologist, 4*, 45–63.

Gfeller, J. D., & Rankin, E. J. (1991). The WAIS-R profile as a cognitive marker of Alzheimer's Disease: A misguided venture? *Journal of Clinical and Experimental Neuropsychology, 13*, 629–636.

Giles, E., Patterson, K., & Hodges, J. R. (1996). Performance on the Boston Cookie Theft Picture Description Task in patients with early dementia of the Alzheimer's type: Missing information. *Aphasiology, 10*, 395–408.

Glosser, G., & Deser, T. (1992). A comparison of changes in macrolinguistic and microlinguistic aspects of discourse production in normal aging. *Journals of Gerontology, 47*, 266–272.

Goodglass, H. (1980). Naming disorders in aphasia and ageing. In L. K. Obler & M. L. Albert (Eds.), *Language and communication in the elderly* (pp. 37–45). Lexington, MA: Heath.

Goodglass, H., & Kaplan, E. (1983). *Boston Diagnostic Aphasia Examination (BDAE)*. Philadelphia: Lea & Febiger.

Goodglass, H., & Kaplan, E. (1987). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea & Febiger.

Gottschaldt, K. (1928). Uber den Einfluss der Erfahrung auf die Wahrnehmung von Figuren. *Psychologische Forschung, 18*, 18–317.

Grant, D. A., & Berg, E. A. (1948). A behavioural analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology, 38*, 404–411.

Greene, J. D. W., Baddeley, A. D., & Hodges, J. R. (1996). Analysis of the episodic memory deficit in early Alzheimer's disease: Evidence from the Doors and People Test. *Neuropsychologia, 34*, 537–551.

Grober, E., Buschke, H., Kawas, C., & Fuld, P. (1985). Impaired ranking of semantic attributes in dementia. *Brain and Language, 26*, 276–286.

Gronwall, D. (1977). Paced auditory serial-addition task: a measure of recovery from concussion. *Perceptual and Motor Skills, 44*, 367–373.

Halligan, P. W., & Marshall, J. C. (1988). How long is a piece of string? A study of line bisection in a case of visual neglect. *Cortex, 24*, 321–328.

Halligan, P. W., & Marshall, J. C. (1989). Line bisection in visuo-spatial neglect: Disproof of a conjecture. *Cortex, 25*, 517–521.

Halligan, P. W., & Marshall, J. C. (1991). Left neglect for near but not far space in man. *Nature, 350*, 498–500.

Halligan, P. W., & Robertson, I. H. (1992). The assessment of unilateral neglect. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 151–175). Hove, UK: Lawrence Erlbaum.

Hart, S., Smith, C. M., & Swash, M. (1988). Word fluency in patients with early dementia of Alzheimer type. *British Journal of Clinical Psychology, 27*, 115–124.

Heaton, R. K. (1981). *Wisconsin Card Sorting Test manual.* Odessa, FL: Psychological Assessment Resources.

Hodges, J. R., Patterson, K., Graham, N., & Dawson, K. (1996). Naming and knowing in dementia of Alzheimer's type. *Brain and Language, 54*, 302–325.

Hodges, J. R., Salmon, D. P., & Butters, N. (1990). Differential impairment of semantic and episodic memory in Alzheimer's and Huntington's diseases: A controlled prospective study. *Journal of Neurology, Neurosurgery and Psychiatry, 53*, 1089–1095.

Hooper, H. E. (1983). *Hooper Visual Organization Test (VOT)* Los Angeles: Western Psychological Services.

Huff, F. J., Boller, F., Lucchelli, F., Querriera, R., Beyer, J., & Belle, S. (1987). The neurologic examination in patients with probably Alzheimer's disease. *Archives of Neurology, 44*, 929–932.

Huff, F. J., Collins, C., Corkin, S., & Rosen, T. J. (1986). Equivalent forms of the Boston Naming Test. *Journal of Clinical and Experimental Psychology, 8*, 556–562.

Huppert, F. A. (1994). Memory function in dementia and normal aging: Dimension or dichotomy? In F. A. Huppert, C. Brayne, & D. W. O'Connor (Eds.), *Dementia and normal aging* (pp. 291–330). Cambridge: Cambridge University Press.

Huppert, F. A., & Beardsall, L. (1993). Prospective memory impairment as an early indicator of dementia. *Journal of Clinical and Experimental Neuropsychology, 15*, 805–821.

Huppert, F. A., Jorm, A. F., Brayne, C., Girling, D. M., Barkley, C., Beardsall, L., & Paykel, E. S. (submitted). Psychometric properties of the CAMCOG and its efficacy in the diagnosis of dementia.

Huppert, F. A., & Kopelman, M. D. (1989). Rates of forgetting in normal ageing: A comparison with dementia. *Neuropsychologia, 27*, 849–860.

Huttenlocher, J., Haight, W., & Bryk, A. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236–248.

Ishiai, S., Okiyama, R., Koyama, Y., & Seki, K. (1996). Unilateral spatial neglect in Alzheimer's disease. A line bisection study. *Acta Neurologica Scandinavica, 93*, 219–224.

Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., & Petersen, R. C. (1996). Neuropsychological tests' norms above age 55: COWAT, BNT, MAE Token, WRAT-R reading, AMNART, STROOP, TMT, and JLO. *The Clinical Neuropsychologist, 10*, 262–278.

Ivnik, R. J., Malec, J. F., Smith, G. E., Tangalos, E. G., Peterson, R. C., Kokmen, E., & Kurland, L. T. (1992). Mayo's older Americans normative studies: WAIS-R norms for ages 56 to 97. *The Clinical Neuropsychologist, 6*, 1–30.

Jobst, K. A., Smith, A. D., Szatmari, M., Esiri, M. M., Jaskowski, A., Hindley, N., McDonald, B., & Molyneux, A. J. (1994). Rapidly progressing atrophy of medial temporal lobe in Alzheimer's disease. *Lancet, 343*, 829–830.

Jones-Gotman, M., & Milner, B. (1977). Design fluency: The invention of nonsense drawings after focal frontal lesions. *Neuropsychologia, 15*, 653–674.

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1978). *The Boston Naming Test*. (Experimental ed.). Boston: Kaplan & Goodglass.

Kaplan, E. F., Goodglass, H., & Weintraub, S. (1983). *The Boston Naming Test* (2nd ed.). Philadelphia: Lea & Febiger.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.

Kemper, S. (1992). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 213–272). Hillsdale, NJ: Erlbaum.

Kempler, D. (1988). Lexical and pantomime abilities in Alzheimer's disease. *Aphasiology, 2*, 147–159.

Kempler, D., Andersen, E. S., & Henderson, V. W. (1995). Linguistic and attentional contributions to anomia in Alzheimer's disease. *Neuropsychiatry, Neuropsychology, and Behavioural Neurology, 8*(1), 33–37.

Kendrick, D. C., Gibson, A. J., & Moyes, I. C. A. (1979). The revised Kendrick Battery: Clinical studies. *British*

Journal of Social and Clinical Psychology, 18, 329–340.
Kirk, A., & Kertesz, A. (1991). On drawing impairment in a Alzheimer's disease. *Archives of Neurology, 48*, 73–77.
Kittner, S. J., White, L. R., Farmer, M. E., Wolz, M, Kaplan, E., Moes, E., Brody, J. A., & Feinleib, M. (1986). Methodological issues in screening for dementia: The problem of education adjustment. *Journal of Chronic Diseases, 39*, 163–170.
Klesges, R. C., Wilkening, G. N., & Golden, C. J. (1981). Premorbid indices of intelligence: A review. *Clinical Neuropsychology, 3*, 32–39.
Knight, R. G., & Godfrey, H. P. D. (1984). Assessing the significance of differences between subtests on the Wechsler Adult Intelligence Scale-Revised. *Journal of Clinical Psychology, 40*, 808–810.
Knight, K. R., & Moroney, B. (1985). An investigation of the validity of the Kendrick Battery for the detection of dementia in the elderly. *International Journal of Clinical Psychology, 7*, 147–151.
Knight, R. G., & Shelton, E. J. (1983). Tables for evaluating predicted retest changes in Wechsler Adult Intelligence Scale scores. *British Journal of Clinical Psychology, 22*, 77–81.
Knopman, D. S., & Ryberg, S. (1989). A verbal memory test with high predictive accuracy for dementia of the Alzheimer type. *Archives of Neurology, 46*, 141–145.
Kopelman, M. D. (1985). Rates of forgetting in Alzheimer-type dementia and Korsakoff's syndrome. *Neuropsychologia, 23*, 623–638.
Kozora, E., & Cullum, C. M. (1995). Generative naming in normal aging: Total output and qualitative changes using phonemic and semantic constraints. *The Clinical Neuropsychologist, 9*, 313–320.
Kukull, W. A., Larson, E. B., Teri, L., Bowen, J., McCormick, W., & Pfanschmidt, M. L. (1994). The Mini-Mental State Examination score and the clinical diagnosis of dementia. *Journal of Clinical Epidemiology, 47*, 1061–1067.
Lachner, G., Satzger, W., & Engel, R. R. (1994). Verbal memory tests in the differential diagnosis of depression and dementia: discriminative power of seven test variations. *Archives of Clinical Neuropsychology, 9*, 1–13.
Leckliter, I. N., Matarazzo, J. D., & Silverstein, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology, 42*, 332–342.
Lee, G. P., Strauss, E., Loring, D. W., McCloskey, L., Haworth, J. M., & Lehman, R. A. W. (1997). Sensitivity of figural fluency on the Five Point Test to focal neurological dysfunction. *The Clinical Neuropsychologist, 11*, 59–68.
Lees, A. J., & Smith, E. (1983). Cognitive deficits in the early stages of Parkinson's disease. *Brain, 106*, 257–270.
Levin, B. E., Llabre, M. M., Reisman, S., Weiner, W. J., Sanchez Ramos, J., Singer, C., & Brown, M. C. (1991). Visuospatial impairment in Parkinson's disease. *Neurology, 41*, 365–369.
Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). New York: Oxford University Press.
Lezak, M. D. (1988a). Neuropsychological tests and assessment techniques. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (Vol. 1, pp. 47–68). Amsterdam: Elsevier.
Lezak, M. D. (1988b). IQ: R.I.P. *Journal of Clinical & Experimental Neuropsychology, 10*, 351–361.
Lezak, M. D. (1991). Identifying neuropsychological deficits. In R. G. Lister & H. J. Weingartner (Eds.), *Perspectives on cognitive neuroscience* (pp. 357–367). New York: Oxford University Press.
Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.
Light, L. L. (1992). The organization of memory in old age. In F. I. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 111–166). Hillsdale, NJ: Erlbaum.
Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging, 3*, 339–355.
Logsdon, R. G., Teri, L., Williams, D. E., Vitiello, M. V., & Prinz, P. N. (1989). The WAIS-R profile: a diagnostic tool for Alzheimer's disease? *Journal of Clinical and Experimental Neuropsychology, 11*, 892–898.
Magni, E., Binetti, G., Bianchetti, A., Rozzini, R., & Trabucchi, M. (1996). Mini Mental State Examination. A normative study in Italian elderly population. *European Journal of Neurology, 3*(3), 198–202.
Malec, J. F., Ivnik, B. J., Smith, G. E., Tangalos, G. G. (1992). Mayo's older American normative studies: Utility of corrections for age and education for the WAIS-R. *The Clinical Neurospychologist, 6* (Suppl.), 31–47.
Martin, A., Wiggs, C. L., Lalonde, F., & Mack, C. (1994). Word retrieval to letter and semantic cues: a double dissociation in normal subjects using interference tasks. *Neuropsychologia, 32*, 1487–1494.
Massman, P. J., & Bigler, E. D. (1993). A quantitative review of the diagnostic utililty of the WAIS-R Fuld profile. *Archives of Clinical Neuropsychology, 8*, 417–428.
Matarazzo, J. D., Daniel, M. H., Prifitera, A., & Herman, D. O. (1988). Inter-subtest scatter in the WAIS-R standardization sample. *Journal of Clinical Psychology, 44*, 940–950.
Matarazzo, J.D., & Herman, D. O. (1985). Clinical uses of the WAIS-R: Base rates of differences between VIQ and PIQ in the WAIS-R standardisation sample. In B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement and applications* (pp. 899–932). New York: Wiley.
Mattis, S. (1976). Mental status examination for organic mental syndrome in the elderly patient. In L. Bellack & T. B. Karasu (Eds.), *Geriatric psychiatry* (pp. 77–121). New York: Grune & Stratton.
Mattis, S. (1988). *Dementia Rating Scale*. Odessa, FL: Psychological Assessment Resources.
Maxwell, A. E. (1960). Obtaining factor scores on the Wechsler Adult Intelligence Scale. *Journal of Mental Science, 106*, 1060–1062.
Mayes, A., & Warburg, R. (1990). Memory assessment in clinical practice and research. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 73–101). Hove, UK: Erlbaum.
Maylor, E. A. (1990). Recognizing and naming faces. Aging, memory retrieval, and the tip of the tongue state. *Journal of Gerontology, 45*(6), 215–226.
McCarthy, R. A., & Warrington, E. K. (1986). Visual associative agnosia: A clinico-anatomical study of a single case. *Journal of Neurology, Neurosurgery and Psychiatry, 49*, 1233–1240.
McKinlay, W. W. (1992). Assessment of the head-injured for compensation. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 381–392). Hove, UK: Erlbaum.
McLean, J. E., Reynolds, C. R., & Kaufman, A. S. (1990). WAIS-R subtest scatter using the profile variability index. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 2*, 289–292.
McSweeny, A. J., Naugle, R. I., Chelune, G. J., & Lüders, H. (1993). "T Scores for Change": an illustration of a regression approach to depicting change in clinical neuropsychology. *The Clinical Neuropsychologist, 7*, 300–312.
Miceli, G., Caltagirone, C., Gainotti, G., Masullo, C., & Silveri, M. C. (1981). Neuropsychological correlates of localized cerebral lesions in nonaphasic brain-damaged patients. *Journal of Clinical Neuropsychology, 3*, 53–63.

Miller, E. (1984). Verbal fluency as a function of a measure of verbal intelligence and in relation to different types of cerebral pathology. *British Journal of Clinical Psychology, 23*, 53–57.

Miller, E. (1992). Some basic principles of neuropsychological assessment. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 7–20). Hove, UK: Lawrence Erlbaum.

Milner, B. (1964). Some effects of frontal lobectomy in man. In J. M. Warren & K. Alkert (Eds.), *The frontal granular cortex and behaviour* (pp. 313–331). New York: McGraw-Hill.

Mitrushina, M., & Satz, P. (1995). Repeated testing of normal elderly with the Boston Naming Test. *Aging: Clinical and Experimental Research, 7*(2), 123–127.

Mittenberg, W., Seidenberg, M., O'Leary, D. S., & DiGiulio, D. V. (1989). Changes in cerebral functioning associated with normal aging. *Journal of Clinical and Experimental Neuropsychology, 11*, 918–932.

Montgomery, K., & Costa, L. (1973). *Neuropsychological test performance of a normal elderly sample*. Paper presented at the Annual Meeting of the International Neuropsychological Society, Mexico City, Mexico.

Montgomery, K. M. (1982). *A normative study of neuropsychological test performance of a normal elderly sample*. Unpublished Master's thesis, University of Victoria, British Columbia, Canada.

Monti, L. A., Reminger, S. L., Gabrieli, J. D. E., Rinaldi, J. A., Wilson, R. S., & Fleischman, D. A. (1996). Differential effects of aging and Alzheimer's disease on conceptual implicit and explicit memory. *Neuropsychology, 10*, 101–112.

Morris, J. C., Heyman, A., Mohs, R. C., Hughes, J. P., van Belle, G., Fillenbaum, G., Mellits, E. D., & Clark, C. (1989). The consortium to establish a registry for Alzheimer's disease (CERAD) 1. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology, 39*, 1159–1165.

Mountain, M. A., & Snow, W. G. (1993). Wisconsin Card Sorting Test as a measure of frontal pathology: A review. *The Clinical Neuropsychologist, 7*, 108–118.

Nebes, R. D., Brady, C. G., & Huff, F. J. (1989). Automatic and attentional mechanisms of semantic priming in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology, 11*, 219–230.

Neils, J., Brennan, M. M., Cole, M., Boller, F., & Gerdeman, B. (1988). The use of phonemic cueing with Alzheimer's disease patients. *Neuropsychologia, 26*, 351–354.

Nelson, H. E. (1990). Review of the Recognition Memory Test. In J. R. Beech & L. Harding (Eds.), *Assessment of the elderly* (pp. 126–127). Windsor, UK: NFER-Nelson.

Nelson, H. E., & O'Connell, A. (1978). Dementia: the estimation of pre-morbid intelligence levels using the new adult reading test. *Cortex, 14*, 234–244.

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test manual* (2nd ed.). Windsor: NFER-Nelson.

Newcombe, F. (1969). *Missile wounds of the brain: A study of psychological deficits*. London: Oxford University Press.

Newcombe, F., Oldfield, R. C., Ratcliff, G. G., & Wingfield, A. (1971). Recognition and naming of object-drawing by men with focal brain wounds. *Journal of Neurology, Neurosurgery and Psychiatry, 34*, 329–340.

Nicholas, M., Obler, L., Albert, M., & Goodglass, H. (1985). Lexical retrieval in healthy aging. *Cortex, 21*, 595–606.

Obler, L., & Albert, M. (1978). *Action Naming Test*. (Experimental ed.). Boston: VA Medical Center.

Obler, L. K., & Albert, M. L. (1985). Language skills across adulthood. In J. Birren & K. W. Schaie (Eds.), *The psychology of aging* (pp. 134–156). New York: Van Nostrand Reinhold.

Obler, L., Albert, M., & Goodglass, H. (1981). *Word finding difficulties in aging and dementia*. Paper presented at the Proceedings of the Gerontology Society of America Meetings. Toronto, ON.

O'Brien, J. T., & Levy, R. (1992). Age associated memory impairment. *British Medical Journal, 304*, 5–6.

O'Carroll, R. (1995). The assessment of premorbid ability: A critical review: *Neurocase, 1*, 83–89.

O'Carroll, R. E., & Badenoch, L. D. (1994). The inter-rater reliability of the Wechsler Memory Scale-Revised Visual Memory test. *British Journal of Clinical Psychology, 33*, 208–210.

O'Carroll, R. E., Baikie, E. M., & Whittick, J. E. (1987). Does the National Adult Reading Test hold in dementia? *British Journal of Clinical Psychology, 26*, 315–316.

O'Carroll, R. E., Conway, S., Ryman, A., & Prentice, N. (1997). Performance on the Delayed Word Recall Test (DWR) fails to clearly differentiate between depression and Alzheimer's disease in the elderly. *Psychological Medicine, 27,* 967–971.

O'Carroll, R. E., Curran, S. M., Ross, M., Murray, C., Riddle, W., Moffoot, A. P. R., Ebmeier, K. P., & Goodwin, G. M. (1994). The differentiation of major depression from dementia of the Alzheimer-type using within-subject neuropsychological discrepancy analysis. *British Journal of Clinical Psychology, 33*, 23–32.

Orgass, B. (1976). Eine Revision des Token Test II. Validitatsnachweis, Normierung und Standardisierung. *Diagnostica, 22*, 141–156.

Osterreith, P. A. (1944). Le test de copie d'une figure complexe. *Archives de Psychologie, 30*, 206–256.

Paolo, A. M., & Ryan, J. J. (1995). Selecting WAIS-R norms for persons 75 years and older. *The Clinical Neuropsychologist, 9*, 44–49.

Parker, D. M., & Crawford, J. R. (1992). Assessment of frontal lobe function. In J. R. Crawford, D. M. Parker, & W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 267–291). Hove, UK: Lawrence Erlbaum.

Patterson, K., Graham, N., & Hodges, J. R. (1994). Reading in dementia: A preserved ability? *Neuropsychology, 8*, 395–407.

Pauker, J. D. (1988). Constructing overlapping cell tables to maximise the clinical usefulness of normative test data: rationale and an example from neuropsychology. *Journal of Clinical Psychology, 44*, 930–933.

Payne, R. W., & Jones, G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13*, 115–121.

Peaker, A., & Stewart, L. E. (1989). Rey's Auditory Verbal Learning Test: A review. In J. R. Crawford & D. M. Parker (Eds.), *Developments in clinical and experimental neuropsychology* (pp. 219–236). New York: Plenum.

Pentore, R., & Venneri, A. (1994). Visuo-spatial neglect in Alzheimer's disease. *Journal of Neurology, 241*(Suppl. 1), 57–58.

Perret, E. (1974). The left frontal lobe of man and the suppression of habitual responses in verbal categorical behaviour. *Neuropsychologia, 12*, 323–330.

Perry, E. K., Tomlinson, B. E., Blessed, G., Bergmann, K., Gibson, P. H., & Perry, R. H. (1978). Correlation of cholinergic abnormalities with senile plaques and mental test scores in senile dementia. *British Medical Journal, 2*, 1457–1459.

Pi, J., Olive, J. M., & Esteban, M. (1994). Mini Mental State examination: Association of the score obtained with the age and degree of alphabetization in an elderly population. *Medicina Clinica, 103*(17), 641–644.

Poppelreuter, W. (1917). *Die psychischen Schadigungen durch Kopfschuss im Kriege*. Leipzig: Leopold Voss.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25–42.

Pratt, M. W., Boyes, C., Robins, S., & Manchester, J. (1989). Telling tales: Aging, working memory, and the narrative cohesion of story retellings. *Developmental Psychology, 25*, 628–635.

Rabbitt, P. (1990). Applied cognitive gerontology: Some problems, methodologies and data. *Applied Cognitive Psychology, 4*, 225–246.

Rabbitt, P. M. A. (1968). Channel capacity, intelligibility and immediate memory. *Quarterly Journal of Experimental Psychology, 20*, 241–248.

Reitan, R. M. (1986). Theoretical and methodological bases of the Halstead-Reitan Test Battery. In I. Grant & K. M. Adams (Eds.), *Neuropsychological assessment of neuropsychiatric disorders* (pp. 3–30). New York: Oxford University Press.

Reitan, R. M., & Wolfson, D. (1994). A selective and critical review of neuropsychological deficits and the frontal lobes. *Neuropsychology Review, 4*, 161–198.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaires de France.

Reynolds, C. R., Willson, V. L., & Clark, P. L. (1983). A four-test short form of the WAIS-R for clinical screening. *Clinical Neuropsychology, 5*, 111–116.

Richards, P. M., & Ruff, R. M. (1989). Motivational effects on neuropsychological functioning: Comparison of depressed versus nondepressed individuals. *Journal of Consulting and Clinical Psychology, 57*, 396–402.

Ripich, D. N., & Terrell, B. Y. (1988). Patterns of discourse: Cohesion and coherence in Alzheimer's disease. *Journal of Speech and Hearing Disorders, 53*, 8–15.

Robbins, T. W., Joyce, E. M., & Sahakian, B. J. (1992). Neuropsychology and imaging. In E. S. Paykel (Ed.), *Handbook of affective disorders* (2nd ed., pp. 289–309). Edinburgh, UK: Churchill Livingstone.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1994). *The Test of Everyday Attention*. Bury St. Edmunds, UK: Thames Valley Test Company.

Robertson, I. H., Ward, T., Ridgeway, V., & Nimmo-Smith, I. (1996). The structure of normal human attention: The Test of Everyday Attention. *Journal of the International Neuropsychological Society, 2*, 525–534.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry, 141*, 1356–1364.

Roth, M., Tym, E., Mountjoy, C. Q., Huppert, F. A., Hendrie, J., Verma, S., & Goddard, R. (1986). CAMDEX: A standardized instrument for the diagnosis of mental disorder in the elderly with special reference to the early detection of dementia. *British Journal of Psychiatry, 149*, 698–709.

Ruff, R. M. (1988). *Ruff Figural Fluency Test administration manual*. San Franscisco: Neuropsychological Resources.

Ruff, R. M., Allen, C. C., Farrow, C. E., Nieemann, H., & Wylie, T. (1994). Figural fluency: Differential impairment in patients with left versus right frontal lobe lesions. *Archives of Clinical Neuropsychology, 9*, 41–56.

Ruff, R. M., Evans, R., & Marshall, L. F. (1986). Impaired verbal and figural fluency after head injury. *Archives of Clinical Neuropsychology, 1*, 87–102.

Ruff, R. M., Light, R. H., & Evans, R. (1987). The Ruff Figural Fluency Test: A normative study with adult. *Developmental Neuropsychology, 3*, 37–51.

Ruff, R. M., Light, R. H., Parker, S. B., & Levin, H. S. (1996). Benton Controlled Oral Word Association Test: Reliability and updated norms. *Archives of Clinical Neuropsychology, 11*, 329–338.

Ryan, J. J., & Paolo, A. M. (1992). Base rates of intersubtest scatter in the old age standardization sample of the WAIS-R. *Archives of Clinical Neuropsychology, 7*, 515–522.

Ryan, J. J., Paolo, A. M., & Brungardt, T. M. (1990). Standardization of the Wechsler Adult Intelligence Scale-Revised for persons 75 years and older. *Psychological Assessment, 2*, 404–411.

Salthouse, T. A. (1988). Effects of aging on verbal abilities: Examination of the psychometric literature. In L. L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 17–35). Cambridge: Cambridge University Press.

Savage, R. D., Britton, P. G., Bolton, N., & Hall, E. H. (1973). *Intellectual functioning in the aged*. London: Methuen.

Schlosser, D., & Ivison, D. (1989). Assessing memory deterioration with the Wechsler Memory Scale, the National Adult Reading Test, and the Schonell Graded Word Reading Test. *Journal of Clinical and Experimental Neuropsychology*, *11*, 785–792.

Schneider, L. S. (1996). An overview of rating scales used in dementia research. *Alzheimer Insights*, *2*(3), 1–7.

Schonfield, A. E. D. (1980). Learning, memory and aging. In J. E. Birren & R. B. Sloane (Eds.), *Handbook of mental health and aging* (pp. 214–244). Englewood Cliffs, NJ: Prentice-Hall.

Shallice, T., & Burgess, P. W. (1991). Deficits in strategy application following frontal lobe damage in man. *Brain*, *114*, 727–741.

Shapiro, M. B. (1973). Intensive assessment of the single case. In P. E. Mittler (Ed.), *The psychological assessment of mental and physical handicaps* (pp. 645–666). London: Tavistock.

Sharpe, K., & O'Carroll, R. (1991). Estimating premorbid intellectual level in dementia using the National Adult Reading Test: A Canadian study. *British Journal of Clinical Psychology*, *30*, 381–384.

Shimamura, A. P. (1994). Neuropsychological perspectives on memory and cognitive decline in normal human aging. *The Neurosciences*, *6*, 387–394.

Silverstein, A. B. (1982). Pattern analysis as simultaneous statistical inference. *Journal of Consulting and Clinical Psychology*, *50*, 234–240.

Silverstein, A. B. (1984). Pattern analysis: The question of abnormality. *Journal of Consulting and Clinical Psychology*, *52*, 936–939.

Snow, W. G., Tierney, M. C., Zorzitto, M. L., Fisher, R. H., & Reid, D. W. (1989). WAIS-R test-retest reliability in a normal elderly sample. *Journal of Clinical and Experimental Neuropsychology*, *11*, 423–428.

Spreen, O., & Strauss, E. (1991). *A compendium of neuropsychological tests. Administration, norms, and commentary*. New York: Oxford University Press.

Squire, L. R. (1987). *Memory and brain*. Oxford: Oxford University Press.

Stebbins, G. T., Wilson, R. S., Gilley, D. W., Bernard, B. A., & Fox, J. H. (1988). Use of the National Adult Reading Test to estimate premorbid IQ in dementia. *The Clinical Neuropsychologist*, *4*, 18–24.

Street, R. F. (1931). *A Gestalt Completion Test*. New York: Bureau of Publications, Teachers College, Columbia University.

Sweet, J. J., Newman, P., & Bell, B. (1992). Significance of depression in clinical neuropsychological assessment. *Clinical Psychology Review*, *12*, 21–45.

Tombaugh, T. N., & McIntyre, N. J. (1992). The Mini Mental State Examination. A comprehensive review. *Journal of American Geriatrics Society*, *40*(9), 922–935.

Trenerry, M. R., Crosson, B., DeBoe, J., & Leber, W. R. (1989). *Stroop Neuropsychological Screening Test Manual*. Odessa, FL: Psychological Assessment Resources Inc.

Tuokko, H., & Woodward, T. S. (1996). Development and validation of a demographic correction system for neuropsychological measures used in the Canadian Study of Health and Aging. *Journal of Clinical and Experimental Neuropsychology*, *18*, 479–616.

Uhlmann, R. F., & Larson, E. B. (1991). Effect of education on the Mini-Mental State Examination as a screening test for dementia. *Journal of American Geriatric Society*, *39*, 876–880.

Van Gorp, W. G., Satz, P., Kiersch, M. E., & Henry, R. (1986). Normative data on the Boston Naming Test for a group of normal older adults. *Journal of Clinical and Experimental Neuropsychology*, *8*, 702–705.

Villardita, C., Cultrera, S., Cupone, V., & Mejia, R. (1985). Neuropsychological test performances and normal aging. *Archives of Gerontology and Geriatrics*, *4*, 311–319.

Walsh, K. W. (1991). *Understanding brain damage: A primer of neuropsychological evaluation*. (2nd ed.). Edinburgh, Scotland: Churchill Livingstone.

Warrington, E. K. (1984). *Recognition memory test*. Windsor: Nelson.

Warrington, E. K., & Taylor, A. M. (1973). Contribution of the right parietal lobe to object recognition. *Cortex*, *9*, 152–164.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised*. New York: The Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale-Revised*. New York: The Psychological Corporation.

Weingartner, H. (1984). Psychobiological determinants of memory failures. In L. R. Squire & N. M. Butters (Eds.), *Neuropsychology of Memory* (pp. 203–212). New York: Guildford Press.

Weingartner, H. (1986). Automatic and effort-demanding cognitive processes in depression. In L. W. Poon (Ed.), *Handbook for clinical memory assessment of older adults* (pp. 218–225). Washington, DC: American Psychological Association.

Weingartner, H., Cohen, R. M., Bunney, W. E., Ebert, M. H., & Kaye, W. (1982). Memory-learning impairments in progressive dementia and depression. *American Journal of Psychiatry*, *139*, 135–136.

Weingartner, H., Cohen, R. M., Murphy, D. L., Martello, J., & Gerdt, C. (1981). Cognitive processes in depression. *Archives of General Psychiatry*, *38*, 42–47.

Welsh, K. A., Butters, N., Hughes, J. P., Mohs, R. C., & Heyman, A. (1992). Detection and staging of dementia in Alzheimer's disease. Use of the neuropsychological measures developed for the consortium to establish a registry for Alzheimer's disease. *Archives of Neurology*, *49*, 448–452.

Wilson, B. A., Alderman, N., Burgess, P., Emslie, H., & Evans, J. J. (1996). *Behavioural Assessment of the Dysexcutive Syndrome*. Bury St Edmunds, England: Thames Valley Test Company.

Wilson, B. A., Cockburn, J., & Halligan, P. (1987). *Behavioural Inattention Test*. Bury St. Edmunds, UK: Thames Valley Test Company.

Wilson, B. A., Cockburn, J. M., Baddeley, A. D., & Hiorns, R. (1989). The development and validation of a test battery for detecting and monitoring everyday memory problems. *Journal of Clinical and Experimental Neuropsychology*, *11*, 855–870.

Wilson, B. A., Evans, J. J., Emslie, H., Alderman, N., & Burgess, P. (in press). The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation*.

Wilson, R. S., Rosenbaum, G., Brown, G., Rourke, D., Whitman, D., & Grisell, J. (1978). An index of premorbid intelligence. *Journal of Consulting and Clinical Psychology*, *46*, 1554–1555.

Woods, R. T., & Britton, P. G. (1985). *Clinical psychology with the elderly*. London: Croom Helm.

Zachary, R. A., & Gorsuch, R. L. (1985). Continuous norming: Implications for the WAIS-R. *Journal of Clinical Psychology*, *41*, 86–94.