18 Observational Methods, Rating Scales, and Inventories

JILL S. FISCHER, H. JULIA HANNAY, DAVID W. LORING, AND MURIEL D. LEZAK

The techniques presented in this chapter tend to be relatively brief. Most are based on observations. Many are not rigorously standardized. Among them are formalized mental status examinations (MSE), elaborations of components of the MSE for identified patient groups or specific diagnostic or treatment questions, screening tests, and schedules for directing and organizing behavioral observations and diagnostic interviews. Some have evolved out of clinical experience, and others were developed for specific assessment purposes. They all provide behavioral descriptions that can amplify or humanize test data and may be useful in following a patient's course or forming gross diagnostic impressions.

THE MENTAL STATUS EXAMINATION

The MSE, a semistructured interview, usually takes place during the examiner's initial session with the patient. It is the only formal procedure for assessing cognitive functions in psychiatric or neurologic examinations. Psychologists often dispense with it since most of the data obtained in the mental status examination are acquired in the course of a thorough neuropsychological evaluation. However, by beginning the examination with the brief review of cognitive and social behavior afforded by the mental status examination, the psychologist may be alerted to problem areas that will need much more detailed study. The MSE will usually indicate whether the patient's general level of functioning is too low for standard adult assessment techniques. It is also likely to draw out personal idiosyncrasies or emotional problems that may interfere with the examination or require special attention or procedural changes. The MSE, whether given as a semistructured interview or as a structured examination using one of the many standardized MSE formats, may be the chief source of data on which determination of a patient's competency for self-care or of legal issues is made (M.P. Alexander, 1988; M. Freedman, Stuss, and Gordon, 1991; S.Y. Kim et al., 2002). However, formalized methods are rapidly evolving to evaluate patients' competency for self-care, management of personal finances (H.R. Griffith et al., 2003; Marson, Sawrie, et al., 2000), or decision making regarding medical treatment (Dymek et al., 2001; Karlawish et al., 2002; Saks et al., 2002).

Mental status information comes from both direct questioning and careful observation of the patient during the course of the interview. Almost every clinical textbook or manual in psychiatry and neurology contains a model mental status examination. Examples of a variety of questions that touch upon many different areas of cognitive and social/emotional functioning and guidelines for reviewing the areas covered by the mental status examination are given in Cummings and Mega (2003), Ovsiew (2002), and Strub and Black (2002, 2003). Different authors organize the components of the mental status examination in different ways and different examiners ask some of the questions differently, but the examination always covers the following aspects of the patient's behavior.

1. *Appearance.* The examiner notes the patient's dress, grooming, carriage, facial expressions and eye contact, mannerisms, and any unusual movements.
2. *Orientation.* This concerns patients' appreciation of time, place, person, and their present situation. Some examiners also inquire about patients' awareness of the examiner's role.
3. *Speech.* Observations are made of both delivery and content of speech. The examiner looks for deviations from normal rate, tone quality, articulation, phrasing, and smoothness and ease of delivery as well as for misuse or confusion of words, grammatical and syntactical errors, perseverations, dysnomia, and other defects in word production and organization.
4. *Thought process.* In patients with aphasic disorders or verbal dyspraxias, and in some with severe func-

tional disturbances such as profound depression with motor slowing, it can be difficult to distinguish speech and thought disorders. In most patients, speech can be evaluated separately from such characteristics of thinking as mental confusion, quality and appropriateness of associations, logic, clarity, coherence, rate of thought production, and such specific thinking problems as blocking, confabulation, circumstantiality, or rationalization.

5. *Attention, concentration, and memory.* In this review of attention span, and of immediate, recent, and remote memory, the examiner inquires about the patient's early and recent history, asking for names, dates, places, and events. Digits forward and reversed, serial subtraction, recall of three or four words immediately and again after an intervening task or five more minutes of interview are typically included in the examination of concentration and memory. Visual memory can be examined by hiding objects or with brief drawing tests (e.g., see Petersen, 1991).

6. *Cognitive functioning.* Estimation of the level of general mental ability is based on quality of vocabulary, reasoning, judgment, and organization of thought as well as answers to questions about topics of general information, fairly simple arithmetic problems, and abstract reasoning tasks. Usually the patient is asked to explain one or two proverbs and to give "similarities" and "differences." When examining patients with known or suspected neurological impairment, the examiner should include simple drawing and copying tasks (e.g., draw a clock and a house, copy a cube or geometric design drawn by the examiner) and a brief assessment of reading and writing.

7. *Emotional state. Mood* (the patient's prevailing emotional tone) and *affect* (the range and appropriateness of the patient's emotional response) need to be distinguished and reported. Mood constitutes the "ground," affect the "figure" of emotional behavior.

8. *Special preoccupations and experiences.* The examiner looks for reports or expressions of bodily concerns, distortions of self-concept, obsessional tendencies, phobias, paranoidal ideation, remorse or suicidal thoughts, delusions, hallucinations, and strange experiences such as dissociation, fugue states, and feelings of impersonalization or unreality.

9. *Insight and judgment.* Questions concerning patients' self-understanding, appreciation of their condition, and their expectations of themselves and for their future elicit information regarding insight. Judgment requires realistic insight. Beyond that, practical judgment can be examined with questions about patients' plans, finances, health needs, and pertinent legal issues (e.g., see Feher, Doody, et al., 1989).

The mental status examination of a reasonably cooperative, verbally intact patient takes 20 to 30 minutes. The examiner's experience and training provide the standards for evaluating much of the patient's responses and behavior, for outside of questions drawn from standardized tests there are no quantitative norms. Thus, the data obtained in the MSE are impressionistic and tend to be coarse-grained, compared with the fine scaling of psychometric tests. It does not substitute for formal testing; rather, it adds another dimension. For many seriously impaired patients, particularly those who are bedridden, who have significant sensory or motor deficits, or whose level of consciousness is depressed or fluctuating, the mental status examination may be not only the examination of choice but also the only examination that can be made of their neuropsychological condition. For example, for severely injured head trauma victims, the mental status examination is often the best tool for following the course during the first six to eight weeks after return of consciousness.

Many of the mental status items can be integrated into an introductory interview covering the patient's history, present situation, and future plans. For example, patients' knowledge about their present income—where it comes from, how much they get from what sources, and their most recent living arrangements—reflects the integrity of recent memory. Patients must make calculations and thus demonstrate how well they can concentrate and perform mental tracking operations if asked to tell the amount of their total income when it comes from several sources, their annual rent or house payments based on the monthly cost, or the amount of monthly income left after housing is paid. Some patients who are concerned about being "crazy" or "dumb" are very touchy about responding to the formal arithmetic questions or memory tests of the MSE. These same patients often remain cooperative if they do not perceive the questions as challenging their mental competence.

RATING SCALES AND INVENTORIES

The content of most scales, inventories, and other patient rating schemes falls into one of three categories: (1) more or less complete mental status examinations that have been given scoring systems; (2) observations by a trained person of some specified class of behavior (e.g., activities, psychiatric symptoms); and (3) observations or reactions of nonprofessional persons familiar with the patient, usually family members. Most of these instruments have been devised with a particular population or diagnostic question in mind and therefore have become associated with that population or question. Moreover, the problems that some of these

scales measure are unique to the population for which they were developed. Therefore, scales and inventories are grouped for review here according to the purpose for which they were originally dedicated.

Rating scales and inventories—particularly ones that were developed early on—typically include scoring schemes that, as likely as not, were devised without benefit of psychometric scaling techniques or substantial reliability or cross-validational studies. Most of the behavioral characteristics that are scored in these instruments tend to separate members of the target population from the population at large at sufficiently respectable rates to warrant their use for gross clinical screening or documentation in research. For clinical purposes, the value of a scale or inventory is more likely to be in the framework it gives to the conduct and evaluation of a brief examination than in its scores.

DEMENTIA EVALUATION

The often very difficult problem of differentiating elderly patients with cognitive or behavioral disturbances due to a progressive dementing disease from those with other neurologic conditions or a psychiatric disorder has inspired many clinicians to systematize the observational schemes that seem to work for them. Most of these instruments were developed to aid in making these difficult discriminations. Thus some contain questions that are best suited for middle-aged and older people or include simplified forms of tasks used in examinations for the general population. Most of them have general applicability including competency evaluations.

Without exception, scales and inventories designed to screen for dementia contain orientation items as these test functions that are sensitive to the most common dementing processes, such as both recent and remote memory, mental clarity, and some aspects of attention. Other areas of common interest are fund of knowledge and language skills. Only the longest scales examine most of the relevant functions; none examines them all. Diagnostic accuracy may be enhanced by combining data from several of these instruments (Eisdorfer and Cohen, 1980; Whelihan, Lesher, et al., 1984).

Thirteen scales for the evaluation of "organic mental status" were briefly described by Kochansky in 1979. Since then, many more have been described in the literature (see Lorentz et al. [2002] and Ruchinskas and Curyto [2003] for reviews). A number of these cognitive screening instruments consist solely of mental status type questions asked of the patient; a few combine such questions with observational ratings. Other scales depend solely upon examiner observations or observer reports. Some scales have had such limited use that they are not in the general assessment repertoire. Only scales in relatively common use are reviewed here. Our focus is on instruments that primarily assess cognitive function, although a few noteworthy measures that assess either the impact of cognitive deficits on daily functioning, or affective and behavioral disturbances associated with common neurologic disorders, will also be featured. For a review of measures—some discussed here—used in Alzheimer's disease drug trials, see Demers et al. (2000a,b). Readers can find rating scales of neurologic function per se in Herndon (1997).

Mental Status Scales for Dementia Screening and Rating

Addenbrooke's Cognitive Examination (ACE) (Mathuranath, Nestor, and Berrios, 2000)

This screening examination is essentially an elaboration of the Mini-Mental Status Examination (MMSE) that was designed to be more sensitive to amnestic syndromes and to isolated frontal or linguistic deficits than most mental status examinations, yet not as complex to administer as the Mattis Dementia Rating Scale or the cognitive section (CAMCOG) of the Cambridge Examination for Mental Disorders of the Elderly. In fact, the ACE bears many similarities to the *Quantitative Mental Status Examination* described by Mahler, Davis, and Benson (1989), which was never widely disseminated.

The ACE consists of six sections—*orientation* (10 items from the MMSE); *attention/mental tracking* (8 points: repetition of three words—*lemon, key,* and *ball*—and five serial seven subtractions); *episodic and semantic memory* (35 points: recall of three words after distraction, learning of a seven-element name and address over three trials, recall of the name and address after a 5-minute delay, and giving the names of four government figures); *verbal fluency* (up to 7 points each for phonemic ["P" words] and semantic ["animals"] fluency); *language* (28 points: naming items depicted in 12 line drawings; comprehension of three simple commands [two spoken and one written], two complex commands, and a three-step command; repeating three words and two phrases; reading two five-item lists composed of either regular and irregular words [each scored all or none]; and writing a sentence); and *visuospatial ability* (5 points: copying intersecting pentagons, copying a cube, and drawing a clock face with numbers and hands set to ten past five). Some of the recent and remote memory items would be more familiar to British citizens than to others, but this instrument could easily be adapted for use elsewhere. The authors estimate

that the ACE takes about 15 to 20 minutes to administer. Scores range from 0 to 100, and an MMSE score can also be calculated.

Test characteristics. The psychometric properties of the ACE were evaluated in a sample of 139 memory clinic attenders (69 with Alzheimer's dementia, 29 with frontotemporal dementia, 14 with vascular dementia, and 27 with other degenerative neurologic disorders) and 127 age- and education-matched patients with nonneurologic illnesses or patient family members (Mathuranath et al., 2000). The ACE had very good internal consistency reliability in this sample (Cronbach's *alpha* = .78). Patients with various forms of dementia earned a mean ACE composite score of 64.8 ± 18.9; the control group's mean score was 93.8 ± 3.5.

Neuropsychological findings. Two ACE cut-off scores were derived. The first (88) was selected because it is two standard deviations below the control group mean. With excellent sensitivity (93%) but only modest specificity (71%), it might be used most appropriately in clinical settings when one wants to avoid overlooking potential dementia cases. Applying this criterion, the ACE identified an impressive 98% of patients with very mild dementia (Clinical Dementia Ratings [CDRs] = 1.0) and 100% of those with moderate to severe dementia. The second cut-off (83) was determined by estimating the probability of diagnosing dementia in the 139 clinic patients; it optimizes sensitivity (82%) and specificity (96%) across a range of dementia prevalence rates and therefore might be useful in identifying subjects for research studies.

Not only was the ACE sensitive to very mild dementia, but it also proved useful in differentiating patients with Alzheimer's dementia from those with frontotemporal dementia based on their patterns of performance on ACE subtests. Mathuranath and colleagues (2000) suggest that calculation of a VL/OM ratio, consisting of the sum of points on the verbal fluency plus language subtests to the sum of points earned on the orientation plus memory tests. Based on the mix of cases in their sample, a VL/OM of >3.2 best differentiated Alzheimer patients from those with other dementias, and a VL/OM of <2.2 as more likely to identify frontotemporal as opposed to other forms of dementia.

Cognitive Capacity Screening Examination (CCSE) (J.W. Jacobs et al., 1977)

This 30-item scale was devised to identify medical patients with brain disorders. In contrast to other brief mental status examinations, items involving attention, mental tracking, and working memory play a prominent role in the CCSE. Consequently, although much less widely used than the Mini-Mental State Examination, this scale is less prone to ceiling effects in higher functioning patients (Hershey et al., 1987). Five CCSE items cover orientation questions (four pertaining to time); 11 involve simple attention (two items) and mental tracking (digits and days of the week reversed and serial subtractions of 7 from 100); three are easy arithmetic problems (e.g., "9 + 3 is __"); six are memory items (two of very short-term recall following a Peterson-Brown type of distraction—scored all or none—and four requiring recall of words after five intervening test items—each word scored individually); these intervening items are five very easy differences (e.g., "The opposite of **up** is __") or similarities (e.g., "**Red** and **blue** are both __").

Test characteristics. A study of a large heterogeneous sample of male veterans referred for psychological consultation or substance abuse treatment found 11 factors for the CCSE, of which three—digit span with interference, complex mental arithmetic, and verbal memory—accounted for the lion's share of the variance in total test scores (D.A. Anderson et al., 2001). The test–retest reproducibility of the CCSE is ±2 points in healthy subjects (J.S. Meyer, Li, and Thornby, 2001).

Based on the scores obtained by samples of medical patients referred for psychiatric consultation, psychiatric inpatients, a consecutive series of medical patients, and 25 hospital staff members, the authors defined a cut-off score of 20, interpreting scores of 19 or lower as indicating cognitive dysfunction. Using this cut-off, from 16% of a psychiatric sample (Beresford et al., 1985) to 53% of neurosurgery patients (Schwamm et al., 1987) had scores in the impaired range, with neurological patients (Hershey et al., 1987; D.M. Kaufman et al., 1979) falling in between.

False positive findings tend to be relatively infrequent and are most likely to occur in patients with hearing or language comprehension deficits associated with focal lesions, relatively mild or circumscribed cognitive deficits, developmental disabilities, or limited education. By raising the cut-off score to 25 and 27 for subjects age ≥50 and <50, respectively, Heaton, Thompson, Nelson, and their coworkers (1990) obtained a false positive rate of 15% in samples of multiple sclerosis patients and normal control subjects. The mean scores for these two groups differed by just 1 point (27.1–28.1), yet this difference was significant ($p <$ .02). False negative results are more common and more likely to occur in patients who have focal lesions or relatively mild or circumscribed cognitive deficits.

Neuropsychological findings. In a sample of patients with migraine or cluster headaches (about two-thirds of whom were less than 50 years old), J.S. Meyer and colleagues set the CCSE cut-off at 27 and obtained 83% sensitivity in detecting cognitive decline (defined as a sustained decrease of >3 points) during headache intervals and 92% specificity for cognitively normal headache-free periods—much greater than the MMSE, with a sensitivity of 49%. Moreover, data from a longitudinal study of patients with memory complaints and a family history of stroke or dementia suggest that a cut-off of 26 works reasonably well in identifying patients who develop dementia in any form over a 3-year period (88% sensitivity and 83.5% specificity)—again considerably better than the MMSE, which has a sensitivity of only 57.1% in identifying these patients.

Cambridge Cognitive Examination (CAMCOG)[1] (Huppert, Brayne, et al., 1995)

This mental status examination is the objective test portion of an instrument developed for the early diagnosis and monitoring of dementia in the elderly, the *Cambridge Mental Disorders of the Elderly Examination-Revised* (*CAMDEX*) (M. Roth, Tym, et al., 1999; M. Roth, Huppert, et al., 1999). The other two portions of the CAMDEX comprise structured interviews with the patient and—separately—with an informant regarding the patient's current psychiatric status, medical history, and family history. Although used primarily in England and Europe, an early study demonstrated that the CAMDEX can be used as effectively in the U.S. (Hendrie et al., 1988). It has been translated into other languages such as Hebrew (Heinik, Werner, et al., 1999).

The CAMCOG's 67 items are grouped into eight subscales: *Orientation* (ten items dealing with time and place); *Language* (seven comprehension items, six naming items, category fluency ["animals"], and four word definitions); *Memory* (recall and recognition of six pictured objects, name and address recall, and ten WIS-A Information type items [e.g., "When did World War I start?"]); *Attention* (counting from 20 to 1 and serial sevens [five subtractions]); *Praxis* (copying geometric figures and following commands); *Calculation; Abstract thinking* (similarities between pairs of items); and *Perception* (e.g., recognition of objects depicted from unusual angles and stereognosis). Eight items do not contribute to the total score but are included to permit calculation of an MMSE total score (five items) or to acquire additional qualitative information (three items). The CAMCOG also incorporates Hodkinson's (1972) modification of the Blessed Dementia Rating Scale. The full CAMCOG takes about 25 minutes and yields a maximum score of 107.

Test characteristics. Unfortunately, no age- and education-stratified norms are currently available. Some investigators have recommended using regression-based formulas to predict CAMCOG scores—with age, social class, marital status, education (or estimated premorbid intellectual level), and "general knowledge" (i.e., performance on 10 WIS Information-type items from the CAMCOG) as predictors—and defining impairment as a predetermined degree of discrepancy between actual and predicted CAMCOG scores (e.g., K. Andersen et al., 1999). Like many other mental status examinations, the CAMCOG is influenced by age and education and, to a lesser extent, sex (Huppert, Brayne, et al., 1995). Of these, age exerts the broadest effects, influencing the total score and all subscale scores—excepting Attention—whereas education principally affects performance on Language and Abstract Thinking. Thus classification errors are inevitable if a single cut-point is used without regard for a patient's age and education (Huppert et al., 1995; Lindeboom et al., 1993).

The CAMCOG correlates strongly with the MMSE—.87 in one study (Blessed, Black, et al., 1991)—as expected, given that all of the MMSE items are embedded in the CAMCOG. Unlike the MMSE though, CAMCOG total scores distribute across a wide range for patients with dementia (Huppert, Brayne, et al., 1995), Parkinson's disease (Hobson and Meara, 1999), and stroke (de Koning, van Kooten, et al., 1998). Interrater reliability on the CAMCOG is high, with an intraclass correlation coefficient of .87 for 10 examiners in one large-scale Danish study (K. Andersen et al., 1999). Test–retest reliability of the CAMCOG is also high—.97 in a sample composed of 53 Alzheimer patients and healthy elderly controls (Lindeboom et al., 1993).

Neuropsychological findings. A cut-off score of 80 was originally recommended when screening for dementia; this cut-off yielded a sensitivity of .92 and a specificity of .96 in a heterogeneous sample of inpatients and outpatients in a geriatric medicine and psychogeriatrics department (M. Roth, Tym, et al., 1986). This cut-off also did quite well in identifying Parkinson patients with dementia, with a sensitivity of .95 and a specificity of .94 in a sample in which close to half of the subjects met the *Diagnostic and Statistical Manual-IV* (DSM-IV) criteria for dementia (Hobson and Meara, 1999). Defining impairment as a total score at least 1.25 standard errors below the predicted score

[1]CAMCOG and CAMDEX are sold by Cambridge Cognition, Tunbridge Ct., Tunbridge Lane, Bottisham, Cambridge CB5 9DU, UK. (e-mail: info@camcog.com)

or a CAMCOG below 74, since none of the nondemented individuals in their pilot study of community-dwelling elderly had scored lower than this, yielded optimal sensitivity (.89) and specificity (.88) (K. Andersen et al., 1999).

Four CAMCOG composite variables—category fluency, memory, general knowledge, and attention—combined with age predicted which subjects were likely to meet criteria for dementia two years later (Nielsen, Lolk, Andersen, et al., 1999). Similarly, other investigators have found that relatively poorer performance on CAMCOG memory items than on nonmemory items predicts who would become demented over the subsequent three years (Schmand, Walstra, Lindeboom, et al., 2000). Among stroke patients, three variables heightened the risk of meeting criteria for dementia three months after a stroke: poorer CAMCOG scores, a right hemisphere stroke, and a hemorrhagic stroke (de Koning et al., 1998).

Other CAMCOG formats. A revised version (CAMCOG-R) adds ideational fluency items ("How many different uses can you think of for a bottle?") and a matrix reasoning test similar to Raven's Progressive Matrices or the WAIS-III Matrices (M. Roth, Huppert, et al., 1999). These can be summed to give an Executive Function score, with a maximum of 28. However, at least in a stroke population, these tests were strongly correlated with the tests of executive function included in the original CAMCOG (category fluency and abstract reasoning), raising questions about the necessity of including them in a screening examination (Leeds et al., 2001). Also included in the revised version are remote memory alternative questions from the 1950s and 1960s to assess more recently born cohorts.

De Koning, Dippel, and their colleagues (2000) developed a 25-item short form to use in screening for poststroke dementia by removing items subject to floor and ceiling effects, removing subscales that did not improve diagnostic accuracy, and eliminating items that diminished subscale homogeneity. With subscales for orientation, memory, perception, and abstraction, this version performs with comparable diagnostic accuracy but takes only about 10 minutes to administer. However promising this shortened instrument might be, it still must be cross-validated in other samples of stroke patients and—if it is to be used in other populations—these patients as well.

Dementia Rating Scale (DRS) (Mattis, 1976, 1988)

This widely used dementia screening scale is also known by the author's name: *Mattis Dementia Rating Scale* (*MDRS*). The MDRS examines five areas that are particularly sensitive to the behavioral changes that characterize senile dementia of the Alzheimer type. Five areas are covered: (I) *Attention* (37 possible points): digits forward and backward up to four; follow two successive commands (e.g., "Open your mouth and close your eyes"); (II) *Initiation and Perseveration* (37 points): name items in a supermarket; repeat series of one-syllable rhymes; imitate double alternating hand movements; copy a row of alternating O's and X's; (III) *Construction* (6 points): copy a diamond in a square; copy a set of parallel lines; write name; (IV) *Conceptual:* four WIS-A type Similarities items; identify which of three items is different; sentence generation; and (V) *Memory:* delayed recall of a five-word sentence; personal orientation; word recognition memory; design recall. A scoring system permits test–retest comparisons of both individual subscales and a total score.

An interesting feature of this scale is that, instead of giving items in the usual ascending order of difficulty, the most difficult item is given first (digit span items excepted). Since the most difficult items on the Dementia Rating Scale are within the capacity of most intact older persons, this feature can be a time-saver. An intact subject would only have to give three abstract answers on the first subtest (Similarities) of the Conceptualization section, for example; the other 26 items in this section would be skipped. Administration with an intact subject can take as little as 20 minutes, whereas with demented patients it is likely to require 30–45 minutes. Comparability between subscales is limited by their differences in the number of items and potential score points.

Test characteristics. Like scores on other mental status instruments, MDRS scores are negatively correlated with age and positively correlated with education (A.L. Bank et al., 2000; Lucas, Ivnik, Smith, et al., 1998b; G. Smith, Ivnik, Malec, and Kokmen, 1994); Vangel and Lichtenberg, 1995). Age effects are most striking in patients with moderately severe dementia (Vitaliano, Breen, Russo, et al., 1984). Appropriate interpretation of individual patients' test scores has been greatly facilitated by the publication of age- and education-stratified normative databases for well-educated healthy older Caucasian adults (Lucas et al., 1998b; Monsch, Bondi, Salmon, et al., 1995), community dwelling older adults with a range of educational backgrounds (R. Schmidt, Freidl, et al., 1994), rural community dwelling older adults with limited education (Marcopulos, McLain, and Giuliano, 1997), and less educated urban medical patients (A.L. Bank et al., 2000; see also Spreen and Strauss, 1998).

Sex (A.L. Bank et al., 2000; Vangel and Lichtenberg, 1995) and race (Woodard, Auchus, et al., 1998) affect

MDRS performance to a much lesser extent than age and education. However, interesting cultural differences have emerged in studies using translated versions of the MDRS. For example, Hispanic Alzheimer patients performed significantly worse than their nonHispanic counterparts on the total MDRS and especially the Conceptualization and Memory subscales (Hohl et al., 1999). As a group, elderly adults from Hong Kong did better than age- and education-matched persons in San Diego on the Construction subscale, whereas those in San Diego had more success on the Initiation-Perseveration and Memory subscales (Chan, Choi, et al., 2001). It is prudent for the examiner to be sensitive to cultural factors that may affect MDRS performance.

For normal older adults, MDRS scores remain reasonably stable over one to two years, although any given individual's scores may fluctuate as much as one standard deviation during this period (G. Smith, Ivnik, Malec, and Kokmen, 1994). Smith and colleagues suggest that MDRS total score declines of more than 10 points are rare, occurring in fewer than 5% of healthy older adults, and should be suspect.

The reliability of the MDRS has been extensively investigated. Test–retest reliability is excellent (.97 for the total score) (Mattis, 1988). Early studies with small patient samples reported a split-half reliability of .90 (R. Gardner et al., 1981) and coefficient *alphas* for individual subscales ranging from .95 (Attention and Conceptualization) to .75 (Memory) (Vitaliano, Breen, Russo, et al., 1984). When a larger, more heterogeneous patient sample (i.e., Alzheimer's disease, vascular dementia, or mild cognitive impairment) was examined, the internal consistency of the MDRS was somewhat lower: coefficient *alpha* for the MDRS total score was .82, while those for most subscales fell into the .75 to .84 range (G. Smith, Ivnik, Malec, and Kokmen, 1994). Initiation-Perseveration was considerably less cohesive (coefficient *alpha* = .44), which should not be surprising given the varied items on this subscale, but raises questions about its interpretability as an individual subscale.

The construct validity of the Attention, Conceptualization, and Memory subscales has been supported in studies of their correlations with Wechsler scale indices (G. Smith, Ivnik, Malec, and Kokmen, 1994) and other tests of similar cognitive functions (Marson, Dymek, et al, 1997). Interpretation of the Construction subscale becomes questionable, however, given that it correlates more strongly with attentional tests than it does with other visuoconstructional measures (Marson, Dymek, et al., 1997; G. Smith et al., 1994). Some authors have pointed out that the MDRS is limited in its assessment of visual construction and suggest that it be supplemented with additional visuoconstructional items when assessing patients who commonly have deficits in this domain, such as Parkinson patients (G.G. Brown, Rahill, et al., 1999).

Factor analyses of the MDRS have yielded varied results. For example, H.R. Kessler, Roth, and their colleagues (1994) found that a two-factor solution gave the best fit for a heterogeneous patient sample with varied neurological and psychiatric diagnoses. Studying Alzheimer patients, Colantonio and colleagues (1993) had derived three factors: memory, construction and conceptualization. Woodard, Salthouse and colleagues (1996) reported that after collapsing Attention with Initiation-Perseveration into a single factor, a modified four-factor version of Mattis' rationally derived subscales provided the best description. And yet a rather different set of five factors has been reported (Hofer et al., 1996). However, this sample was small and included more healthy subjects than dementia patients—which raises questions about factor stability. G. Smith and his colleagues (1994) advised caution in interpreting subscales other than Memory and Conceptualization which, in light of the findings reported here, appears to be wise. A recently revised MDRS manual provides additional reliability and validity data and guidelines for clinical interpretation (Jurica et al., 2002).

Neuropsychological findings. Unlike some other brief mental status examinations, the MDRS total score does well in identifying Alzheimer patients, separating mildly impaired Alzheimer patients from control subjects with perfect accuracy in one study (Prinz, Vitaliano, et al., 1982). Originally, a cut-off score of 137 (out of 144) was proposed as a "red flag" for suspected impairment, but this was based on a small sample consisting of 11 healthy adults and 20 patients with heterogeneous neurological conditions affecting brain function. This cut-off was later revised downward to 123, a score two standard deviations below the mean score of Montgomery and Costa's sample of 85 healthy older adults (cited in Mattis, 1988). This revised cut-off performed reasonably well (83% sensitivity; 100% specificity) in a sample of 41 patients with Alzheimer-type or vascular dementia and 22 healthy controls (van Gorp, Marcotte, et al. 1999).

The danger of using cut-off scores derived from demographically dissimilar samples was amply illustrated in a study which found that close to half of a sample of older rural community-dwelling adults were misclassified as impaired when the conventional cut-off of 123 was used (Marcopulos, McLain, and Giuliano, 1997). Vangel and Lichtenberg (1995) reported that a cut-off score of 125 produced acceptable sensitivity (.85) and specificity (.90) in a sample of urban elderly medical patients, but higher cut-off scores may be ap-

propriate when well-educated patients are evaluated. For example applying a cut-off score of 129 to a sample with a 2:1 ratio of Alzheimer patients to healthy controls yielded optimal sensitivity (.98) and specificity (.97) in a highly educated, predominantly Caucasian sample (Monsch, Bondi, Salmon, et al., 1995). This cut-off score also performed well when cross-validated in a separate community-dwelling sample of older adults, about 15% of whom had Alzheimer's disease (91% of patients and 93% of healthy controls correctly classified). An even higher cut-off (133) was necessary to achieve optimal sensitivity (.96) and specificity (.92) in separating another highly educated sample of mildly impaired Alzheimer patients (those with "intact" MMSEs of 24 or higher) from matched healthy persons (Salmon, Thomas, et al., 2002).

MDRS total scores have been used to stage dementia patients in terms of level of impairment (Salmon, Thal, et al., 1990; Shay et al., 1991). Different patterns of subscale performance may help distinguish control subjects from mildly impaired Alzheimer patients (Hochberg et al., 1989; Vitaliano, Breen, Russo, et al., 1984) and mildly impaired patients from moderately impaired patients (Hochberg et al., 1989). In one study, three MDRS subscales (Initiation-Perseveration, Construction, and Memory) discriminated significantly between control subjects and mildly impaired patients and between mildly and moderately impaired patients, whereas Attention and Conceptualization discriminated only between mildly and moderately impaired patients (Vitaliano, Breen, Russo, et al., 1984). Hochberg and her colleagues (1989) found that the high degree of sensitivity of the Initiation-Perseveration subscale to Alzheimer's disease severity depended mostly on verbal fluency (articles of clothing), accounting for 78% of the variance in predicting patients' self-care behavior; adding verbal imitation raised the amount of variance accounted for to 92%.

The pattern of MDRS subscale performance may also help differentiate patients with differing neuropathological conditions. In fact, neuroimaging studies have demonstrated differential correlations between specific MDRS subscales (e.g., Memory) and brain regions known to be associated with these functions (e.g., hippocampal volumes) (Fama, Sullivan, Shear, et al., 1997). In an early study, patients with frontal involvement were impaired only on the Initiation-Perseveration subscale, whereas Korsakoff patients did most poorly on the Memory subscale (Janowsky, Shimamura, Kritchevsky, and Squire, 1989). As a group, Alzheimer patients are almost always more impaired on the MDRS Memory subscale than patients with any other dementia etiology. Autopsy studies have shown that Alzheimer patients without evidence of Lewy bodies performed significantly worse in life on the MDRS Memory subscale than did either Alzheimer patients with Lewy body pathology (D.J. Connor et al., 1998) or frontotemporal dementia patients (Rascovsky et al., 2002). In contrast, Alzheimer patients with Lewy body pathology did worse on the Initiation-Perseveration subscale (D.J. Connor et al., 1998).

Alzheimer patients also do worse on the MDRS Memory subscale than patients with so-called subcortical pathologies, such as those with Parkinson's disease (Cahn-Weiner, Grace, et al., 2002; Paolo et al., 1995) or vascular dementia, who typically perform poorly on Construction (Lukatela et al., 2000), or patients with progressive supranuclear palsy (Rosser and Hodges, 1994) or Huntington's disease (Paulsen, Butters, et al., 1995; Rosser and Hodges, 1994; Salmon, Kwo-on-Yuen, et al., 1989) who do worst on the Initiation-Perseveration subscale. These observations fit nicely with a neuroimaging study demonstrating that MDRS Memory performance was most strongly related to whole brain volume, whereas Construction and Initiation-Perseveration subscale performances were more closely linked with subcortical hyperintensities (Paul, Cohen, et al., 2001).

Both total scores and subscale scores on the MDRS were positively related to the ability to perform basic and instrumental activities of daily living—although not behavior problems—in Alzheimer patients (Teri, Borson, et al., 1989; Vitaliano, Breen, Russo, et al., 1984). The MDRS has also been shown to predict rehabilitation outcome (e.g., return to prior living situation) (MacNeill and Lichtenberg, 1997). Total MDRS scores have been used to predict length of survival in Alzheimer patients (G. Smith, Ivnik, Malek, and Kokmen, 1994) and medically ill patients (Arfken et al., 1999).

The Extended Scale for Dementia (ESD). This revision of the DRS divides up the orientation item so that time, place, and age are scored separately and it adds several items: "Information" (e.g., "How many weeks [months] are there in a year?"); "Count Backwards" and "Count by 3's"; "Simple Arithmetic"; a "simple" paired-association learning test; a "simple version" of Block Design taken from the Wechsler Intelligence Scale for Children; and the two graphomotor items of the original test combined, making a total of 23 items (Hersch, 1979). After six weeks, test–retest correlations were .94 for 24 dementia patients.

However, the ESD's sensitivity of .93 in distinguishing dementia patients from normal control subjects in the 65 and older age range dropped to .75 for persons under age 65 (Lau et al., 1988). Age-dependent cut-off scores were applied to maintain the specificity rate at

.96 for both age groups. Over six months, both Alzheimer and vascular dementia patient groups had significant score declines even though the groups were small. Another small study suggested that dementia patients deteriorated at similar rates, regardless of the underlying pathology (Alzheimer's disease, dementia with Lewy bodies, and a combination of the two) (Helmes, Bowler, et al., 2003).

Based on a factor analysis of the responses of 219 outpatients with Alzheimer's disease that yielded three factors (conceptualization, construction, and memory), Colantonio and colleagues (1993) devised an abbreviated 86-item test with a reorganized scoring system. The full scale remains much more widely used however. The MDRS has also been adapted for use with diverse populations, including Spanish-speaking and Chinese adults.

Mini-Mental State (MMS) or Mini-Mental Status Examination (MMSE) (M.F. Folstein, Folstein, and McHugh, 1975)

This formalized mental status examination is probably the most widely used brief screening instrument for dementia whether used either alone or as a component of such examination protocols as the CERAD battery (J.C. Morris, Heyman, et al., 1989). Originally devised to facilitate differential diagnosis of hospitalized psychiatric patients, it is routinely used to assess cognitive abilities in epidemiological studies—both cross-sectional and longitudinal (Crum et al., 1993; Kase, Wolf, et al., 1998). It is also routinely used to select patients for dementia treatment trials (M.J. Knapp et al., 1994; Raskind et al., 2000; S.L. Rogers, Farlow, et al., 1998). The MMSE assesses a restricted set of cognitive functions simply and quickly (see Fig. 18.1). The standardized administration and scoring procedures are easily learned, with administration by a seasoned examiner taking about five to ten minutes. A total of 30 points are possible.

Early factor analyses of the MMSE together with other tests identified three factors, labeled differently but essentially consisting of verbal functions, memory abilities, and construction (Giordani et al., 1990; J.C. Morris, Heyman, et al., 1989). When MMSE item responses of a large sample of older adults were analyzed independently, five distinct though related domains emerged: concentration or working memory (serial 7s and spelling 'world' backwards); language and praxis (naming, following commands, and construction); orientation; memory (delayed recall of three items); and attention span (immediate recall of three items) (R.N. Jones and Gallo, 2000). Very similar factors were derived in an analysis of the MMSE item responses of psychiatric inpatients (Banos and Franklin, 2002), providing empirical support for Folstein's rational groupings of MMSE items (M.F. Folstein, Folstein, and McHugh, 1975).

Test characteristics. MMSE scores are strongly influenced by both age and education, decreasing with age and increasing with education (J.C. Anthony et al., 1982; Tombaugh and McIntyre, 1992). Individuals with less education tend to make errors on the first serial subtraction, spelling "world" backwards, repeating phrases, writing, naming the season, and copying (R.N. Jones and Gallo, 2002). Clinically useful age- and education-stratified norms have been published for the MMSE (Bravo and Hebert, 1997; Crum et al., 1993; Tombaugh, McDowell, et al., 1996). Sex has a negligible impact on overall MMSE scores (Tombaugh and McIntyre, 1992), although differences are evident on a few individual MMSE items (e.g., women are more likely to err on serial 7s, whereas men are more prone to errors on spelling "world" backwards and other language items) (R.N. Jones and Gallo, 2002).

Ethnicity also affects MMSE performance. For example, African Americans and Hispanics are more likely than European Americans to be erroneously identified as demented (J.C. Anthony et al., 1982; Auerbach and Faibish, 1989; Espino et al., 2001; Mulgrew et al., 1999). Ethnic differences—at least in Mexican Americans—appear to be largely a function of acculturation: barrio-residing Mexican Americans score lower than their counterparts who live in transitional or suburban neighborhoods (Espino et al., 2001).

Test–retest reliability over a 24-hour period in the original standardization sample of nondemented psychiatric inpatients was high, whether the examiner was the same both times ($r = .89$) or different ($r = .83$) (M.F. Folstein, Folstein, and McHugh, 1975). Test–retest reliability over a 4-week period was nearly perfect for the dementia patients in Folstein's sample ($r = .99$). (For more test–retest reliability data, see McCaffrey, Duff, and Westervelt, 2000b.)

Neuropsychological findings. By and large, the effectiveness of the MMSE in identifying cognitively compromised patients depends upon the composition of the groups under study (Tombaugh and McIntyre, 1992). In the original MMSE validation study, none of the 63 normal elderly patients scored below 24, which subsequently became the de facto criterion for identifying cognitive impairment. The MMSE is most effective in distinguishing patients with moderate or severe deficits from control subjects (Filley, Davis, et al., 1989; M.F. Folstein, Folstein, and McHugh, 1975). It is less effective in separating mildly demented patients from nor-

Patient _____

Examiner _____

Date _____

MINI MENTAL STATE

Score		
		Orientation
( )		What is the (year) (season) (month) (date) (day)? (5 points)
( )		Where are we? (state) (county) (town) (hospital) (floor) (5 points)
		Registration
( )		Name 3 objects: 1 second to say each. Then ask the patient to repeat all three after you have said them. 1 point for each correct. Then repeat them until he learns them. Count trials and record _____. (3 points)
		Attention and Calculation
( )		Serial 7's. 1 point for each correct. Stop at 5 answers. Or spell "world" backwards. (Number correct equals letters before first mistake - i.e., d l o r w = 2 correct). (5 points)
		Recall
( )		Ask for the objects above. 1 point for each correct. (3 points)
		Language Tests
( )		name - pencil, watch (2 points)
( )		repeat - no ifs, ands or buts (1 point)
( )		follow a 3 stage command: "Take the paper in your right hand, fold it in half, and put it on the floor." (3 points)

FIGURE 18.1 Mini-Mental State. (From Folstein et al., 1975)

(continued)

mal subjects (Galasko et al., 1990; R.G. Knight, 1992), identifying cognitively impaired medical inpatients (J.C. Anthony et al., 1982; Auerbach and Faibish, 1989), or identifying patients with focal or lateralized lesions (Dick et al., 1984; Naugle and Kawczak, 1989; Schwamm et al., 1987).

Applying a cut-off score of 24 without regard for the examinee's age or educational background, or to patients with subtle or focal cognitive deficits, is bound to lead to classfication errors. When the conventional cut-off of 24 was applied to samples of patients referred for dementia evaluations, the MMSE had good specificity but limited sensitivity: .90 and .69 in one study (Feher and Martin, 1992) and .96 and .63 in another (Kukull et al., 1994), respectively. The ideal *screening* test should emphasize sensitivity even if this comes at

Mini Mental State
Page 2

Score

Read and obey the following:

( ) CLOSE YOUR EYES. (1 point)

( ) Write a sentence spontaneously below. (1 point)

( ) Copy design below. (1 point)

( ) TOTAL 30 POINTS

The above test does not include abstraction. You may want to test this for your own information:

Proverbs
Similarities

FIGURE 18.1 Mini-Mental State (continued).

the expense of specificity, in contrast to *diagnostic* tests which should favor specificity over sensitivity. Toward this end, Kukull and colleagues recommended raising the MMSE cut-off to 26 or 27 to increase the MMSE's sensitivity in symptomatic populations.

The MMSE is sensitive to dementia severity (J.C. Morris, Heyman, et al., 1989; Teng, Chui, Schneider, and Metzger, 1987), although individual MMSE items perform differently at earlier and later stages of the illness (Fillenbaum, Wilkinson, et al., 1994). For example, performance on the three-word delayed recall item of the MMSE predicted which community-dwelling older adults would develop Alzheimer's disease in one study (Small et al., 2000) and was the most sensitive item for distinguishing mild to moderate dementia (Galasko et al., 1990; Teng, Chui, Schneider, and Metzger, 1987). Language items—excepting the 3-stage command, which also has mental tracking and sequencing components—had the least sensitivity in the early stages of dementia (Feher, Mahurin, and Doody, 1992). These findings suggest that, in some contexts, a very brief two-item screen—using three-word recall and either orientation to time (the second most sensitive item in the Galasko and Fillenbaum studies) or copying (the second most sensitive item in the Teng study)—might perform as well as the full MMSE. Similarly, another study identified six MMSE items that did nearly as well as the entire scale in identifying patients with dementia (Callahan et al., 2002).

The MMSE performance of healthy older adults is reasonably stable over time, following a slight improvement between the first and second testings as a result of experience with the test (Jacqmin-Gadda et al., 1997). In contrast, the MMSE performance of patients with Alzheimer's disease deteriorates over time—at an average rate of 3.26 points per year (95% confidence interval: 3.06 to 3.46) in one study (R.S. Wilson, Gilley, et al., 2000b). MMSE change is not linear across the range of test scores as it is subject to both ceiling and floor effects (Mungas and Reed, 2000). Moreover, MMSE change is not consistent from one Alzheimer patient to the next (Doody et al., 2001; Mendiondo et al., 2000) though the rate of decline for a specific patient at a given stage of dementia is reasonably predictable (Doody et al., 2001).

MMSE total scores do not differentiate patients with

Alzheimer's disease from patients with other dementias, but some investigators suggest that patterns of performance on individual MMSE items may help distinguish patients with different dementia etiologies. Thus patients with pathologically confirmed dementia with Lewy bodies performed poorly relative to patients with pathologically confirmed Alzheimer's disease on the attention and construction items of the MMSE and did relatively better on the MMSE memory items (Ala et al., 2001). Parkinson patients struggled the most with construction (both mechanics of writing a sentence and copying), while patients with ischemic vascular disease had difficulty with both attention and construction, and Alzheimer patients did worst on temporal orientation and delayed recall (Jefferson et al., 2002). Orientation to date and three-word delayed recall also distinguished Alzheimer patients from Huntington patients with early disease who had relatively greater difficulty with serial sevens (Brandt, Folstein, and Folstein, 1988). However Huntington patients with advanced disease did worse than Alzheimer patients on registration (immediate recall) of three words and writing.

MMSE performance predicts important functional outcomes such as medication adherence (Salas et al., 2001), length of hospital and rehabilitation stay, rehabilitation course and outcome, and risk of death (see reviews by Ruchinskas and Curyto, 2003 and Tombaugh and McIntyre, 1992). MMSE scores have also been used to model the costs of care in Alzheimer's disease, estimated to be approximately $2000 (in 1995 dollars) for each 1-point decrement in MMSE scores in one study (L. Jonsson et al., 1999).

Variants of the Mini-Mental State Examination. Numerous modifications of the MMSE have been proposed, some minor and others more extensive. For example, the Galasko group (1990) and others have observed that spelling "world" backwards and performing serial 7s are not interchangeable tasks, so they suggested replacing both with the "months backward task." Leopold and Borson (1997) proposed retaining the "world" item and having higher functioning individuals not only spell it backwards, but also put its letters in alphabetical order. On finding that the addition of cumulative recall over two delayed recall trials at five minute intervals improved the detection of patients with mild cognitive impairment substantially (sensitivity of 96.2% with a specificity of 90.4%) Loewenstein, Barker, and their colleagues (2000) suggested adding delayed recall trials to the MMSE.

The *Modified Mini-Mental State (3MS)* is the most widely used revision of the MMSE (Teng and Chui, 1987). These authors added four new items (listing four-legged animals and identifying similarities between three pairs of items), modified the administration order and content of other items (e.g., adding cued recall and recognition items to the memory assessment), and developed a more detailed scoring system (e.g., copying pentagons is allotted up to 10 points rather than a simple "pass/fail"). These additions extend the score range to 0–100. The 3MS is slightly more sensitive than the MMSE to cognitive impairment in stroke patients, though not appreciably better in terms of overall classification accuracy when conventional cut-offs are used (i.e., below 79 on the 3MS and below 24 on the MMSE) (Grace et al., 1995). Age and education adjusted norms for the 3MS have been developed (Bravo and Hebert, 1997; Tombaugh, McDowell, et al., 1996). The 3MS itself has been revised (*3MS-R*) with publication of normative data on the 3MS-R for 2913 healthy individuals spanning a broad age range that includes subjects more than 100 years old.

The *Cognitive Abilities Screening Instrument (CASI)* was developed for use in cross-national studies of community-dwelling older adults (Teng, Hasegawa, et al., 1994). Its 25 items come from the MMSE, the 3MS, and the Hasegawa Dementia Rating Scale. The CASI can be administered in 15–20 minutes. It has been translated into Japanese, Chinese, Vietnamese, and Spanish and has been used in a number of international studies. Nine domain subscale scores can be calculated, although the authors caution about the limited range and potential unreliability of most subscale scores. Total scores range from 0 to 100, and an MMSE score can be derived as well. A CASI cut-off score of <86 has both high sensitivity (96.5%) and specificity (92%); a slightly lower cut-off score (<81) optimizes specificity (98.9%) while still maintaining an acceptable sensitivity (82.5%) (Graves, Teng, et al., 1992). Age-stratified norms are available.

The *Severe Mini-Mental State Examination (SMMSE)* was designed to facilitate the testing of more severely impaired patients (Harrell et al., 2000). This 30-point instrument can be used until patients become mute or have no functional language. It consists of personal information (giving one's first name, last name, and complete birth date); three-word repetition; two single-step commands; three naming items; two construction items (drawing a circle to command, copying a square); writing one's name; category fluency ("animals"); and spelling "cat". Interrater reliability was exceptionally high (.99). Test–retest reliability over a 5-month period was quite good (.80), considering the long interval with deteriorating patients. The SMMSE appears to be particularly useful in assessing patients whose MMSE scores are below 10 or who are considered to have at least "moderately severe" dementia (Global Deterioration Scale of 5 or higher, or Clinical

Dementia Rating of 4 or higher). It tests a restricted set of cognitive functions simply and quickly.

The 7-Minute Screen (7MS) (P.R. Solomon, Hirschoff, et al., 1998)

The 7MS was designed as a rapid screening procedure for identifying those in the early stages of Alzheimer's disease. Rather than taking an existing mental status examination as its starting point, the 7MS combines four tests: a 16-item enhanced cued recall procedure initially described in longer form by Grober and Buschke (1987); a semantic fluency task (animal naming); the Benton Temporal Orientation Test (Benton, Sivan, Hamsher, et al., 1994); and clock drawing (setting the hands to "twenty to four"), with a simplified 7-point version of the Freedman scoring procedure (M. Freedman, Leach, et al., 1994).

Age, education, and sex had no appreciable effects on test scores. A complex algorithm was developed for combining scores from the four tests into a single score that can be interpreted as the odds of having Alzheimer's dementia. Both interrater reliability for the overall score and test-retest reliability over a one to two month interval were high in 25 randomly selected Alzheimer patients and 25 control subjects (.92 and .91, respectively). Sensitivities and specificities for detecting dementia were also impressive in the larger sample of 60 patients with Alzheimer's disease and 30 healthy control subjects (>.90), even for patients with less severe Alzheimer's disease. The major limitations of the 7MS appear to be the small, homogeneous sample on which it was validated and the complex scoring algorithm required to obtain the total score.

Short Portable Mental Status Questionnaire (SPMSQ) (Pfeiffer, 1975)

This brief screening measure was published the same year as the MMSE but it was developed specifically for use with geriatric patients. The SPMSQ has played a key role in large-scale epidemiological studies designed to identify risk factors for cognitive and functional impairment, such as the National Institute of Aging's program, Established Populations for Epidemiological Studies of the Elderly (EPESE) (Chodosh, Reuben, et al., 2002; Fillenbaum, Landerman, Blazer, et al., 2001). The SPMSQ is a ten-question, ten-point test that is even more heavily weighted toward orientation than the MMSE: seven of its items involve orientation (e.g., date, place, mother's maiden name), two tap memory for current and previous presidents, and the last assesses concentration and mental tracking with serial threes. Test–retest reliability was .82 and .83 for two small groups of elderly control subjects (Pfeiffer, 1975) and .85 for nursing home patients (Lesher and Whelihan, 1986). A telephone version of the SPMSQ has been developed (Roccaforte et al., 1994). (The SPMSQ is not to be confused with the similarly structured and titled "Mental Status Questionnaire" [R.L. Kahn and Miller, 1978], also a ten item brief screening measure composed of orientation and general information items, but one that has been used much less in recent years.)

Test characteristics. Age affects SPMSQ performance, as it does performance on most brief screening instruments. Between ages 65–69 and 85–89, the average number correct for community-dwelling subjects dropped from 7.8 to 6.05 (Scherr et al., 1988); others have also shown that age has a pronounced effect on SPMSQ scores in these later years (Fillenbaum, Landerman, and Simonsick, 1998). Criteria for discriminating between intact subjects and three levels of impairment severity were based on a sample of almost 1,000 community-dwelling elderly Caucasian and African-American persons from the southern U.S, taking both education and race into account.

The specificity of the SPMSQ is very high (e.g., 96% in a clinical sample of 133 elderly patients, 40% of whom carried a diagnosis of dementia) (Pfeiffer data cited in Lorentz et al., 2002). However, like most screening instruments, its sensitivity is limited—peaking at 67% when the 10th percentile cut-off was applied to the clinical sample but dropping to 26% in an institutionalized sample. In a regression analysis, 47% of the variance in the SPMSQ was explained by only three items (date of birth, naming the previous president, and naming the day of the week), leading to the conclusion that these three items might well do the job of all ten (Fillenbaum, 1980).

Neuropsychological findings. Given its almost exclusive focus on orientation, the SPMSQ does not identify mildly impaired or early dementia patients to any reliable degree (G. Berg, Edwards, et al., 1987; Fillenbaum, 1980; Pfeiffer, 1975). One large epidemiological study demonstrated that community-dwelling elderly individuals who scored <7 on the SPMSQ were 2.60 (women) to 2.72 (men) times as likely to develop limitations in their ability to perform basic activities of daily living over the subsequent three years as those with higher scores (Moritz et al., 1995). In a study of over 2,500 hospitalized patients, those whose SPMSQ performance was mildly impaired were 2.8 times as likely as unimpaired individuals to have a first time admission to a nursing home within three months of dis-

charge, while those whose SPMSQ performance was moderately to severely impaired were 6.7 times as likely to be admitted to a nursing home (Sands et al., 2003). These findings suggest that the SPMSQ may be better suited to population-based screening to identify individuals at risk for functional impairment, who can then be closely monitored, than it is to the clinical assessment of individual patients.

Telephone Interview for Cognitive Status (TICS) (Brandt, Spencer, and Folstein, 1988)

This test was the first of several telephone instruments developed to provide follow-up documentation on patients who had been seen in clinic or for research but who had not returned for later examinations. Other telephone screening instruments for dementia such as the *Minnesota Cognitive Acuity Screen* (*MCAS*) (Knopman, Knudson et al., 2000) and the *TELE* (Gatz, Reynolds, et al., 2002; Jarvenpaa et al., 2002), have been published but have not yet seen widespread use. The TICS has been incorporated into several large epidemiological studies, including the National Academy of Sciences Registry of Aging Twin Veterans (Brandt, Welsh, et al., 1993; Gallo and Breitner, 1995) and the Nurses' Health Study beginning in 1995 (Grodstein, Chen, Pollen, et al., 2000; Grodstein, Chen, Wilson, et al., 2001). It has also been used to screen patients for a recent clinical trial of rofecoxib for amnestic mild cognitive impairment (Lines et al., 2003).

The TICS covers domains similar to the MMSE but affords a more sensitive assessment of memory. In its original form the TICS had 11 items and included an assessment of immediate—but not delayed—recall as none of the Alzheimer's dementia patients in the pilot study could recall any items after a delay. A subsequent modification of the instrument (TICS-m) incorporated delayed recall to increase its sensitivity in early dementia (K.A. Welsh, Breitner, and Magruder-Habib, 1993). Several items on the TICS-m test for orientation and general fund of knowledge (name, date, telephone number, President, and Vice President, for a total of 14 points); three items involve language (following a command to tap the phone five times, repetition, responsive naming, for a total of 8 points); two are mental tracking tasks (counting backwards and subtraction, for a total of 7 points); one requires the subject to generate word opposites (of "west" and of "generous", for 2 points); and one involves immediate and delayed recall of a 10-word list (20 points total). The maximum score is 51. A computer-assisted telephone interview version of the TICS-m has recently been developed (Buckwalter et al., 2002).

Test characteristics. TICS scores were modestly correlated with education for patients but not for control subjects, whose range of scores was more restricted. Test–retest reliability of the TICS after one to six weeks was .96 for 34 Alzheimer patients (Brandt, Spencer, and Folstein, 1988) and was comparably high in stroke patients over a one month retest interval (D.W. Desmond et al., 1994). A factor analysis of TICS-m responses in 4000 twin pairs identified four factors: memory (20 points); language/attention (17 points); personal orientation (10 points); and general information (4 points) (Brandt, Welsh, et al., 1993). A subsequent factor analysis of the TICS-m responses of over 6000 subjects responding to an advertisement for those with memory complaints yielded similar findings, excepting that the personal orientation and general information factors combined into a single factor (Lines et al., 2003).

In the original validation study, the TICS was given to both normal subjects and previously diagnosed Alzheimer patients who had scored at least 20 points on the Mini-Mental State. Not surprisingly, TICS scores were strongly correlated with MMSE scores (r = .94) (Brandt, Spencer, and Folstein, 1988), a finding that was later replicated in an Italian sample (Ferrucci et al., 1998). In the Brandt group's study, patient scores ranged from 0 to 31, those for control subjects were in the 31 to 39 range: applying a cutting score of 30, only one patient was misclassified, for a sensitivity of 94% and a specificity of 100%. Subsequent studies confirmed the ability of the TICS to detect Alzheimer patients with excellent accuracy (>99% sensitivity and 86% specificity when a cut-off of <28 was used), even in population studies with low base rates of Alzheimer's disease (Gallo and Breitner, 1995). With the data evaluated both cross-sectionally and longitudinally, the TICS differentiated healthy controls from those with mild or ambiguous cognitive impairment and from patients with dementia (Plassman, Newman, et al., 1994). The validity of the TICS-m was further substantiated in a recent study of patients three months after they had sustained subarachnoid hemorrhages (Mayer et al., 2002). Patients who scored <30 on the TICS-m were rated as significantly more handicapped overall and less independent in performing daily activities. They also reported greater anxiety, more depression, and poorer overall quality of life.

Briefer screening instruments

The introduction of medications for dementia in the mid-1990s heightened interest in very brief screening instruments, ones that could be administered in less

than five minutes and might be suitable for primary care and general neurology practices. The briefest of these—taking under two min—include three word recall, clock drawing tests (pp. 553–556), the Time and Change Test, which consists of clock reading and making change for a dollar (Froehlich et al., 1998), and the WORLD Test (Leopold and Borson, 1997), which asks subjects to spell "world" forward and backward and then arrange its letters in alphabetical order (Cullum, Thompson, and Smernoff, 1993). Unfortunately, none of these very brief tests has acceptable psychometric properties as a stand-alone screen for dementia (Lorentz et al., 2002).

The following two slightly longer screens are more promising.

Memory Impairment Screen (MIS). This is a four-item delayed free and cued recall procedure that incorporates category cues to facilitate acquisition and recall (Buschke, Kuslansky, et al., 1999). Subjects are shown a standard sheet of $8^1/_2 \times 11$" paper on which four words appear in large (24-point) uppercase letters, with each word derived from a different category. The subject is asked to read the items aloud and, when the examiner gives a category cue, to point to and read the item belonging to that category. After a two or three minute distraction period during which the subject counts from 1 to 20 forward and backward, the subject is asked to recall the four words in any order. Category cues are given for any items that are not spontaneously recalled. The total MIS score is twice the number of items retrieved on free recall (because it is assumed that these items would be retrieved on cued recall as well), plus the number of items retrieved on cued recall, for a total of 8 possible points. Two reasonably comparable ($r = .69$) alternate forms are available.

In a validation study with 483 community-dwelling elderly individuals, of whom 50 (10.4%) had dementia (Alzheimer's disease diagnosed in 39), the MIS proved surprisingly accurate in identifying patients with any form of dementia (sensitivity = .80 and specificity = .96 using a cut-off score of 4) or with Alzheimer's dementia (sensitivity = .87 and specificity = .96, also with a cut-off score of 4). Age, education, and gender did not significantly affect performance. In contrast, a standard three-word recall test had considerably poorer sensitivity (.65) and specificity (.85) as a screen for Alzheimer's dementia (Kuslansky et al., 2002). Buschke and his coworkers (1999) provide detailed tables on the performance of different cut-off scores as well as the probability of accurately identifying patients with dementia given differing base rates in the population. This enables clinicians and researchers to select the cut-off score that best meets their needs for optimizing sensitivity or specificity in a given population.

The Mini-Cog. This test combines uncued recall of three unrelated words (using words from the Cognitive Abilities Screening Instrument) with a clock drawing test (Borson, Scanlon, et al., 2000). The clock drawing test serves as the distractor between subjects' initial registration of the words (scored 0–3) and their subsequent recall of these words (also scored 0–3). Clock drawing is scored using the CERAD templates, yielding scores ranging from 0 (*normal*) to 3 (*severely impaired*) (Borson, Brush, et al., 1999). Scanlan and Borson (2001) found that inexperienced raters did nearly as well as experienced raters in scoring clock drawing. They suggested that any differences could be minimized with training in identifying clocks meeting criteria for mild impairment. Unlike the MMSE and the CASI, performance on the Mini-Cog was not influenced by education.

A classification algorithm for the Mini-Cog assigns subjects who recall none of the words to the "demented" group, those who recall all three words to the "nondemented" group, and those who recall one or two words as either "nondemented" if they perform normally (i.e., score 0) on the clock drawing test or "demented" if they exhibit any impairment (i.e., score 1–3) on clock drawing (Borson, Scanlon, et al., 2000). The Mini-Cog was initially validated in a heterogeneous sample of 249 community-dwelling older adults, about half of whom spoke—and were tested in—languages other than English and about half of whom met standard criteria for dementia (71% probable Alzheimer's dementia). Using this algorithm, the Mini-Cog demonstrated excellent sensitivity (99%) and specificity (93%), outperforming either test on its own. The sensitivity and specificity of the Mini-Cog were less impressive (approximately .75 and close to .90, respectively) in an epidemiological sample of over 1000 older adults in which the dementia prevalence was much lower (6.3%), but it performed as well as either the MMSE or a standard neuropsychological battery, identifying many subjects whose impairments were not recognized by their physicians (data reported in Lorentz et al., 2002).

Mental Status and Observer Rating Scale Combinations

Some assessment instruments include both a mental examination and a standardized observer- or informant-based rating format. In some instruments these two kinds of examination approaches are offered in separate sections. Structured patient interviews, however,

may provide examiners the opportunity of rating their observations while assessing specific cognitive functions.

Alzheimer Disease Assessment Scale (ADAS) (W.G. Rosen, Mohs, and Davis, 1984, 1986)

The *ADAS-Cognitive* subscale (*ADAS-Cog*) was the primary cognitive outcome measure in clinical trials that led to U.S. Food and Drug Administration approval of tacrine, the first medication approved for treatment of Alzheimer's disease (K.L. Davis et al., 1992). It soon replaced other clinical trial outcome measures that were either psychometrically deficient (e.g., Sandoz Clinical Assessment Geriatric [SCA-G] scale: Shader et al., 1974) or restricted in scope (e.g., the Selective Reminding Test), and has been used in all the major dementia treatment trials. In fact, the ADAS-Cog is one of two primary outcome measures required for clinical trials of new medications for Alzheimer's disease in the United States, the other being a clinician rating of global function. (The entire ADAS is usually administered in clinical trials, but the *ADAS-Noncognitive* subscale is considered a secondary outcome measure.) Numerous translations are available. For an overview of issues in selecting clinical outcome measures for dementia clinical trials, see Demers et al. (2000a,b) and Winblad et al. (2001).

ADAS-Cognitive subscale. W.G. Rosen and her colleagues (1984) selected items for the ADAS-Cog based on what they perceived to be the principal features of cognitive dysfunction in Alzheimer patients. Items cover *language ability* (25 possible points for naming objects and fingers and observer-rated comprehension of spoken language, expressive language, and word finding); *memory* (27 points for recall of instructions, word list recall and recognition); *praxis* (10 points), consisting of "constructional praxis" (copying geometric figures) and "ideational praxis" (preparing envelope to send to oneself); and *orientation* (8 points). Factor analyses of large data sets have essentially confirmed the conceptual framework underlying the ADAS-Cog, identifying three reproducible factors: memory, language, and praxis (Y.S. Kim et al., 1994; Talwalker et al., 1996).

The ADAS-Cog takes about 30–35 minutes to administer. Individual item scores are based on errors and generally range from 1 to 5, although some items have smaller or larger score ranges. The total ADAS-Cog score ranges from 0 to 70, with higher scores indicating greater impairment. The addition of a digit cancellation task, word learning with delayed recall, and a maze task has been recommended to improve sensitivity of the ADAS-Cog in assessing patients with mild Alzheimer's disease or those with mild cognitive impairment considered at risk of developing Alzheimer's disease (Mohs et al., 1997).

ADAS-Noncognitive subscale. The noncognitive portion of the ADAS consists of 10 items covering concentration, motor disturbances (tremors, pacing, and motor restlessness), appetite change, mood disturbance (tearfulness and depressed mood), behavioral disturbance ("uncooperativeness"), and psychotic symptoms (delusions and hallucinations). Some investigators have suggested dropping three of these items: concentration (beause of its high correlation [.78] with the ADAS-Cog), appetite disturbance (because it is not one of the cardinal behavioral disturbances in Alzheimer's disease), and tremor (because it is not characteristic of Alzheimer's disease). This would create a seven-item ADAS-Noncog that more purely reflects behavioral disturbances typical of Alzheimer's disease (D.B. Marin et al., 1997).

Ratings on the ADAS-Noncognitive are based on a clinician's observations, interview with the patient, and interview with a caregiver or other knowledgeable informant. Scores on individual items are rated from 0 (no impairment) to 5 (greatest impairment) for the week preceding the assessment; behavioral descriptors anchor the scale. With a maximum summation score of 50 on the full ADAS-Noncog, higher scores reflect more aberrant behavior. The ADAS-Noncognitive takes about four to six minutes to complete.

Test characteristics. Age and education had statistically significant effects on ADAS-Cog performance (Doraiswamy et al., 1995, 1997b). Scores declined with increasing age most noticeably in less educated subjects.

Interrater reliability coefficients for individual ADAS items ranged from .65 to .99 in Rosen's original sample (W.G. Rosen, Mohs, and Davis, 1986). The interrater reliability of the total ADAS was .82 to .83 in a subsequent study, with the ADAS-Cog subscale being considerably more reliable (.82–.90) than the ADAS-Noncognitive subscale (.42–.45) (Standish et al., 1996). Standardization of test administration and scoring, along with rigorous examiner training, substantially improved the interrater reliability of the ADAS-Noncognitive subscale (.85–.89). Over a one-month interval, test–retest item reliability coefficients for Alzheimer patients were in the .51 to 1.0 range in the Rosen group's original sample, with the ADAS-Noncognitive subscale producing the lower coefficients. In a separate study, test–retest reliability for the ADAS-Cog subscale alone was excellent (.91 over a 6-week period) (Talwalker et al., 1996). As expected, the

ADAS-Cog total score correlated strongly (−.76) with the MMSE; moreover, it did a better job than the MMSE in separating patients with different levels of cognitive impairment.

Neuropsychological findings. The ADAS in general—and the ADAS-Cog in particular—easily differentiated 15 Alzheimer patients from 15 elderly controls (W.G. Rosen, Mohs, and Davis, 1984). In fact, each individual ADAS-Cog item on its own successfully differentiated these groups. Group differences on the ADAS-Noncognitive were smaller in magnitude and statistically significant on only three items. The ability of the ADAS-Cog to differentiate patients with Alzheimer's disease from elderly controls was subsequently replicated by Zec, Landreth, and colleagues (1992) in a larger sample. The ADAS-Cog subscale can also successfully distinguish patients who differ in their dementia severity: for example, it discriminated patients with moderate dementia (GDS = 4) from those with moderately severe dementia (GDS = 5), with the orientation item being the best discriminator at these levels of dementia severity.

Alzheimer patients obtained consistently higher (i.e., worse) scores on both ADAS subscales at 12- and 18-month retests, while normal elderly patients' scores remained essentially unchanged (W.G. Rosen, Mohs, and Davis, 1986). The rate of deterioration is more pronounced on the ADAS-Cog as opposed to the ADAS-Noncognitive subscale and is greatest among patients with moderate to severe—as opposed to mild or very severe—impairment at baseline.

Blessed Dementia Scale (BDS) (Blessed, Tomlinson, and Roth, 1968)

This two-part scale was originally called simply the "Dementia Scale," but many users added the senior author's name to avoid confusion with other similarly named instruments. It was originally designed to evaluate the relationship between mental deterioration in the elderly and pathological changes in brain tissue observed on autopsy. The first part, the *Blessed Rating Scale* (*BRS*), registers changes in behavior and daily functioning reported by informants. The second part, the *Blessed Information-Memory-Concentration Test* (*BIMC*), consists of many of the most commonly used mental status questions examining the areas announced in the test's title (see below). A six-item mental status test taken from this portion of the BDS—(the *Orientation-Memory-Concentration Test*—also usually carries Blessed's name (see p. 715). All three instruments have had wide application, but only occasionally are the rating scale and one of the two mental status tests used together. In recent years, Mungas and Reed (2000) used sophisticated psychometric methods to produce a 25-item cognitive screening instrument composed of 10 items from the Blessed Rating Scale, 12 from the Blessed Information-Memory-Concentration Test, and three from the MMSE. This new instrument is not only brief and easy to administer but also statistically reliable. Unfortunately, it has not been widely adopted despite its obvious psychometric appeal.

Blessed Rating Scale (BRS). This scale has been variously referred to in the literature as the "Dementia Score" (Hachinski, Iliff, et al., 1975; see Table 18.1), the "Dementia Rating Scale (DRS)" (Eastwood et al., 1983), "Part I of the Blessed Dementia Rating Scale (BDRS)" (Y. Stern, Mayeux, Sano, et al., 1987), and the "Blessed Dementia Scale (BDS)" (J.C. Morris, Heyman, et al., 1989). Here it is called the *Blessed Rating Scale* (*BRS*) as the most descriptive and least confusing title.

TABLE 18.1 Dementia Score

Feature	*Score*
CHANGES IN PERFORMANCE OF EVERDAY ACTIVITIES	
1. Inability to perform household tasks	1
2. Inability to cope with small sums of money	1
3. Inability to remember short list of items, e.g., in shopping	1
4. Inability to find way about indoors	
5. Inability to find way about familiar streets	1
6. Inability to interpret surroundings	1
7. Inability to recall recent events	1
8. Tendency to dwell in the past	1
CHANGES IN HABITS	
9. Eating	
Messily with spoon only	1
Simple solids, e.g., biscuits	2
Has to be fed	3
10. Dressing	
Occasionally misplaced buttons, etc.	1
Wrong sequence, commonly forgetting items	2
Unable to dress	3
11. Sphincter control	
Occasional wet beds	1
Frequent wet beds	2
Doubly incontinent	3
12. Increased rigidity	1
13. Increased egocentricity	1
14. Impairment of regard for feelings of others	1
15. Coarsening of affect	1
16. Impairment of emotional control	1
17. Hilarity in inappropriate situations	1
18. Diminished emotional responsiveness	1
19. Sexual misdemeanor (appearing de novo in old age)	1
20. Hobbies relinquished	1
21. Diminished initiative or growing apathy	1
22. Purposeless hyperactivity	1

From Hachinski et al. (1975) *Archives of Neurology 32*, p. 633. © 1975, American Medical Association.

The BRS measures how well patients have functioned in their usual environment during the preceding six months. Information typically comes from family informants or caregivers, but medical records can be used as well. Summing the 22 items together yields scores ranging from 0 to 28, with higher scores indicating greater incapacity. As a rule of thumb, persons receiving scores less than 4 are considered to be unimpaired, scores of 4 to 9 indicate mild impairment, and scores of 10 and higher are in the moderate to severe impairment range (Eastwood et al., 1983). Based on clinical experience, a slightly higher threshold (15) for moderate impairment has been suggested (Y. Stern, Mayeux, Sano, et al., 1987).

Test characteristics. The test–retest stability of the BRS over 4 weeks in a sample of 68 nondemented elderly subjects was estimated to be .79 (Erkinjuntti et al., 1988). The first 11 items alone show a satisfactory test–retest reliability over 4 weeks ($r = .68$). These items can be used alone to distinguish those with dementia of varying severities. This version of the BRS was adopted for CERAD (*BDRS-CERAD version:* J.C. Morris, Heyman, et al., 1989), with each item phrased positively instead of negatively (e.g., "1. *ability* to perform household tasks," as opposed to "*inability* to perform household tasks") for a total possible score of 17.

Neuropsychological findings. In the original study of 60 elderly persons who had come to autopsy, some had functional psychiatric diagnoses, some were delirious, some were demented, and a small number of physically ill patients served as control subjects (Blessed, Tomlinson, and Roth, 1968). Patients diagnosed with senile dementia were more impaired on the BRS than those in the other groups, with the correlation between the BRS total score and the mean senile plaque count reaching .77.

When repeated over time, the BRS can be used to monitor dementia progression (J.C. Morris, Heyman, et al., 1989; Y. Stern, Mayeux, Sano, et al., 1987) by documenting the behavioral alterations that accompany cognitive deterioration (Van Gorp and Cummings, 1989). A longitudinal study using the BRS documented the variability that is often observed clinically in the timing and magnitude of change across different aspects of behavior (Y. Stern, Hesdorffer, Sano, and Mayeux, 1990). Examining changes on four BRS factor scores derived from a principal components analysis of patients' individual item scores (i.e., Cognitive, Personality Change, Apathy/Withdrawal, and Basic self-care), the Stern group noted that cognitive deficiencies affecting instrumental ADLs were evident early and worsened throughout the disease course, whereas changes in Basic self-care did not occur until four to five years into the illness. Increases in Personality Changes and in Apathy/Withdrawal became more common as the disease progressed but these behavioral changes tended to fluctuate more than the cognitive symptoms.

Information-Memory-Concentration Test (BIMC). This part of the Blessed scale contains three sections. The "Information Test" (15 points) inquires into the patient's personal orientation. "Memory" (16 points) asks for recall of remote memories—both "personal" (e.g., school attended) and "non-personal" (e.g., date of World War II)—and includes a name and address to be learned for recall five minutes later. "Concentration" consists of three items, months backwards, and counting from 1 to 20 and 20 to 1, with each scored 0–2 for a total of 6 possible points. A perfect performance earns a score of 37.

When given to nursing home patients, both two to four week test–retest and split-half reliability coefficients were very satisfactory (.88 and .89, respectively). Blessed and his colleagues (1968) reported that the BIMC score had a correlation of $-.59$ with senile plaque count in their population of elderly patients. This finding was replicated exactly in a study that also included mentally intact subjects along with Alzheimer patients and other demented patients (Katzman, Brown, Fuld, et al., 1983). Among Alzheimer patients, an average annual decline in the BIMC score of 4.4 was found, independent of age, except for the most intact whose initial rate of decline was less (Katzman, Brown, Thal, et al., 1988). For individual patients, however, the rate at which scores declined was quite variable.

Orientation-Memory-Concentration Test (OMC). Upon observing that six items from the BIMC and the Mental Status Questionnaire (MSQ) correlated more highly with the total BIMC than the total MSQ score, Katzman, Brown, Fuld, and their colleagues (1983) selected them for a brief mental status screening test. They include orientation for time (month, year, and time of day), counting from 20 to 1, months backward, and repeating a brief phrase. Points are given for failures—with individual items differentially weighted—for a total possible score of 24. Calling this test the *Short Orientation-Memory-Concentration Test* (*SOMCT*), Lesher and Whelihan (1986) reported limited internal consistency (split-half correlation of .37, not surprising for such a brief and heterogeneous set of items) but good test–retest reliability ($r = .80$). The Katzman group found that over 90% of intact elderly subjects earned weighted error scores of 6 or less; error scores greater than 10 are strong indicators of dementia.

Brief Cognitive Rating Scale (BCRS) (Reisberg, Schneck, Ferris, et al., 1983)

This two-part scale rates both responses to mental status questions and qualitative characteristics observed in a semistructured assessment interview. Whenever possible, the interview is conducted with a spouse or caregiver present to provide realistic information when the patient's self-reports are inaccurate. The first part consists of five "Axes": I. Concentration and calculating ability; II. Recent memory; III. Remote memory; IV. Orientation; V. Functioning and self-care. Each axis has a 7-point rating scale with descriptors ranging from "No objective or subjective evidence of deficit . . . " to descriptions of severe impairment in that domain. Scores of 1 and 2 are considered to be within the range of intact functioning, while scores of 4 or greater indicate moderate to severe dementia. Scores for the five axes in the first part of the BCRS can be averaged and interpreted on a 7-point *Global Deterioration Scale* (*GDS*) for which each score level indicates the same degree of severity as the axis score level (Reisberg and Ferris, 1982; Reisberg, Ferris, de Leon, and Crook, 1982). Intercorrelations among the first five axes ranged from .83 to .97, indicating considerable overlap in ratings of these functions (Reisberg, Ferris, Borenstein, et al., 1986). On the basis of assessments of 50 subjects (a relatively intact sample heavily skewed toward lower GDS scores), correlations of Axes I through V with neuropsychological tests and test items in common use were all positive and significant ($p < .001$).

The second part of the BCRS—"Language, Motoric, and Mood Concomitants"—is named for each of its three "axes" which also have 7-point rating scales ranging from highest, "No subjective or objective [problems in that area]," to lowest, "Inability to perform the functions under consideration." (W.G. Rosen and her coworkers [1986] cautioned that the language scale does not adequately cover speech and language, noting that speech comprehension is not included among the descriptors, for example.) The three axes comprising the second part of the BCRS are separated from the first five because the authors did not consider them to be as closely or regularly associated with disease progression in Alzheimer patients as the first five axes. Individual correlations of Axes VI, VII, and VIII with the summed score for Axes I through V (GDS) were in the .71–.88 range.

The principal application of the BCRS has been the use of its first five axes to derive the Global Deterioration Scale. The assumption underlying the development of the GDS and other global rating scales is that all of the functions covered in Part I of the BCRS will deteriorate at a similar rate in Alzheimer's disease, an assumption that does not hold in many individual cases. Nonetheless, the GDS and other global rating scales are widely used in clinical dementia research and clinical trials of antidementia medications to provide an index of overall level of functioning, or stage of dementia, and change over time. (For an astute review of the psychometric properties of the GDS and two other commonly used global rating scales for dementia, the Clinical Global Impression (CGI) scales and Clinical Dementia Rating (CDR), see Oremus et al., 2000.)

Scales for Rating Observations

These scales can focus on many different aspects of mood, behavior, and functional abilities. Behavioral and mood problems in Alzheimer's disease are common and have a profound effect on the level of care that a patient must have and caregiver burden, not to mention the cost of such care. Assessment of mood, behavior, and functional abilities in Alzheimer patients is complicated by the fact that the patient may not be able to provide reliable responses, particularly in the later stages of the disease. Consequently, the clinician must base ratings on direct observation of the patient's behavior (as on the ADAS-Noncognitive subscale or the Blessed Rating Scale described earlier), or on information derived from an interview with a relative or other knowledgeable informant. A review of measures of functional abilities is beyond the scope of this book. For interested readers, two frequently used instruments are the *Barthel Index* (Mahoney and Barthel, 1965) with its recent modification (Novak et al., 1996) and the *Functional Independence Measure + the Functional Assessment Measure* (*FIM/FAM*) (Uniform Data Systems, 1987, 1993).

Behavioral Pathology in Alzheimer Disease Rating Scale (BEHAVE-AD) (Reisberg, Borenstein, Franssen, et al., 1987)

Potentially remediable behavioral disturbances common in Alzheimer's disease are the subject of this rating scale which reviews seven categories of behavior symptoms: Paranoid and Delusional Ideation: Hallucinations; Activity Disturbances (e.g., wandering); Aggressivity; Diurnal Rhythm Disturbances; Affective Disturbances; Anxieties and Phobias. The symptoms in these categories often create problems for caregivers but may be ameliorated pharmacologically. Each of the 25 symptoms is rated on a 4-point scale (from 0 = Not present, to 3 = Present—at a level intolerable to caregiver). The rating form also provides space for elaborating details of some of these problems. Information

for ratings comes from patients' spouses and caregivers, and from clinical observations. Unlike many scales, the BEHAVE-AD rates the impact of behavior on caregivers. Five factors accounting for 40% of the variance were identified: agitation/anxiety, psychosis, aggression/fear of being left alone, depression, and activity disturbance/delusion that one's house is not one's home (Harwood, Ownby, et al. 1998). A version is available with a symptom frequency-weighted score, the *BEHAVE-AD-FW,* which measures both the magnitude and prevalence of behavioral symptoms (Monteiro et al., 2001).

Ratings of a group of 120 Alzheimer patients at different stages of the disease, from mild to dilapidated, brought out the typical course of development and eventual disappearance of these symptoms, with most having their peak occurrence in the late middle stages of the disease (Reisberg, Franssen, et al., 1989). A longitudinal study using the BEHAVE-AD showed that activity disturbance was a common and relatively persistent symptom in the mild stages of Alzheimer's disease (Eustace et al., 2002). Anxiety, paranoid ideation, and aggression were moderately persistent; but depressive symptoms usually lasted less than one year. Patients with frontotemporal dementia had significantly worse global BEHAVE-AD scores with more verbal outbursts and inappropriate activity compared to Alzheimer patients (Mendez, Perryman, Miller, and Cummings, 1998). The BEHAVE-AD often is used as an outcome measure in dementia treatment trials (Brodaty et al., 2003).

Geriatric Evaluation by Relative's Rating Instrument (GERRI) (G.E. Schwartz, 1983)

This scale was conceived to assess behavioral functioning in elderly persons showing signs of mental decline. The 49 items cover a broad spectrum of behaviors observable in the home. Persons in close contact with the patient (usually a relative or caregiver) rate the patient on each item by means of a 5-point scale ranging from "Almost All the Time" to "Almost Never" with a "Does Not Apply" option. Correlational analyses identified three item clusters: Cognitive Functioning (21 items), Social Functioning (18 items), and Mood (10 items). Using two sets of informants for 45 dementia patients at different severity levels, the total score interrater reliability was .94; for the three clusters, it was .96, .92, and .66, respectively. GERRI scores varied significantly with severity rating scores (Global Deterioration Score), the Cognitive and Social clusters discriminating significantly between three levels of dementia severity ($p < .0001$). In a large sample of dementia patients, GERRI scores correlated significantly ($r = .40$) with ADAS-Cog scores (Doraiswamy, Bieber, et al., 1997b).

The GERRI has been used as an outcome measure in treatment trials with geriatric and dementia patients (Le Bars et al., 2002). R.S. McDonald (1986) cautioned that untrained and emotionally close observers such as relatives may be biased in their observations, but acknowledged the advantages of an observer reporting on patient behavior—and behavioral changes—in the natural setting of the home.

The Neuropsychiatric Inventory (NPI) (Cummings, Mega, Gray, et al., 1994)

Developed to assess a wide range of behaviors common in dementia patients, ten behavior domains are evaluated: delusions, hallucinations, dysphoria, anxiety, euphoria, agitation/aggression, apathy, irritability/lability, disinhibition, and aberrant motor behavior. An informant, preferably the daily caregiver, is asked scripted questions about the patient's behavior during the previous month. Each section has screening questions and if the behavior has occurred, more detailed questioning probes the *frequency* on a 4-point scale and *severity* on a 3-point scale. Two additional scales were later added to assess sleep and appetite/eating disorders (Cummings, 1997). Also, added to each domain is a 6-point caregiver distress scale which ranges from 0 (no distress) to 5 (very severe distress). It was suggested for the original scale that the interview can be brief (7–10 min), but some caregivers elaborate their answers and require considerably more time.

Test characteristics. Interrater reliability and internal consistency of the scale were high (Cummings, Mega, Gray, et al., 1994). Test–retest reliability by a second interviewer within 3 weeks generally was adequate, with the lowest correlations (.53 for frequency, .51 for severity) for Irritability/lability. The NPI's correlation with the BEHAVE-AD was .66 for the total score. Most subscales correlated well with the corresponding BEHAVE-AD subscale except NPI dysphoria which had a .33 correlation with BEHAVE-AD Affective Disturbances. The authors state that the Dysphoria scale items were selected to represent core psychological and behavioral manifestations of depression and to exclude dementia symptoms. Three factors characterized the behavior symptoms of a large group of dementia patients: mood/apathy, psychosis, and hyperactivity (Aalten et al., 2003).

Neuropsychological findings. All behavior problems assessed by the NPI were greater in Alzheimer patients compared to age-matched control subjects (Mega,

Cummings, et al., 1996). The most common was apathy, which was exhibited by 72% of patients, followed by agitation, which was displayed by 60%. The NPI differentiated the behavioral symptoms of Alzheimer's and Parkinson's diseases (Aarsland et al., 2001) as Alzheimer patients had more aberrant motor behavior, agitation, disinhibition, irritability, euphoria, and apathy, while more hallucinations were reported for the Parkinson patients. The NPI has also been used to assess psychiatric symptoms in patients with multiple sclerosis (Diaz-Olavarrieta et al., 1999). Symptoms were present in 95% of patients, the most common being depression (79%) and agitation (40%). Although euphoria was once described as a common characteristic of patients with multiple sclerosis, only 13% showed this symptom. Euphoria was more common in patients with moderately severe frontotemporal MRI abnormalities. The NPI has also been used to assess psychiatric symptoms in other, mostly subcortical, neurodegenerative disorders (Litvan, Cummings, and Mega, 1998).

A self-administered NPI. A paper-and-pencil caregiver questionnaire, the *NPI-Q,* has been developed (Kaufer, Cummings, et al., 2000). The questionnaire format saves time for the examiner as, the authors say, most caregivers can complete the form in five minutes or less. Adequate convergent valdity and test–retest reliability were obtained. Correlations between the NPI and the NPI-Q were high (.90). More symptoms were reported on the NPI-Q than the NPI.

TRAUMATIC BRAIN INJURY

Although behavioral rating scales and inventories in general use can be adapted for traumatically brain injured patients, many of their particular issues have led to the development of specialized assessment instruments. Perhaps the most important of these issues is predicting outcome, since most TBI victims have their future before them. Many aspects of outcome are closely associated with the severity of damage such that particular attention has been given to assessing initial severity on the basis of clinical observations. Some measures can be used both to define severity of injury and to establish improvement and/or deterioration over time (e.g., Glasgow Coma Scale). A second issue has been the assessment of a condition in which rapid change is the rule, as is the case particularly in the first few months after return to consciousness. Not infrequently an examiner will have begun an examination of such a patient on a Thursday or Friday and had to discontinue a test before completing it only to find, on the following Monday or Tuesday, that the patient's new performance level has rendered the original data obsolete. Moreover, in the early stages, the rate of change becomes an important feature in itself. Still another issue concerns the enormous intraindividual variability in performance levels that characterizes so many head injury patients. A thorough neuropsychological examination of some patients may require use of many different measures ranging in complexity and sophistication from infant scales to college aptitude tests, depending on severity and time since injury. Social adjustment is another issue that must be dealt with in assessments as some TBI patients, especially those who survived a severe injury, regain most of their premorbid physical competencies and many of their original cognitive abilities while judgment, self-control, and social skills and sensitivity remain impaired. The disparities between what these patients are capable of doing and what they are competent to do result in patterns of social maladaptation peculiar to them which the usual inventories of behavioral or social problems do not handle well.

Recent changes in the World Health Organization (WHO) system for evaluation of diseases and disorders are likely to affect our choice of outcome measures for TBI. The *International Classification of Impairments, Disabilities, and Handicaps* (*ICIDH*; World Health Organization, 1980) resulted from the need to assess the effectiveness of health care. Gray and Hendershot (2000) note that such evaluation was relatively uncomplicated when the health care system was dealing with acute disease or when the patient was cured or died. Now many patients live with chronic diseases and disorders, and the consequences need evaluation. "Impairments, disabilities, and handicaps" were thus included in the medical model. This classification has met with criticism for a variety of reasons, including overlap and ambiguity in the relationships between impairment, disability, and handicap, and not enough consideration of environmental and other factors (Gray and Hendershot, 2000; Pfeiffer, 1998; Fougeyrollas, 1995; Whiteneck, Fougeyrolloas, and Gerhart, 1997). Also the model emphasized the negative aspects of disease and disorders, not the competencies of individuals.

The current model, the *International Classification of Functioning, Disability, and Health* (*ICFDH*) (World Health Organization, 2001) emphasizes health and health-relevant components of well-being. It has two parts: Part 1. *Functioning and disability* incorporates the components "body functions and structures" and "activities and participation". Part 2. *Contextual factors* include the components "environmental factors" and "personal factors". Domains, constructs, positive aspects, and negative aspects are described for each of these four components with qualifiers for some components. The component "body functions and

structures" is associated with changes in body functions and body structures: the positive aspect is functional and structural *integrity*, the negative aspect is *impairments*. *Localization* is a qualifier. The component "activities and participation" is associated with life tasks and actions; *capacity* (executing tasks in a standard environment) and *performance* (executing tasks in the current environment). The positive aspects are activities and participation; the negative aspects are activity limitation and participation restriction. *Assistance* is a qualifier. The component "environmental factors" involves external influences in functioning and disability; specifically, the facilitating or hindering impact of features of the physical, social, and attitudinal world: the positive aspect is *facilitators* and the negative aspect is *barriers/hindrances*. A qualifier is *extent* or *magnitude* and a second qualifier is *subjective satisfaction*. The first three components are quantified on the same scale from 0 (no problem) to 4 (complete problem). Finally, the component "personal factors" is associated with internal influences on functioning and disability respectively, that is, the impact of attributes of the person. However, positive and negative aspects of personal factors are termed "nonapplicable." It is puzzling that personal factors are defined as ". . . the particular background of an individual's life and living, and comprise features of the individual that are not part of a health condition or health states;" yet these features include age, sex, race, even coping style, overall behavior pattern, psychological assets, and other variables that are clearly related to health condition and health states, similar to environmental factors. No classification of personal factors is attempted, just as positive and negative aspects and qualifiers are not given. Assessing personal factors and integrating them into this evaluation are to be done by the clinician. Much work still needs to be done on the ICF but the emphasis on environmental factors is one of many improvements (Gray and Hendershot, 2000) and the discussion of personal factors is a beginning. Response to the ICF has been generally positive, although problems are being noted in its application (Chopra et al., 2002; Dahl, 2002; Willems and de Kleijn-de Vrankrijker, 2002). New measures are likely to be developed (Steiner et al., 2002) and changes will be made to current outcome measures in order for them to be consistent with this classification (Stineman et al., 2003).

Evaluating Severity

Glasgow Coma Scale (Teasdale and Jennett, 1974)

Although it has "coma" in its title, this brief assessment technique can be used to describe all posttraumatic states of altered consciousness from the mildest confusional state to deep coma (see Table 18.2). A coma score, the sum of the highest score in each dimension, can be calculated. In evaluating injury severity, a GCS range of 3 to 8 is considered severe, 9 to 12 is moderate, and 13 to 15 is mild (Rimel, Giordani, et al., 1982; see Table 18.3). Coma has been defined as occurring when the GCS is ≤8 in a patient without spontaneous eye opening, ability to obey commands, or comprehensible speech (H.S. Levin, Williams, et al., 1988). The simplicity of the GCS allows it to be used reliably by emergency medical technicians in the field as well as by nursing personnel and doctors (Menegazzi et al., 1993). The inclusion of three response dimensions makes it possible to evaluate level of conscious-

TABLE 18.2 Glasgow Coma Scale

The Glasgow Coma Scale Response Chart (GCS)		
Examiner's Test	*Patient's Response*	*Score*
Eye opening		
Spontaneous	Opens eyes normally	4
Speech	Opens eyes when asked in loud voice	3
Pain	Opens eyes to pain (e.g., pinch)	2
Pain	Does not open eyes	1
Verbal		
Speech	Carries on a conversation correctly and demonstrates intact orientation	5
Speech	Speaks, seems confused and disoriented	4
Speech	Talks to examiner but speech makes no sense	3
Speech	Makes unintelligible sounds	2
Speech	Makes no noise	1
Best motor response		
Commands	Follows simple commands	6
Pain	Pulls examiner's hand away on painful stimuli (localizes pain source)	5
Pain	Pulls a part of body away on painful stimuli (withdraws)	4
Pain	Flexes body inappropriately to pain (abnormal flexion)	3
Pain	Decerebrate posturing (abnormal extension)	2
Pain	No motor response to pain	1
		Range 3–15

TABLE 18.3 Severity Classification Criteria for the Glasgow Coma Scale (GCS)

Classification	*GCS*		*Coma Duration*
Mild	≥13	or	≤20 minutes
Moderate	9–12	or	No longer than within 6 hours of admission
Severe	≤8*	or	>6 hours after admission

*Patients with GCS ≤8 are considered to be in coma (M.R. Bond, 1986).

TABLE 18.4 Frequency of "Bad" and "Good" Outcomes Associated with the Glasgow Coma Scale (24-Hour Best Response)

Coma Response Sum	n	*Dead/ Vegetative (%)*	*Moderate Disability/Good Recovery (%)*
≥11	57	7	87
8–10	190	27	68
5–7	525	53	34
3, 4	176	87	7

Adapted from Jennett (1979)

ness when vision or speech, for example, is compromised by factors other than impaired consciousness. Moreover, it can be used repeatedly to provide longitudinal data on the course of improvement during the earliest posttrauma period. Its greatest virtue is that it has proven to be a good predictor of outcome (e.g., Jennett, Teasdale, and Knill-Jones, 1975; H.S. Levin, Grossman, Rose, and Teasdale, 1979; see Table 18.4), albeit not always a strong predictor (Zafonte, Hammond, et al., 1996). It is also useful in predicting outcome from other medical conditions (Bhagwanjee et al., 2000; Gotoh et al., 1996; Mullie et al., 1988; Plum and Carona, 1975).

The Glasgow Coma Scale has been just about universally accepted as the standard measure for determining severity of injury in patients whose consciousness is compromised. The mortality rates for patients (seen at medical centers) with a GCS score ≤8 for more than four hours run in the 50 to 88% range (Eisenberg, 1985; Teasdale and Mendelow, 1984). Older age at injury is highly related to mortality and morbidity among those having a GCS of 3–8 (Kilaru et al., 1996; Quigley et al., 1997). GCS scores are significantly related to depth of lesions. Lesions in deep central gray matter or the brain stem tend to be associated with a lower GCS than cortical or subcortical white matter lesions (H.S. Levin, Williams, et al., 1988). In children and adolescents with moderate to severe TBI, depth of lesion was most predictive of the Disabilty Rating Scale (DRS) score at time of discharge from rehabilitation, while GCS better predicted the one year DRS score (Grados et al., 2001). At one month postinjury, of patients given a neuropsychological test battery, those with moderately severe injuries (GCS = 8–10) performed, on the average, less well than those with milder injuries (GCS ≥ 11) who, in turn, performed below levels obtained by matched control subjects; most coma survivors were still untestable at one month (Dikmen, McLean, Temkin, and Wyler, 1986). However, after three months, the GCS did not distinguish between mildly and moderately injured patients with respect to rates of return to employment (Rimel, Giordani, et al., 1982). Community integration and vocational outcome (J. Fleming et al., 1999) as well as patient and family reports of quality of life and social adjustment (P.S. Klonoff, Costa, and Snow, 1986) relate directly to initial GCS measures.

Despite its demonstrated usefulness, questions arise as to which GCS measurement indicates severity of injury: the emergency medical service GCS (taken at the scene or in the ambulance), the initial Emergency Room GCS (frequently called the postresuscitation GCS), the Best Day-1 GCS, the Worst Day-1 GCS, the Best Day-1 motor score, or GCS 6 hours post injury? All of these have been used in studies. Often rehabilitation researchers use the GCS on admission to their facility to indicate severity of injury but this GCS may or may not represent the initial severity of injury. Each of these measures provides a "snapshot" of what may happen during the critical first 48 to 72 hours postinjury, especially with the more serious injuries. J.M. Williams (1992) noted that differences in scale range for the three tested response modalities can bias the evaluation, depending on which modalities are operative. It would be fair to say that clinicians such as intensivists and neurosurgeons treating these patients do not rely on one GCS but often on hourly GCS scores, continuous clinical monitoring data, and serial CT scans to determine the status of the patient and necessary treatment on an ongoing basis during this period.

The GCS also has some inherent problems. Some trauma patients are lucid initially at the scene of the accident but have to be sedated for agitation or anesthetized and intubated for medical emergencies. These circumstances artificially lower their GCS on admission to the ER. Others deteriorate on transport to the hospital or in the ER or in neurosurgery intensive care, and earlier scores may not be representative of the eventual severity of injury. If a patient goes to surgery and is anesthetized, the GCS drops to 3 for several hours. A patient with a relatively mild head injury that would not produce a low GCS score may have a period of time on the record when the GCS score is low, suggesting to the naive reader that there was some neurological deterioration. Moreover, intoxicated patients may produce unreliable GCS scores with impaired consciousness attributed inappropriately to head trauma severity in some cases, to alcoholic stupor in others. Alcohol reduces admission GCS (M.P. Kelly, Johnson, et al., 1997; Sloan et al., 1989). The effects of alcohol are likely to be seen in the first six hours after injury. Our [hjh] experience with drugs given to patients while they are in intensive care suggests that some drugs do not affect the GCS (e.g., mannitol), some have large effects

(e.g., entomidate), and some have additive effects (e.g., hydrocodeine). Drug use by patients and metabolic alterations due to injuries not directly involving the brain can also affect level of consciousness resulting in a misleading GCS score (Stambrook, Moore, Lubrusko, et al., 1993). All of these effects need to be taken into account in trying to understand variations in the GCS of a patient over time.

Eisenberg (1985) noted two other important problems with the GCS: Some examination modalities may not be measurable during the first few days when patients who are intubated or have a tracheotomy cannot talk, eyes swollen from facial injuries (*ecchymosis*) will not open, and paralysis or immobilization for treatment purposes precludes limb movement. Of real concern is the way in which components of the GCS are scored under such circumstances in various medical centers. A national telephone survey of Level I trauma centers (Buechler et al., 1998) found that 26% of centers gave intubated patients 1 point for the verbal component added to the eye and motor scores, 23% scored a total GCS of 3, 16% estimated GCS with "T" given for the verbal component (16%), 10% gave "unknown" as the score, another 10% gave a score of 15, and for 15% the method of scoring was unknown. Such wide GCS scoring variations even among Level I trauma centers raises questions about institutional, state, and national databases; epidemiological and outcome research could be adversely affected by such scoring variations. The second problem noted by Eisenberg concerns the sacrifice of a richer data base for higher interexaminer and intersite reliability; but loss of information about when and how the GCS was scored will lower predictive accuracy. While it is a generally useful guideline to injury severity, the times that the GCS was measured and the circumstances surrounding the first few hours and days after injury must be taken into account in determining how much weight to give it as a predictor in the individual case.

Rancho Los Amigos Scale: Levels of Cognitive Functioning (Hagen, 1984; Hagen, Malkmus, et al., 1979)

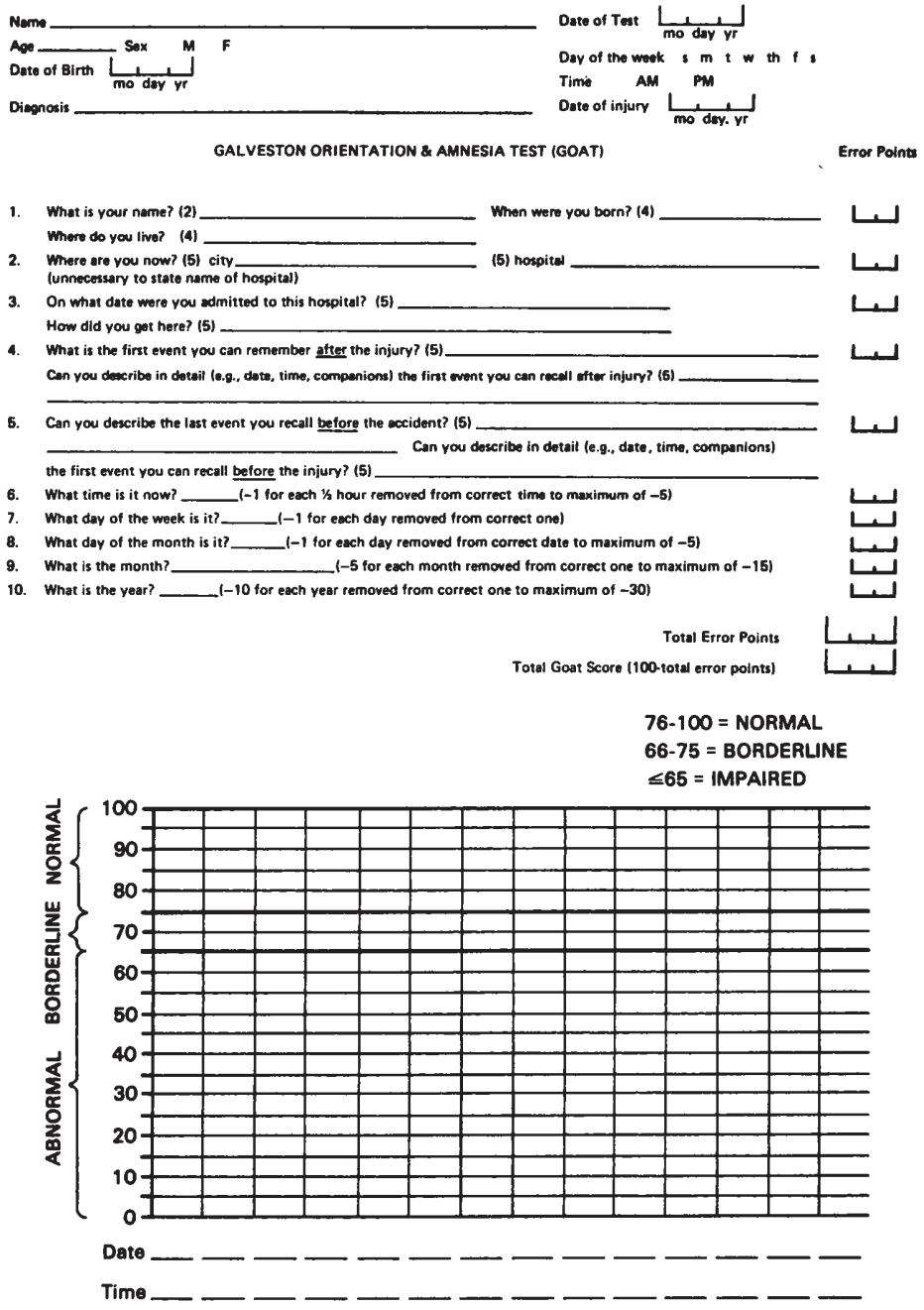
This scale, typically referred to as the "Rancho scale," has been used to track improvement (Kay and Lezak, 1990), for evaluating potential (Story, 1991), for planning and placement purposes (Mysiw et al., 1989), and to measure outcome and treatment effects (Lal et al., 1988; Razack et al., 1997). Its main focus is on cognitive functioning in the broadest behavioral sense. It differentiates eight levels of functioning covering much of the observable range of psychosocially relevant behaviors following TBI (see Table 18.5). An often implicit assumption that clinicians make about this scale is that the course of improvement following head trauma will follow the levels outlined therein. It was developed for use by clinical and rehabilitation staff.

TABLE 18.5 The Eight Levels of Cognitive Functioning of the "Rancho Scale"

1. *No Response:* The patient is in deep coma and completely unresponsive.
2. *Generalized Response:* The patient reacts inconsistently and nonpurposefully to stimuli in a nonspecific manner.
3. *Localized Response:* The patient reacts specifically but inconsistently to stimuli, orienting, withdrawing, or even following simple commands.
4. *Confused-Agitated:* The patient is in a heightened state of activity with severely decreased ability to process information.
5. *Confused, Inappropriate, Non-agitated:* The patient appears alert and is able to respond to simple commands fairly consistently; however, with increased complexity of commands or lack of any external structure, responses are nonpurposeful, random, or at best fragmented toward any desired goal.
6. *Confused-Appropriate:* The patient shows goal-directed behavior but is dependent on external input for direction.
7. *Automatic-Appropriate:* The patient appears appropriate and oriented within hospital and home settings, goes through daily routine automatically, but frequently robot-like, with minimal to absent confusion, and has shallow recall of what he/she has been doing.
8. *Purposeful and Appropriate:* The patient is alert and oriented, is able to recall and integrate past and recent events, and is aware of and responsive to his environment.

Reprinted from Kay and Lezak (1990)

The three highest levels of the Rancho scale tend to reflect cognitive improvement as measured by language skills (Wiig et al., 1988). Thus, patients at level VI were less able to understand metaphoric expressions or to compose sentences from sets of words than those at level VII, but these language tests did not differentiate level VII from level VIII patients. Low Rancho scale levels on admission to rehabilitation hospitals indicate patients at risk for abnormal swallowing, aspiration, delay in initiation of oral feeding, and delay in total oral feeding (L.E. MacKay et al., 1999a,b). The Rancho scale can discriminate between patients returning to competitive employment and those requiring vocational training or supported work but is not sensitive to differences in lower levels of vocational potential (Mysiw et al., 1989). Sohlberg and Mateer (1989) observed that while useful in giving a general indication of a patient's cognitive and behavioral status, the actual details of the patient's functioning cannot be deduced from the patient's level. They further note that this scale implies similar rates of improvement on different kinds of functions, when this is more often not the case.



Name _____ Date of Test ⌊_____⌋ mo day yr

Age _____ Sex M F

Date of Birth ⌊_____⌋ mo day yr

Day of the week s m t w th f s

Time AM PM

Diagnosis _____ Date of injury ⌊_____⌋ mo day. yr

GALVESTON ORIENTATION & AMNESIA TEST (GOAT) Error Points

1. What is your name? (2) _____ When were you born? (4) _____ ⌊__⌋
   Where do you live? (4) _____
2. Where are you now? (5) city _____ (5) hospital _____ ⌊__⌋
   (unnecessary to state name of hospital)
3. On what date were you admitted to this hospital? (5) _____ ⌊__⌋
   How did you get here? (5) _____
4. What is the first event you can remember after the injury? (5) _____ ⌊__⌋
   Can you describe in detail (e.g., date, time, companions) the first event you can recall after injury? (6) _____
   _____
5. Can you describe the last event you recall before the accident? (5) _____ ⌊__⌋
   _____ Can you describe in detail (e.g., date, time, companions)
   the first event you can recall before the injury? (5) _____
6. What time is it now? _____(−1 for each ½ hour removed from correct time to maximum of −5) ⌊__⌋
7. What day of the week is it? _____(−1 for each day removed from correct one) ⌊__⌋
8. What day of the month is it? _____(−1 for each day removed from correct date to maximum of −5) ⌊__⌋
9. What is the month? _____(−5 for each month removed from correct one to maximum of −15) ⌊__⌋
10. What is the year? _____(−10 for each year removed from correct one to maximum of −30) ⌊__⌋

Total Error Points ⌊_____⌋

Total Goat Score (100-total error points) ⌊_____⌋

76-100 = NORMAL
66-75 = BORDERLINE
≤65 = IMPAIRED

ABNORMAL BORDERLINE NORMAL

100
90
80
70
60
50
40
30
20
10
0

Date _____

Time _____

FIGURE 18.2 Galveston Orientation and Amnesia Test (GOAT) record form.

Galveston Orientation and Amnesia Test (GOAT) (H.S. Levin, O'Donnell, and Grossman, 1979)

The GOAT is a short mental status examination devised to assess the extent and duration of confusion and amnesia following TBI (see Fig. 18.2). Like the GCS, it was designed for repeated measurements and can be used many times a day and repeated over days or weeks as necessary. Eight of the ten questions involve orientation for time, place, and person. The two questions asking for the first event the patient can remember "after injury" and the last event "before the accident" relate specifically to anterograde and retrograde amnesia, respectively. The error scoring system has a score range from −8 to 100. This test can serve two purposes. In light of the relationship between early return of orientation and good outcome—and its converse—it can serve as an outcome predictor. It also provides

a fairly sensitive indicator of level of responsivity in recently brain injured patients.

H.S. Levin and colleagues (1979) recommend that formal mental ability testing begin only after the patient achieves a GOAT score of 75 or better (within the "normal" range), i.e., when orientation is relatively intact. However, Hannay and Sherer (1996) found that most of their severely injured patients (at least 70%) could complete relatively simple tests (sentence comprehension; auditory and visual attention tasks; digit span) once their GOAT reached 40 but completion rates were lower for tests such as Trail Making A (50%) and B (29%). When these patients reach a GOAT of 40 (on the average at one month postinjury), a high percentage of them have recovered remote memory for personal information (name, date of birth, street address, and city) and the year, but not knowledge for events surrounding the injury or other items of temporal orientation (Hannay and Sherer, 1996). Levin and his coworkers noted that problems with amnesia are apt to persist after orientation has returned to normal. They suggested showing a calendar to aphasic and intubated patients when asking about temporal orientation. A preliminary study of the use of a multiple-choice GOAT with aphasic patients suggests that this response format results in a noticeably easier task for nonaphasic TBI patients and additionally, that the GOAT can underestimate the level of orientation and memory of aphasic TBI patients (Jain et al., 2000).

The cut-off score actually represents a level of orientation exceeded by 92% of a standardization sample of patients aged 16–50 with mild TBI (H.S. Levin, O'Donnell, and Grossman, 1979). This sample was chosen because it would control for demographic and personal characteristics that predispose one to closed head injury. Neurological examination was normal but 32% had a linear skull fracture and 24% had surgery for a depressed skull fracture. Interrater reliability in the original study was reported as .99, but it takes some training for examiners to be consistent in obtaining the information for and then scoring the amnesia items correctly.

GOAT measurements of posttraumatic amnesia (PTA) show strong associations with the severity of injury (GCS), and with a measure of long-term outcome (Glasgow Outcome Scale, GOS) (H.S. Levin, O'Donnell, and Grossman, 1979; Ellenberg et al., 1996), and the Disability Rating Scale and Functional Independence Measure (Zafonte, Mann, et al., 1997). This instrument's usefulness was supported by a study that found that only 52 of 102 head injury patients could estimate the duration of their PTA; and of these, only 30 of the 50 with mild injuries made this estimation (C.A. Bailey et al., 1984). However, those who made these estimations tended to be reasonably accurate as the correlation between GOAT data and patients' estimations was .85. The most typical sequence of reorientation is for person, place, and time, in that order (High et al., 1990). Eighty-eight percent of acutely hospitalized head injury patients showed a "backward displacement of the date," believing it was earlier than it actually was.

Oxford Test (Artiola i Fortuny, Briggs, Newcombe, et al., 1980); *Westmead PTA Scale* (Shores, Marosszeky, Sandaman, and Batchelor, 1986)

The Oxford Test for measuring the duration of PTA was probably the first quantitative test that involved formal testing of memory as well as a questionnaire about personal demographics (e.g., age, marital status, number of children, occupation), orientation in time and space, and last memories before the accident and first memories after the accident. Each day the patient is shown a different set of three colored pictures and asked to recall them or recognize them among a set containing five distractor items. The patient is also tested each day for recall and, if necessary, recognition of the examiner's first name and face ("Have you seen me before?"), using a photograph of the previous day's examiner when there is a change. Recognition of the examiner's name involves three names, two phonologically similar or with the same number of syllables as that of the examiner. A perfect score for three consecutive days signals the end of PTA on the first of the three days. The authors noted that this daily examination technique also identified mental status changes indicating deterioration in the patient's condition.

Success on the formal memory testing was as effective in determining the status of PTA as were the usual questions about personal history, orientation, and events surrounding the accident, in this case by neurosurgeons (Artiola i Fortuny et al., 1980). However, formal testing is less open to misinterpretation than questions such as the first event remembered after the injury. This procedure is recommended for research, especially multicenter trials in which the examiners at different centers have slightly different training and criteria for judging the correctness of the response to such questions.

The *Westmead Scale* (Shores et al., 1986) was based on the Oxford Test and provided a standardized set of procedures and a scoring form that tracked daily performance. The scale first asks seven questions about age, date of birth, month, time of day, day of week, year, and name of place, giving 1 point for each correct answer. A point is given for correct recall or recognition of the examiner's face and name and for each of the three pictures of objects, producing a total possible score of 12. As with the examiner's name, recognition of the face involves pictures of the original examiner

and two other faces. Recognition of objects involves six distractors, rather than the five of the Oxford Test. The same three object pictures are used every day until a perfect score of 12 is achieved. Thereafter, the object pictures are changed daily until the patient's recall is perfect for three consecutive days. This procedure was designed to ensure that new learning is taking place. PTA is judged to have ended on the first of three consecutive days for which the patient scores 12.

Patients who were in PTA on the Westmead, out of PTA, and orthopedic control subjects were given the Selective Reminding Test to assess learning and memory as part of the initial validation study (Shores et al., 1986). Patients still in PTA showed essentially no learning of the word list over trials while patients out of PTA learned but were still somewhat amnesic and exhibited poorer learning over trials than the orthopedic controls. The duration of PTA as measured by the Westmead was a significant predictor of severe TBI outcome in terms of verbal learning ($r = .44$) and nonverbal problem solving ($r = .37$), slightly better than the GCS on admission, and markedly better than duration of coma, which did not predict cognitive outcome (Shores, 1989). In another study, PTA duration predicted learning and memory on the Rey Auditory–Verbal Learning Test ($r = .34$), especially using a square root transformation of PTA duration ($r = .44$), and information processing speed measured by the Paced Auditory Serial Addition Test and Symbol Digit Modalities Test ($r = .29$), prediction again being better with a square root transformation ($r = .35$) (Haslam et al., 1994). With the exception of the GCS on admission and subarachnoid hemorrhage, no injury variables, including the nature of the trauma, hemorrhages, hematomas, or coma duration, were related to verbal learning or to information processing speed. Also, the duration of PTA minus coma duration (*postcoma disturbance* [*PCD*]) proved to be a good a predictor. A study with hospitalized children indicated that relatively few normal 6 to 7 year-olds (15%) met the Westmead's PTA criteria in four days of testing whereas over 90% of children in age groups from 8 to 15 did, suggesting that the adult Westmead procedure can be used with children over 7 years old (Marosszeky et al., 1993). Indices for consistency of "recovery" and duration to "recovery" have been developed for charting improvement of different components of orientation and memory (K. McFarland et al., 2001).

Choosing Outcome Measures

All health care professionals understandably would prefer to have brief measures of outcome that they can administer at bedside, in the office, or over the telephone. However, the focus of outcome evaluation of the TBI patient changes over time, especially for the severely injured patient who may start in a coma and must be evaluated by relatively simple measures that involve basic visual, verbal, and motor responses, such as the Glasgow Coma Scale (Teasdale and Jennett, 1974); and who later resumes a relatively normal life but continues to have some difficulties. Long-term follow-up measures for assessing the patient some time after return to the community will differ in their format and content from measures used when the patient leaves the acute care or rehabilitation hospital. Not only does the content change, but items included in tests used earlier can have ceiling effects. K.M. Hall, Bushnik, and their coworkers (2001) determined which of 10 outcome measures were useful for long-term follow-up (an average of five years postinjury), i.e., do not have marked ceiling effects. They found that the Functional Independence Measure memory item and the Functional Assessment Measure employment item (Uniform Data Systems, 1987, 1993; see also K.M. Hall, Hamilton, Gordon, et al., 1993), the Disability Rating Scale level of functioning and employability items (Rappaport, Hall, et al., 1982), all of the Neurobehavioral Functioning Inventory scales (depression, somatic difficulties, memory/attention, communication, aggression, motor) (Kreutzer, Marwitz, et al., 1996), the Patient Rating Competency Scale (Prigatano and Altman, 1990; Fordyce and Rouche, 1986), all Community Integration Questionnaire scales (home integration, social integration, productivity) (Willer, Rosenthal, and Kreutzer, 1993), and the Craig Handicap and Reporting Technique (Whiteneck, Charlifue, Gerhart, et al., 1992) cognition and occupation scales provided a useful range of scores across patients (defined as <25% of the data at any one score). Scores from the Glasgow Outcome Scale (Jennett and Bond, 1975), the Supervision Rating Scale (Boake, 1996b), and the Level of Cognitive Functioning Scale (Hagen, Malkmus, et al., 1979) did not have enough variability to be useful for the variety of outcomes that occur. Ideal measures will have good reliability and predictive validity and document motor, cognitive, psychosocial, and behavioral changes; strengths and weaknesses; ability to carry out activities in various environments; integration and participation in society; the environmental and personal factors that act as facilitators and hindrances; and quality of life (well-being and life satisfaction) at different times after injury. For these reasons, evaluation of TBI and other patients is likely to include some of the instruments discussed below at different times in the patient's course. Measures of severity and global measures are appropriate for assessing level of functioning when a TBI patient is in the acute and postacute stages,

and later on to measure changes (progress or deterioration). Measures of reintegration and participation in society, psychosocial adaptations, and quality of life are introduced later on. Representative measures from these somewhat different domains have been included here.

Outcome Evaluation

Global measures

The choice of a global measure of outcome for following the progress of a TBI patient as well as determining the effectiveness of treatments in randomized controlled trials and clinical research in general continues to be controversial. Much of the discussion focuses on the relative merits of the Glasgow Outcome Scale and the Disability Rating Scale (S.C. Choi et al., 1998; Contant et al., in press; Narayan, et al., 2002; Teasdale, Pettigrew, et al., 1998). Neither measure is particularly good at characterizing individual outcomes with less serious TBI or residual subtle impairments and disabilities.

Glasgow Outcome Scale (GOS) (Jennett and Bond, 1975; M.R. Bond, 1990)

This scale complements the Glasgow Coma Scale by providing criteria for evaluating the "goodness" of outcome. It has five levels: (1) *Death* (due to brain damage. This typically occurs within the first 48 hours after injury. It is rare that death after 48 hours, of persons who improved to an outcome level of 4 or 5, will be attributable to primary brain damage); (2) *Persistent vegetative state* (PVS) (absence of cortical function); (3) *Severe disability* (conscious but disabled; these patients are "dependent for daily support."); (4) *Moderate disability* (disabled but independent); (5) *Good recovery* (resumption of "normal life" is the criterion rather than return to work which, the authors noted, can be misleading when economic factors prevent an able person from finding employment or particularly favorable circumstances allow a relatively disabled person to earn money). Sometimes 1 is assigned to death and 5 to good outcome and sometimes the numbers have been assigned in the reverse direction. The clinician and researcher must be careful to find out which way the numbers are assigned to categories before interpreting a score of 2 as PVS or as moderate disability, a problem that has contributed to some misunderstandings in the literature (Contant et al., in press).

Although the GOS is attractive in its simplicity, this same quality makes it difficult to categorize many patients who are semidependent or independent. Interrater reliability is obviously not a problem for the Death and PVS categories. Valid ratings may not be obtained for the other categories if examiners do not ask appropriate questions of the patient, caretakers, or family [hjh]. Disagreements between raters are most likely to occur for the "moderate disability" rating (D.N. Brooks, Hosie, et al., 1986), which has been considered too inclusive (H.S. Levin, Benton, and Grossman, 1982) and too coarse-grained (Walsh, 1991) to provide more than suggestive categorization. Even with an expanded format (to eight categories, by adding an extra level each to the categories Severe, Moderate, and Good [Jennett, Snoek et al., 1981]), the *extended GOS (GOSE)* is insufficiently refined to accommodate the varieties and complexities of posttraumatic outcomes (*Lancet* Editors, 1986; B. [A.] Wilson, 1988). Moreover, an examination of interrater reliability indicated that agreement between experienced patient observers was considerably higher for the original five category scale (*Kappa* GOS = .77, *Kappa* GOSE = .48) (Maas et al., 1983). Intraobserver reliability was also better for the five category scale, but these higher *Kappa* values varied from .89 to .40 while those for the eight category scale were in the .82 to .22 range.

Structured interviews are now available for both the GOS and GOSE (J.T.L. Wilson, Pettigrew and Teasdale, 2000) with explicit criteria for categorizing individuals. The inclusion of a series of specific information-gathering questions and criteria for classifying patients should improve the agreement between ratings made by the different clinicians seeing these patients as well as the validity of the ratings. For example, data on agreement in the ratings made by a nurse and a psychologist produced a weighted *Kappa* of .89 for the GOS and .85 for the GOSE. J.T. Wilson, Edwards, et al. (2002) reported rating reliabilities a *Kappa* of .82 and .94 for the GOS, *Kappa* and repeated rating reliabilities of .89 for the GOS and .98 for the GOSE with a two week interval. They also compared the ratings obtained from a structured interview of patients conducted on the telephone by an experienced nurse and a postal version filled out by the patients about one week later with much lower agreement for the GOS than the GOSE. This is perhaps not surprising since TBI patients may be unaware of the severity of their difficulties or even that they have those difficulties.

Jennett and Bond (1975) advised that, "aspects of social outcome should be included . . . such as leisure activity and family relationships" in making outcome determinations. However, they did not offer a solution to the complex classification problem presented by so many patients whose level of social or emotional functioning is very different from the level of their cognitive skills, sensory-motor competence, or daily activi-

ties. Neither the GOS nor the GOSE has the gradation of scores necessary to provide information about the changes that take place within the severe, moderate, and good outcome categories. Especially, these scales cannot register the subtle deficits and changes that are experienced by less severely injured patients and that continue to interfere to some degree with many aspects of their lives, even though they appear to be doing well on the surface, having returned to work or school and looking after themselves independently.

Disability Rating Scale (DRS) (Rappaport, Hall, Hopkins, et al., 1982)

The DRS was designed to assess disability in severe TBI patients as they progress from coma back to the community (Rappaport, Hall, et al., 1982). It is not very sensitive to preinjury demographic variables (Hedrick et al., 1995). The total score ranges from 30 (death) to 0 (no disability) and represents the sum of scores for eight items (Table 18.6). The first three items are almost identical to the GCS and thus allow for the assessment of an individual with compromised consciousness. There are some important differences, however. While the best response for an item on the GCS is given the highest number, the same response on the DRS is given the lowest number. Also, the motor response item of the GCS ranges from 6 (obeying commands to 1 (none) while the same response on the DRS ranges from 0 (obeying commands) to 5 (none). Since the GCS is ordinarily determined just before the DRS by clinicians in the acute or subacute situation, it is important that they be careful in translating scores from the GCS to scores on similar items on the DRS. Furthermore, the verbal response is evaluated in a slightly different way on the DRS. An intubated patient or one

TABLE 18.6 Disability Rating Scale

Arousability, Awareness, and Responsibity		
Eye Opening	*Communication Ability (Verbal, Written, Letterboard or Sign)*	*Best Motor Response*
0 Spontaneous	0 Oriented	0 Obeying
1 To speech	1 Confused	1 Localizing
2 To pain	2 Inappropriate	2 Withdrawing
3 None	3 Incomprehensible	3 Flexing
	4 None	4 Extending
		5 None

Cognitive Ability for Self-Care Activities (Does patient know how and when? Ignore motor disability?)		
Feeding	*Toileting*	*Grooming*
0 Complete	0 Complete	0 Complete
1 Partial	1 Partial	1 Partial
2 Minimal	2 Minimal	2 Minimal
3 None	3 None	3 None

Level of Functioning (Consider Both Physical and Cognitive Disability)	*"Employability" (As a Full-time Worker, Homeworker, or Student)*
0 Completely independent	0 Not restricted
1 Independent in special environment	1 Selected job, competitive
1 Mildly dependent	2 Sheltered workshop, noncompetitive
2 Moderately dependent	3 Not employable
3 Markedly dependent	
4 Totally dependent	

Categorizations of Outcome Scores (Limitations, Severity)		
0 None	4–6 Moderate	17–21 Extremely severe
1 Mild	7–11 Moderately severe	22–24 Vegetative state
2–3 Partial	12–16 Severe	25–29 Extreme vegetative state
		30 Dead

From Rappaport, Hall, Hopkins, et al. (1982)

with a tracheotomy is given a score of 1 on the verbal response of the GCS but could earn any possible score for the analogous communication ability item of the DRS since a written, letter board, or sign response is credited. As the patient comes out of coma and begins to be able to complete basic activities of daily living, items evaluate the level of these abilities, ignoring motor disabilities but taking into account demonstrated knowledge of how and when. Dependence is assessed with the level of functioning item which considers both physical and cognitive ability to be independent. Finally, the employability item refers to functioning as a full-time worker, homemaker, or student depending on which is most appropriate to rate. More detailed information is provided for each of these items than was originally provided in the description of each GOS category. However, a list of appropriate questions to ask the patient, caretakers, and family members in order to obtain valid information would be helpful as well and would likely increase interrater reliability.

Test characteristics: comparing DRS and GOS. The DRS has some advantages over the GOS, in part because it has a range of scores for seven of the ten suggested levels of disability, with only death, mild, and none each represented by a single score. The range is particularly wide in the severe disability category. In contrast, the GOS has a single score for each category. The measure that has a finer gradation of scores is likely to provide better prediction (Contant et al., in press). The DRS has shown better predictive success than the GOS in a number of studies: predictions from acute care variables to outcome at three and six months postinjury (Struchen et al., 2001); prediction from outcome at discharge from acute care one, three, and six months postinjury to psychosocial outcome at six months (McCauley, Hannay and Swank, 2001); and change during rehabilitation (K. Hall, Cope, and Rappaport, 1985; Rappaport, Hall, et al., 1982). Neither measure is particularly good at depicting outcome in individuals with less serious TBI or residual subtle deficits (Pender and Fleminger, 1999).

Neuropsychological findings. DRS scores are correlated with auditory, visual, and somatosensory evoked potentials (Rappaport, Hemmerle and Rappaport, 1990, 1991; Rappaport, Herrero-Backe, et al., 1989). The DRS admission score in rehabilitation (Ponsford, Olver, et al., 1995; Cifu, et al., 1997; Gollaher et al., 1998) and the DRS discharge score (Cifu et al., 1997; Gollaher et al., 1998) are predictive of later employment. Interrater reliabilities of .97–.98 (Gouvier, Blanton, et al., 1987; Rappaport et al., 1982) and a test–retest reliability of .95 (Gouvier et al., 1987) have been reported. Predictive validity (Eliason and Topp, 1984; Gouvier et al., 1987) and concurrent validity (Gouvier et al., 1987; K. Hall, Cope, and Rappaport, 1985; K.M. Hall, Hamilton, et al., 1993) with other functional measures are high. DRS scores at six months after rehabilitation are strongly related to executive functioning and memory (Hanks, Rapport, et al., 1999). Anosmia occurs with longer coma, more neuropsychological deficits, and greater functional problems on the DRS (Callahan and Hinkebein, 1999).

Evaluation of the Psychosocial Consequences of Head Injury

An appreciation of the effects of TBI on personal and social adjustment and of their impact on family, friends, and the community has led a number of workers to develop schedules and scales for standardizing the examination and documentation of these problems. Some were designed as questionnaires for relatives, some as clinical rating scales, and for some the information is obtained from all possible sources. Although most of these scales were developed for research purposes but may be useful in the individual case for tracking the evolution of problems or their solutions, the Mayo-Portland Adaptability Inventory (see pp. 729–731) in particular was also developed for the individual case, to bring to light psychosocial issues that may be overlooked without the guidelines it provides [mdl]. Lacking comparative studies, no "best" scale or rating method has been identified, leaving examiners to decide which one(s) seems to suit their needs. The inventories reviewed here are among those most used with TBI and represent the variety of approaches to documenting these problems.

Katz Adjustment Scale: Relative's Form (KAS-R) (M.M. Katz and Lyerly, 1963)

The original purpose of this scale was the assessment of the personal, interpersonal, and social adjustment of psychiatric patients in the community, but much of it is appropriate for neuropsychologically impaired patients as well (e.g., Hanks, Rapport, et al., 1999; McSweeny et al., 1985). The issues this scale deals with are particularly relevant to TBI survivors living with their families or in noninstitutionalized settings. The authors' rationale for assessing the patient's adjustment from a relative's perspective is that "the patient's overall functioning is . . . intimately linked with the working out of mutually satisfactory relationships within the family." Additionally, the informant can provide an intimate view of the patient's day-to-day activities. Moreover, as is the case with psychiatric patients, some brain

impaired patients cannot respond reliably to a self-rating inventory or may be unable to cooperate with this kind of assessment at all. Thus the only way to get dependable information about them is through an informant. The questionable objectivity of a close relative led to the development of items concerning specific behaviors.

The scale consists of five inventories, or subscales, each designed to assess a different aspect of the patient's life or the relatives' perception of it. Form R1 asks for "Relatives Ratings of Patient Symptoms and Social Behavior." It includes 127 questions about such indicators of patient adjustment as sleep, fears, quality of speech, and preoccupations, for rating on a scale ranging from "1–almost never" to "4–almost always." Forms R2 and R3, "Level of Performance of Socially-expected Activities" and "Level of Expectations for Performance of Social Activities," use the same 16 items dealing with such ordinary activities as helping with household chores, going to parties, and working. Form R2 requires the informant to indicate the patient's level of activity for each item on a 3-point scale on which a rating of 1 is given for "not doing," 2 for "doing some," and 3 for "doing regularly." A 3-point scale is used for Form R3 too, but the rating criteria are reworded to include the informant's expectations of the patient, i.e., 1–"did not expect him [*sic*] to be doing," etc. The 22 items of Forms R4 and R5 have to do with how patients spend their free time. Like Forms R2 and R3, these two inventories share the same items which list specific leisure activities such as watching television, shopping, or playing cards, plus a 23rd item asking for activities to be listed that were not included in the previous items. These too are on 3-point scales. Form R4 asks for the frequency of activity (1–"frequently" to 3–"practically never"). R5 inquires about the relative's level of satisfaction with the patient's activities (1–"satisfied with what he does here" to 3–"would like to see him do less"). McSweeny and his colleagues added a "does not apply" response alternative to each scale except R1.

Test characteristics. Three major factors yielding 12 factor scales have been extracted from Form 1 (M.M. Katz and Lyerly, 1963; see Table 18.7). For both relatives and patients (using an appropriately reworded version of Form R1), with an eight week interval, test–retest correlations were significant (.65–.88) on three global factors (I. Social Obstreperousness, II. Acute Psychoticism, II. Withdrawn Depression), with the lowest correlations occurring on Factors II and III, which contain only 14 and 10 items, respectively (Ruff and Niemann, 1990).

TABLE 18.7 Item Clusters and Factors from Part 1 of the Katz Adjustment Scale

Item Clusters	*Factors*
Belligerence (BEL)	I. Social Obstreperousness
Verbal expansiveness (EXP)	
Negativism (NEG)	
General psychopathology (PSY)	
Anxiety (ANX)	II. Acute Psychoticism
Bizarreness (BIZ)	
Hyperactivity (HYP)	
Withdrawal (WDL)	III. Withdrawn Depression
Helplessness (HEL)	
Suspiciousness (SUS)	
Nervousness (NER)	
Confusion (CON)	
Stability (STA)	

Reprinted from Grant and Alves (1987)

Neuropsychological findings. The KAS-R can provide discriminating information about TBI patients, although not always on the same factor scales or to the same degree. Eight factor scales differentiated severely head-injured patients from patients without TBI, identifying 75% and 96% of these patients, respectively (W.A. Goodman et al., 1988). TBI patients had higher average scores on those factor scales that did not discriminate statistically between the two groups: Belligerence, Negativism, Bizarreness, and Hyperactivity (Fordyce et al., 1983). A group of TBI patients whose average injury duration was 25 months had significantly higher scores than recently injured patients (≤6 months) on the Belligerence score, along with higher scores on the Withdrawal and Retardation and the General Psychopathology scales. Hinkeldey and Corrigan (1990) found a similar pattern of abnormal ratings for patients one to five years postinjury. Looking at patients two to four years postinjury, P.S. Klonoff, Snow, and Costa (1986) also found Belligerence—and Negativism—among others, to be significant problem areas as reported by relatives of TBI patients. Klonoff and her colleagues calculated a dissatisfaction index (R3 − R2) which characterized these patients' relatives' responses although responses on KAS-R forms R2 to R5 did not, in themselves, differ significantly from age-graded norms. The form R1 scales that correlated significantly with employment status were Belligerence (−.22), Verbal expansiveness (0.29), Helplessness (−.21), and Confusion (−.19) (Stambrook, Moore, et al., 1990). Of these, Belligerence contributed significantly to a step-wise equation for predicting vocational status. Form R2, in which social performance is re-

ported, had the highest correlation (.30) with ratios of employment status. Hanks, Temkin, et al. (1999) obtained data on the KAS for 157 TBI (78% mild) and 125 general trauma controls. At one year postinjury, the TBI group reported many adjustment problems, typical of TBI, compared to the normative sample (M.M. Katz and Lylerly, 1963), but they did not differ from the trauma control group. Moderate TBI patients had more problems than those with mild or severe TBI. Within the TBI group, cognitive clarity, dysphoric mood, and emotional stability improved while anger management, antisocial behaviors, and self-monitoring worsened. Pender and Fleminger (1999) consider the KAS-R gives more information on post TBI personality change than on outcome.

Revisions of the KAS-R. H.F. Jackson and his coworkers (1992) modified the KAS-R so that two ratings were made for each item: how the injured persons were before the injury and how they are now (*KAS-R1*). An analysis of change scores generated 30 first order and seven second order factors. Classification of TBI patients with varying degrees of severity of injury and of spinal cord patients proved to be more accurate with their factors (60.9%) than with those of the authors (47.2%). Goran and Fabiano (1993) removed redundant items from the KAS-R1 and those not contributing to the stability of previously established psychological factors. The remaining 79 items had internal consistency *alpha* values of .75 to .93 for the component groups. With the exception of the components—Belligerence, Verbal Expansiveness, and Emotional Sensitivity—internal consistency was the same or better for relatives of TBI patients. More research is needed to establish the usefulness of this revision (see also Pender and Fleminger, 1999).

The Mayo-Portland Adaptability Inventory (MPAI) (Lezak and Malec, 2003)[1]

The MPAI is a revision and elaboration of the *Portland Adaptability Inventory* (*PAI*) (Lezak, 1987b; Lezak and O'Brien, 1988, 1990), developed to increase the sensitivity of its parent inventory. This set of three subscales was constructed to provide a systematic record of the personal and social maladaptations that tend to prevent many patients with *acquired brain injuries* (*ABI*) from resuming normal family relationships and social activities. While the MPAI retains the three-subscale format of the PAI, subscale names and contents differ somewhat from the original inventory but include the 24 PAI items. The original items were reworded as necessary to ensure that all ratings were made on the basis of current functioning. Now the 29 items (of which one, #28, comes in two parts) make up the three subscales (*Ability, Adjustment,* and *Participation;* see Table 18.8). Six additonal items which ask about "Preexisting and associated conditions" (e.g., drug and alcohol use) do not enter into the scoring or statistical evaluations of the MPAI. The *Manual for the Mayo-Portland Adaptability Inventory* (Malec and Lezak, 2003) provides detailed scoring criteria for each item.

Items are rated on a 5-point scale, from 0 (e.g., Item 16, **Pain and headache**: "0—No significant pain reported" to "4—Pain complaints are totally or almost totally disabling"). Wording of the scale varies according to the issue under consideration, but most ratings follow the same pattern in which 1 indicates a mild

[1]Copies of the MPAI may be obtained from the web site for the Center for Outcome Measures in Brain Injury (www.tbims.org/combi/mpai) or from James F. Malec, Ph.D., PM&R-1D-St. Mary's, Mayo Clinic Rochester, MN 55905. The MPAI is in the public domain and may be copied freely.

TABLE 18.8 Mayo-Portland Adaptability Inventory (MPAI) Items by Subscales

Ability Index	*Adjustment Index*	*Participation Index*
Mobility	Anxiety	Initiation
Use of Hands	Depression	Social Contact
Vision	Irritability/Anger	Leisure/Recreational
Motor Speech	Pain and Headache	Self Care
Communication	Fatigue	Residence
Attention/Concentration	Sensitivity to Mild Symptoms	Transportation
Memory	Inappropriate Social Interaction	Work/School
Fund of Information	Impaired self-awareness	Money Management
Novel Problem-Solving	Family/Significant relationships	
Visuospatial Abilities		
Dizziness		

From Malec and Lezak (2003)

problem or condition that "does *not* interfere" with functioning; 2 indicates a mild problem that interferes "5% to 24% of the time;" and 3 is given for a "moderate" problem or condition that interferes "25% to 75% of the time." For some items, the 5-point scale is worded to be parallel to the "% of time" scaling (e.g., Item 28a, **Paid employment**: 0-Full time [>30 hrs/wk], 1-Part-time [3–30 hrs/wk] without support; 2-Full-time or part-time with support; 3-Sheltered work; 4-Unemployed; employed < 3 hrs/wk).

When the MPAI is given to patients or personal associates (significant others [SO])—usually a spouse, partner, parent, or adult child—a clinical staff person should review the guidelines with the rater and be availale for questions. Patients with severe cognitive deficits should give MPAI ratings only with a staff person writing in the responses. Clinical staff ratings can provide information on patient progress. Ratings by patients and their families can alert clinicians to specific problems and achievements. As an outcome measure, the MPAI covers the full range of issues relevant to patient functioning after rehabilitation and in the community.

Test characteristics. The MPAI has undergone multiple revisions, based on analyses of responses from several large samples. Two data sets formed the bases for evaluating MPAI (MPAI-4th revision). One was a national sample of 386 patients with acquired brain injuries (ABI) (M_{age} = 38 ± 12.4, 73% male, 88% ABI, 23% < 12 years education, 80% white, with a severity range of mild [5%], moderate [29%], severe [44%], and unknown [15%]). A Mayo sample consisted of 134 ABI patients (M_{age} = 39 ± 13.5, 61% male, 65% TBI, 18% <12 years' education, 92% white, with a severity range of mild [29%], moderate [12%], severe [44%], and unknown [15%]). Total raw scores can be converted to *T*-scores (M = 50, SD =10) by using tables for staff ratings from either of these samples. Subscale tables based on Mayo staff ratings are included in the manual as are *T*-score tables for both Total and subscale raw scores made by both the brain-injured patients and their significant other derived from the Mayo samples.

Subscale items were identified following Rasch analysis of previous (and very similar) versions of the MPAI (Malec, Moessner, et al., 2000), selected on a "rational" basis; i.e., items that corresponded to clinical experience (Malec and Lezak, 2003; see Table 18.8). Item reliability for a three-rater composite (patient, SO, staff) was .99 for a sample of 134 Mayo clinic ABI outpatients. For each subscale, item reliabilities derived from the National sample were .99 for Total, Ability, and Adjustment subscales, .98 for Participation. For each subscale index for the Mayo Sample of 134 ABI outpatients, the three-rater composite was .99 for Total, Ability, and Participation, .97 for Adjustment. On the Mayo sample, for the first 29 items, item agreement (±1 point) between all rating group pairs was ≥66% on all but one item (impaired self-awareness) and ≥70% on 20 items. Concurrent validity of staff responses to the MPAI was demonstrated in moderately high correlations with the Disability Rating Scale and the Rancho Scale (Malec and Thompson, 1994). Factor analysis demonstrated "an underlying unitary dimension representing outcome after TBI that includes indicators of ability, activity, and participation" (Malec, Kragness, et al., 2003). Principal components identified by factor analysis (Bohac et al., 1997) may be informative in interpreting the multifactorial structure of ABI outcome (see Malec and Lezak, 2003). However, for practical purposes, the strong internal consistency of the rational subscales (R_{xx} [*Alpha*] = .80 for Ability, .76 for Adjustment, .83 for Participation) recommends that subscale integrity be maintained. The considerable interdependence between capacity and function was reflected in some items correlating highly with two subscale indices (e.g., Self-care correlation with Participation was .61, with Ability it was .57).

Other MPAI versions. Rasch analysis of the MPAI refined prediction of outcome by removing items that did not contribute to the total score (Malec, Moessner, et al., 2000). This resulted in a *22-item MPAI* that had similar predictive validity to the 30-item MPAI. The MPAI-22 has since been shown to be sensitive to change in rehabilitation and prediction from preadmission score to level of initial vocational placement and vocational status one year later (Malec, 2001; Malec, Buffington, Moessner, et al., 2000).

The *M2PI* is just the eight-item Participation subscale (Malec and Lezak, 2003). A series of correlations (mostly above .70) with different group evaluators (patients, SOs, and staff) and with the full-scale 3-Rater Composite Index suggest that it can be used as an outcome measure. Its brevity requires minimal personal or telephone contact thus lending itself to treatment follow-up or research programs.

A French version, *Inventaire d'Adaptabilité Sociale de Mayo-Portland,* is being developed in collaboration with Drs. Pierre North and Jean-Michel Mazaux. In a preliminary study involving 15 young (ages 21–36) rehabilitation patients with severe TBI, MPAI scores identified as significant problems with fatigue, dizziness, attention and concentration, recall of old information, problem solving, anxiety and irritability, return to work or school, social contact, and participation in leisure activities (Selmaoui, 2002). A comparison with the Neurobehavioral Rating Scale-R (French version) showed that these two instruments appear to be "com-

plementary, the NRS-R looking mostly at impairments . . . the MPAI looks mostly at cognitive and behavioral disability and handicaps" (J.-M. Mazaux, personal communication, June, 2003). Specific differences between these scales were in "fatigue" which in the MPAI referred to physical fatigue, in the NRS-R to "mental fatigability;" planning capacity was not examined in this French version; variables concerning work, social contact, and leisure were not examined in the NRS-R. The author concludes that the clinical utility of the MPAI resides in its "global evaluation of the diversity of problems—physical, cognitive, emotional, behavior, and social of TBI patients" (trans., mdl).

Neurobehavioral Rating Scale (NRS) (H.S. Levin, High, Goethe, et al., 1987; see also I. Grant and Alves, 1987)

This 27-item modification of the *Brief Psychiatric Rating Scale* (*BPRS*) (Overall and Gorham, 1962) was developed specifically for TBI patients. Its use requires a trained examiner to follow detailed guidelines (given in H. S. Levin, Overall, et al., 1984). BPRS items more appropriate for a psychiatric population were dropped (e.g., mannerisms and posturing, grandiosity), and others particularly relevant to head injury were added (e.g., Inaccurate Insight, Poor Planning, Decreased Initiation/motivation). Like its parent instrument, ratings are made on a 7-point scale from "not present" to "extremely severe." The format allows for profiles to be drawn for each patient, for groups or group comparisons, or for a single patient over time. Unfortunately, the items are listed in what appears to be a random order (e.g., 16. Suspiciousness; 17. Fatigability; 18. Hallucinating Behavior; 19. Motor Retardation, etc.) so that commonalities between these characteristics and symptoms cannot be grasped at a glance. Some of the items are based on a short interview while the rest are derived from patient observation during the interview and formal examination. It would be best to complete this scale after the examination. The NRS has proved to be useful in studies of Alzheimer's disease (Sultzer et al., 2003; B.G. Pollock et al., 2002; Harwood, Sultzer, and Wheatley, 2000). Dombovy and Olek (1996) included items of the NRS in a telephone follow-up procedure involving an interview with the caregiver, a cost-effective way of determining the status of many TBI survivors.

Interrater reliability examined with two pairs of observers rating either 43 or 34 patients proved to be high in an initial study (r = .90, .88, respectively) (H.S. Levin, High, Goethe, et al., 1987). A replication of this study involving 44 TBI patients produced an interrater reliability coefficient of .78; a repeated evaluation of 37 of these patients one week later found a similar level of interrater reliability (r = .76) (Corrigan, Dickerson, et al., 1990). Four factors emerged on analysis of a group of patients examined at different times postinjury and with different severity levels: I. Cognition/Energy, II. Metacognition, III. Somatic/Anxiety, IV. Language. Five items either loaded on more than one factor (Inattention/Reduced Alertness and Decreased Initiative) or did not load on any (Guilt, Hallucinations, Lability of Mood) (H.S. Levin, High, Goethe, et al., 1987). The item cluster of Factors II and IV differentiated the mildly injured groups from patients with moderate and severe injuries but not the latter two groups. Factor I items differentiated only mildly from severely impaired patients. The Cognitive/Energy item was a predictor of social outcome in Vilkki, Ahola, et al. (1994). Pender and Fleminger (1999) recommend the NRS as "probably the standard" measure against which all newcomers to behavior change scale development should be compared.

The *Neurobehavioral Rating Scale–Revised* (*NRS-R*) (H.S. Levin, Mazaux, et al., 1990) was developed to increase reliability and content validity. Several changes were made (H.S. Levin et al., 1990; McCauley, Levin, et al., 2001). Items on difficulty with mental flexibility and irritability were added; tension and anxiety were merged into one item; and inattention became reduced alertness and attention. The Likert rating scale was reduced to four categories (absent, mild, moderate, and severe). Answers to a structured interview of 15–20 min provide rating for about two-thirds of the items; one-third are based on examiner observations, unlike the GOS and DRS which use all available information.

Interrater reliability by item on data from 70 patients ranged from a *Kappa* of .22 for difficulty in planning to .77 for memory difficulties (median *Kappa* = .40). Factorial validity of NRS-R data on 286 TBI patients assessed at least one month (mild) or three months (moderate and severe) postinjury produced five factors: Intentional Behavior, Emotional State, Survival-Oriented Behavior/Emotional State, Arousal State, and Language (Vanier, Mazaux, et al., 2000). Interrater reliability for the factor scores was reasonable (.56 to .81). Associations of factor scores with GCS and coma duration, while significant in many cases, were fairly low (.12–.33). An exploratory factor analysis (McCauley, Levin, et al., 2001) of data from 210 moderate or severe TBI patients six months postinjury identified five factors: Executive/Organization, Positive Symptoms, Negative Symptoms, Mood/Affect, and Oral/Motor. These factors had good internal consistency (.62–.88) and modest but significant correlations with GCS scores (.17–.24) once again. Correlations of the Executive/Organization and Oral/Motor factors

and to some degree Mood/Affect and Negative Symptoms with several domains of neuropsychological functioning (verbal and visual memory, speed dependent visuomotor tracking, manual dexterity, and speeded language production were significant (.24–.70). The NRS-R total score correlated at .72 with the GOS and at .74 with the DRS at 6 months postinjury. A principal components analysis (Rapoport, McCauley, et al., 2002) of three month follow-up data from 115 mild/moderate patients from Toronto and the 392 patients from the McCauley, Levin, et al. (2001) study produced three factors: Cognitive, Emotional, and Hyperarousal. Severity of injury was significantly related to NRS-R total score as was the three month GOS score. Postresuscitation GCS scores were significantly related to the cognitive factor (.47) and weakly to the hyperarousal factor (.27).

Community participation

Craig Handicap Assessment and Reporting Technique (CHART) (Whiteneck, Charlifue, Gerhart, et al., 1992)

The CHART was designed to quantify the extent of handicap (community participation). It assesses the six dimensions of handicap, now referred to as "participation" (World Health Organization, 1980, 2001): (1) Physical independence—ability to sustain a customarily effective independent existence; (2) Mobility—ability to move about effectively in surroundings; (3) Occupation—ability to occupy time in the manner customary to that person's age, gender, and culture; (4) Social integration—ability to participate in and maintain customary social relationships; (5) Economic self-sufficiency—ability to sustain customary socioeconomic activity and independence. The original CHART consisted of 27 items but the addition of dimension (6) "Cognitive independence" brought it to 32 items (Mellick et al., 1999). This dimension, involving ability to orient in relation to surroundings, was not included in the original version because it was considered difficult to quantify.

The CHART assesses each dimension based on reports of how the individual functions from day to day. While the GOS estimates the capacity to work, for instance, the CHART directly asks how many hours a week the individual works. Each dimension is scored from 0–100, with 100 representing no handicap compared to a sample of able-bodied individuals. Ponsford, Olver, Nelms, and their colleagues (1999) find the CHART useful with TBI patients but have dropped the "Economic self-sufficiency" items because some patients find them intrusive, and further because this scale is not informative when patients receive substantial benefits (as in Australia). The CHART was designed as an interview that can be done in person or by telephone and takes about 15 min to give. The CHART was developed for use with spinal cord–injured (SCI) individuals but has been applied to TBI survivors. It is used when the individual is in the community, not in a hospital, since it is a measure of participation in the community.

The CHART was originally normed on 88 able-bodied individuals and 100 spinal cord injured (SCI) persons (Whiteneck, Charlifue, et al., 1992). For 135 SCI patients, one week test–retest reliability was .93 overall with the coefficients ranging from .80 for economic self-sufficiency to .95 for mobility. Patient–SO agreement ranged from .84 for mobility to .28 for social integration (total score agreement of .84). The latter coefficient rose to .57 when only patients with spouses were considered, presumably because the spouse was more knowledgeable of this aspect. (Patient–proxy agreement for the CHART total score was .70 in Cusick et al., 2001.) CHART scores for subgroups rated as having low or high handicap by rehabilitation professionals differed significantly, providing an indication of its validity. Rausch analysis of CHART items produced 11 handicap strata with a .99 item separation reliability. Corrigan, Smith-Knapp, and Granger (1998) found that inpatient rehabilitation discharge scores on the CHART were moderately predictive of CHART scores (.45) over a 5-year period as opposed to the Uniform Data Systems' (1987) Functional Independence Measure motor (.77) and cognitive (.69) scores (which were less likely to change over time as might be expected since they mainly refer to physical status).

With TBI patients, Boake and High (1996) compared the association of CHART scales and DRS and GOS scores to four outcome indicators (self-care independence, travel, employment, and friendship). CHART physical independence and DRS and GOS scores were strongly related to self-care. CHART scales have a strong association only to the related outcome indicator, although it was less for mobility and travel (.22) than for physical independence and self-care (.43), occupation and employment (.33), and social integration and friendship (.32). DRS and GOS scores had strong associations only with self-care and travel. O'Neill, Hibbard, Brown, et al. (1998) found that TBI patients' levels of employment, education, marital status, and sex were related to social integration scores on the CHART one year postinjury. C.A. Curran and colleagues (2000) found orthopedic and TBI patients with serious injuries to be similar in physical independence, mobility, occupation, and social integration but the TBI patients had significantly lower cognition scores. These groups also had similar depression, state, and trait anxiety scores. In general, higher depression and trait anxiety were

associated with lower mobility and cognition scores and, to a lesser degree, with lower occupation and social integration scores.

Community Integration Questionnaire (CIQ) (Willer, Rosenthal, Kreutzer, et al., 1993)

The CIQ was specifically designed as a telephone interview to evaluate community integration in TBI survivors. The CIQ consists of 15 questions that assess Home integration (H), Social integration (S), and Productive activities (P). Six questions have a 3-point scale, ranging from "doing the activity yourself alone" to "yourself and someone else" to "someone else." Six questions have a 3-point scale for times per month from "5 or more," "1–4 times" or "never." The remaining three items have individualized ratings. The total score range is 0 to 29 for maximum integration. The patient can also give written responses to the CIQ, although help may be needed; a significant other can complete it if necessary. Normative data for various demographic groups are needed. A revised CIQ-2 is in development.

Test characteristics. The authors' initial small study ($n = 16$) with a 10-day interval produced a test–retest reliability of .91 for patients and .97 for SO assessment of the patient. The same study measured concurrent validity with the CHART and CIQ. CHART Occupation was significantly related to CIQ Productive activities (for the patient, $r = .66$; for the SO, $r = .75$), as might be expected since they involve the same domains. The Social integration scale was not significantly related, perhaps because of the CHART's low ceiling in this area. Patient–SO agreement was evaluated by Sander, Seel, et al. (1997), who reported *Kappa* coefficients of .42 (shopping) to .94 (school) on the 15 items for 122 patients with a range of injury severity. The Home integration scale produced differences that were attributable to two items, "meal preparation" and "housekeeping," the patients rating themselves higher than did SOs. Agreement was lower for an earlier study of 148 TBI patients and SOs using the intraclass coefficient (Tepper et al., 1996). Acceptable internal consistency has been reported (Corrigan and Deming, 1995; Willer, Ottenbacher, and Coad, 1994). Factor analysis on data from 312 patients with primarily severe TBI found the same three factors (H, S, P) but two items were moved: "financial management" from Social integration to Home integration and "travel" from Productive activities to Social integration (Sander, Fuchs, et al., 1999). This study also established concurrent validity as CIQ total score and scale scores had significant correlations with DRS level of functioning (.25–.47) and employability (.37–.58), Uniform Data Systems' (1987) Functional Assessment Measure Community access (.27–.47) and Employability (.41–.60) scales, and Functional Independence Measure Social interaction (.24–.34) scale. Questions about the distribution of CIQ scores have arisen (Corrigan and Deming, 1995) and not resolved satisfactorily (Willer, Ottenbacher, and Coad, 1994).

Neuropsychological findings. Patients with more severe injuries have lower CIQ scores (Colantonio, Dawson, and McLellan, 1998). CIQ scores are related to premorbid factors, severity of injury, disability level, and cognition (J. Fleming et al., 1999; C.P. Kaplan, 2001; Novack et al., 2001; Rosenthal, Dijkers, et al., 1996) as well as measures of executive functioning and verbal memory (Hanks, Rapport, et al., 1999) and depression (H.S. Levin, Brown, et al., 2001). The Trail Making Test and Rey's Auditory–Verbal Learning Test predicted outcome on the CIQ (S.R. Ross, Millis, and Rosenthal, 1997). TBI patients' communication problems appear in the CIQ's numerous aspects of discourse related to social integration (Galski, Tompkins, and Johnston, 1998). The CIQ is sensitive to time of initiation for treatment (Seale et al., 2002). Some change over time in CIQ scores has been noted by K.M. Hall, Mann, et al. (1996) and Corrigan, Smith-Knapp, and Granger (1998).

Environmental factors

Craig Hospital Inventory of Environmental Factors (CHIEF), CHIEF Manual (Craig Hospital Research Department, 2001)

The CHIEF was developed to assess the frequency and magnitude of perceived *barriers/hindrances* that interfere with the lives of disabled individuals. The 25 questions cover five domains: Physical and structural (e.g., design and layout of buildings, temperature, terrain, noise), Work and school (e.g., availability of education and training, format of material, special adapted devices), Attitudes and support (e.g., community attitudes towards disabled persons, encouragement or support at school or work), Services and assistance (e.g., programs and services in the community), and Policies (in government, education, and employment). Frequency of a problem is rated on a 5-point scale from "never" to "daily;" magnitude is rated as "little" or "big problem." Only the patient is supposed to respond to the CHIEF, not a significant other. It takes about ten minutes, can be self-administered or done as an interview in person or by telephone.

A sample of 409 disabled individuals (124 patients with SCI, 120 with TBI, 165 with other disabilities) was recruited for a validation study of the psychome-

tric characteristics of the test. A two-week test–retest reliability study with a subset of 103 participants found an internal consistency correlation of .93 for the TBI group (n = 44). Family members or friends of 125 individuals not included in the test–retest reliability study completed the CHIEF in order to determine patient–SO agreement. The TBI sample's internal consistency correlation was .59 for barrier frequency and .72 for magnitude for 54 subject pairs. Factor analysis generated the five factors given above. Differences in frequency and magnitude of environmental barriers between groups with various impairments and activity limitations are reported in the manual as well as norms for disabled, non-disabled, SCI, TBI, and other diagnoses. A CHIEF short form of 12 items has been created with norms also in the manual. It remains to be seen how useful this measure of environmental barriers will be. The study of how the environment can affect outcome is a newly developing area which should see increasing development in the future.

EPILEPSY PATIENT EVALUATIONS

Scales and inventories for documenting the behavior of epilepsy patients have been used for two quite different purposes. One has been to document the behavioral and psychosocial consequences of epilepsy surgery. The other is for behavioral description, often in evaluating outcomes of clinical drug trials (Kline Leidy et al., 1998). Although some studies have used instruments from the general psychometric repertoire (e.g., R. Martin, Meador, et al., 2001), specialized questionnaires and scales have been developed specifically for this population. A brief survey of different representative instruments is presented below.

A-B Neuropsychological Assessment Schedule (ABNAS) (Aldenkamp, Baker, Pieters, et al., 1995)

This self-administered measure, previously called the Neurotoxicity Scale (Aldenkamp, Baker, Pieters, et al., 1995), enables patients to report on the adverse effects of antiepileptic drugs on cognition. The 24 questions are rated from 0 (no problem) to 3 (a serious problem). The inventory was originally validated on healthy control subjects taking a benzodiazepine and endorsing items relating to "fatigue and slowing" (Aldenkamp et al., 1995). "Fatigue and slowing" was also the dominant area endorsed by patients with poorly controlled epilepsy but this finding was unrelated to seizure frequency, drug dosing (high vs. low), or monotherapy vs. polytherapy (Aldenkamp and Baker, 1997). In general, the global ABNAS score is considered to be the primary variable reflecting perceived cognitive effects, with excellent reliability (Cronbach's *alpha* = .96) (J. Brooks et al., 2001).

Epilepsy Foundation of America (EFA) Concerns Index (Gilliam, Kuzniecky, Faught, et al., 1997)

This scale was developed by asking patients with chronic epilepsy to list in order of importance their concerns about living with recurrent seizures. Twenty questions assess different domains including driving, autonomy, work, education, family, seizure effects, medication effects, mood and anxiety, and social activities. Ratings are made on a 5-point scale, then summed to yield an overall *Concerns Index* which ranges from 20 to 100. Cronbach's *alpha* was .94, indicating a highly reliable instrument.

For patients who had previously undergone surgery for poorly controlled epilepsy, reponses regarding mood, employment, driving, and antiepileptic drug cessation were related to quality of life perception (Gilliam, Kuzniecky, Meador, et al., 1999). In contrast to studies using the Quality of Life in Epilepsy questionnaire, seizure freedom was not a predictor of postoperative quality of life. The EFA Concerns Index provides disease-specific quality of life information that is complementary to that obtained using more generic health related quality of life scales (Viikinsalo et al., 1997).

Liverpool Assessment Battery (G.A. Baker, Smith, et al., 1993)

This battery of measures assesses health related quality of life in epilepsy using eight different instruments of which four predate this battery and have been used elsewhere with different kinds of groups. The four developed by the test authors are the *Seizure Severity–PERCEPT* and *Seizure Severity–ICTAL* scales which ask for the patient's perception of the physical characteristics of seizure severity (G.A. Baker, Smith, et al., 1991); an *Adverse Events Profile* enquiring about medication side effects; and *The Impact of Epilepsy* scales concerning the social aspects of epilepsy and treatment on everyday functioning (Jacoby et al., 1993). Mood and other psychological factors are examined with the *Affect Balance Scale* (Bradburn, 1969) and the *Hospital Anxiety and Depression Scale* (Zigmond and Snaith, 1983), both of which are independently established tests. Coping ability is tested with the *Rosenberg Self-Esteem Scale (SES)* (M. Rosenberg, 1965) and the *Mastery Scale* (Pearlin and Schooler, 1978), which is designed to measure the degree to which patients feel in control of their own life as op-

posed to being fatalistically determined. The *Impact of Epilepsy* scales emphasize the social aspects of epilepsy and treatment on everyday functioning (Jacoby et al., 1993). Portions of this battery have been reported in different combinations in the literature, both in clinical drug trials (G.A. Baker, Smith, et al., 1993) and patient studies (Jacoby et al., 1993; Kellett et al., 1997).

Quality of Life in Epilepsy (QOLIE) (Devinsky et al., 1995)

This questionnaire was developed using the Epilepsy Surgery Inventory as its base (Vickrey et al., 1992), which itself includes the Rand Study 36-item Healthy Survey (Ware and Sherbourne, 1992), with additional specific epilepsy related questions. Thus, it follows the current practice for quality of life measures to use a generic instrument with disease-specific additions (G.A. Baker, 2001). In addition to assessing general quality of life, the QOLIE includes epilepsy specific domains: attention, concentration, memory, seizure worry, medication effects, and work and driving limitations.

The three versions of the QOLIE differ in length: The 89-item version containing 17 scales is intended primarily for research; the 31-item test is applicable to either research or clinical evaluations; the ten-item scale is intended for clinical practice. Although copyrighted, all versions of the test are available without charge. For the 89-item version, reliability coefficients using Cronbach's *alpha* for the 17 scales ranged from .78 to .92, with test–retest reliabilities from .58 to .86. The only scales below $r = .70$ were the two involving role limitation: pain and medication effects. Intraclass correlations ranged from .58 to .85.

To determine the magnitude of change needed to infer improved quality of life, QOLIE scores were compared to patient ratings; a 10.1 point change was required for the QOLIE-89 and an 11.8 point change for the QOLIE-31 (Wiebe, Matijevic, et al., 2002). Moreover, both measures discriminated medium from large changes in quality of life. In a surgical population, patients who became seizure free reported higher QOLIE scores (31 and 89 forms) than those who did not (Birbeck et al., 2002; Markand et al., 2000). The QOLIE-89 can be reliably administered by telephone (Leidy et al., 1999).

Side Effect and Life Satisfaction (SEALS) (Gillham, Baker, et al., 1996)

The SEALS inventory is a 38-item, questionnaire for patients designed to measure satisfaction with medications for seizure control. Questions ask for responses based upon feelings and behavior experienced during the previous week. Answers are placed on a 4-point Likert scale ranging from 0 (never) to 3 (many times). The questionnaire yields five summary measures—worry, temper, cognition, dysphoria, and tiredness—in addition to an overall SEALS score. The SEALS appears sensitive to differential cognitive side effects of drugs; e.g., patients taking carbamazepine had more side effects than those taking lamotrigine, which led to greater patient dropout on the former medication (Gillham, Kane, et al., 2000). In a validation study comparing responses of 307 patients with poorly controlled seizures on SEALS and on two scales measuring emotional status and one for cognitive functioning, significant correlations ranging from .51 to .84 were present for all SEALS scores with the other questionnaires. The authors concluded that this is a valid test for both clinical investigations of antiepileptic drugs and long-term epilepsy management (Gillham, Bryant-Comstock, and Kane, 2000).

Washington Psychosocial Seizure Inventory (WPSI) (Dodrill, 1986; Dodrill, Batzel, et al., 1980)

This 132-item True–False patient questionnaire was developed to document social maladaptations that tend to be associated with chronic epilepsy. The seven psychosocial scales relate closely to important aspects of the patient's life: Family Background (primarily pertaining to family and predisposing influences), Emotional Adjustment; Interpersonal Adjustment; Vocational Adjustment; Financial Status; Adjustment to Seizures; and Medicine and Medical Management. Using responses by 100 adult seizure patients, these scales were based upon item relationships with professional ratings (Dodrill, Batzel, et al., 1980). Higher scores indicate more problems.

Reliability coefficients were calculated for each scale and for an "Overall Psychosocial Functioning" scale, which includes some of the items contributing to other scales (Dodrill, Batzel et al., 1980). On 30-day follow-up, test–retest reliability coefficients were in the .66 to .87 range, split-half reliabilities ranged from .68 to .95; "Medicine and Medical Management" had the lowest correlations. Responses were evaluated by comparing them with ratings made by significant others and by professional examiners. The highest correlations between ratings and scale scores appeared for the Vocational scale ($r = .69$ with significant others' ratings, $r = .74$ with professional examiners' ratings); the lowest (.11, .33, for significant others and professional examiners, respectively) were on the "Adjustment to Seizures" scale.

Higher WPSI scores were associated with poorer neuropsychological test performance (Dodrill, 1986).

Seizure patients had significantly higher scores on the Emotional Adjustment scale than control subjects and also reported a great deal of difficulty adjusting to their illness (Tan, 1986). Invalid profiles were produced by approximately one-third of the epilepsy patients (24/68) and one-sixth of the control subjects (7/42), raising questions about the appropriateness of the validity measures. Moreover, of the normal control subjects whose inventory profiles were valid, 46% met the criterion for problems in "Emotional Adjustment," suggesting that this scale may not meet generally accepted standards for emotional disorders.

Using this scale, Trostle and colleagues (1989) found that community-dwelling people who were not seeking professional assistance for epilepsy-related problems obtained significantly lower scores on the WPSI than seizure patients seen in the clinic. The degree of psychosocial difficulty documented by the WPSI depends not only on patients' seizure frequency but also on the culture of their community (Swinkels et al., 2000). In patients undergoing surgery for poorly controlled seizures, better psychosocial functioning predicts better postoperative seizure control (Wheelock et al., 1998). Depression following anterior temporal lobectomy can be predicted, in part, by baseline WPSI emotional adjustment scores (Derry and Wiebe, 2000).

QUALITY OF LIFE

Satisfaction With Life Scale (SWLS) (Diener, Emmons, Larsen, et al., 1985)

Subjective well-being seems to have two components: an affective component (pleasant and unpleasant affect) and a cognitive component (life satisfaction) (Andrews and Withey, 1976; Corrigan, Bogner, et al., 2001; Pavot and Diener, 1993). The SWLS is designed to measure life satisfaction which has been defined as "a global assessment of a person's quality of life according to his chosen criteria" (Shin and Johnson, 1978, p. 478). Diener et al. (1985) suggested that life satisfaction derives from the individual's judgment of what is important, not what the examiner considers important. Even if two individuals value the same aspects of life (e.g., health, energy, finances), they may differ in their emphasis on them. On this basis, the authors developed a simple five-item scale that uses a Likert rating going from 1 (strongly disagree) to 7 (strongly agree) and results in a score from 5 (low satisfaction) to 35 (high satisfaction) (Table 18.9).

Normative data are available in many studies and include samples of American, French-Canadian, Russian, Chinese, and Korean groups; disabled college students; nurses and health workers; older Americans and French-Canadians; religious women (nuns); printing trade workers; military wives and nurses; VA inpatients; Dutch medical outpatients; abused women; clinical clients seeing psychologists (inpatients and outpatients); and elderly caregivers (Pavot and Diener, 1993). Concerns have been raised that self-report measures of well-being can be influenced by transient factors such as momentary mood, physical surroundings, and even the item that precedes a single-item measure of well-being and life satisfaction; but such effects have not been found for multi-item measures (Pavot, Diener, et al., 1991).

TABLE 18.9 Satisfaction With Life Scale (SWLS)

1. In most ways my life is close to my ideal
2. The conditions of my life are excellent
3. I am satisfied with my life
4. So far I have gotten the important things I want in life
5. If I could live my life over, I would change almost nothing

From Diener et al. (1985)

In an initial study of 176 undergraduates the mean score was 23.5 ± 6.43 (Diener et al., 1985). The two-month test–retest reliability for 76 students was .82, similar to the correlations of .89 reported for a two-week retest (Alfonso and Allison, 1996). Criterion validity was moderately strong as measured by correlations between SWLS and other measures of well-being and life satisfaction for samples of 176 and 163 undergraduates. Ratings of life satisfaction by 53 elderly individuals based on interview produced strong interrater reliability (.73). Internal consistency (item-total correlations) for the five items in the scale was also good (.61–.81). Others have reported substantial item-factor loadings (Arrindell et al., 1999). Factor analyses of the SWLS consistently produce a single factor accounting for over 60% of the variance (cf. Pavot and Diener, 1993; Arrindell et al., 1999).

Many different variables relate to SWLS scores (e.g., sex, marital status, health, and such personality variables as self-esteem, euphoria, dysphoria, and neuroticism) (Arrindell et al., 1999). Higher life satisfaction at one and two years postinjury has been associated with not having a preinjury history of substance abuse, having gainful employment, and a higher GCS score in 218 TBI patients (Corrigan, Bogner, et al., 2001). At one year it was associated with trauma admission GCS score and at two years, with depressed mood and social integration. Life satisfaction was relatively stable for two years, only changing significantly with marital status and depressed mood over time. Mean scores of 20.3 and 20.8 for the first and second years, respectively, represent a neutral rating in the scale. Bogner

and coworkers (2001) reported similar effects of substance abuse with telephone interviews of 168 TBI patients one year after injury. Much lower mean life satisfaction scores have been found for TBI patients with PTSD (12.88) than those without it (19.07) (Bryant, Marosszeky, et al., 2001). Lowered life satisfaction in spinal cord patients two years after injury was associated with being male, unemployed, having poor perceived health, decreased mobility, and decreased social integration (Putzke, Richards, et al., 2002).

PSYCHIATRIC SYMPTOMS

Brief Psychiatric Rating Scale (BPRS) (Overall and Gorham, 1962)

This 18-item instrument has enjoyed wide use with psychiatric disorders (e.g., Belanoff et al., 2002; Umbricht et al., 2002). Although the BPRS had been used with TBI patients (e.g., H.S. Levin and Grossman, 1978), the Neurobehavioral Rating Scale modification is usually preferred for these patients (see pp. 731–732). Each item of the BPRS represents a "relatively discrete symptom area"; most of the items were derived from psychiatric rating data. Ratings are made on a 7-point scale from "Not Present" to "Extremely Severe." The scale is intended for use by psychiatrists and psychologists. Although many of the items are more appropriate for a psychiatric population than for brain impaired patients (e.g., Guilt feelings, Grandiosity), there are also items involving symptoms that are prominent features of some neurological conditions (e.g., Motor retardation, Conceptual disorganization, Blunted affect). Others, although usually considered psychiatric symptoms, also appear in many patients with organic brain damage (e.g., Uncooperativeness, Depressive mood, Suspiciousness).

Interrater reliabilities have ranged from .67 to .75 (Hafkenscheid, 2000). Five factors were reported on ratings of a large number of schizophrenic patients: Anxiety–Depression, Anergia, Thought Disturbance, Activation, and Hostile–Suspiciousness (R.S. McDonald, 1986). A four factor model was reported for a group of recent-onset schizophrenics (Van der Does et al., 1993): Positive Symptoms, Negative Symptoms, Disorganization, and Depression described a group of recent-onset schizophrenics. A similar model identified factors derived from a sample of more chronic patients as Thought Disturbance, Anergia, Disorganization, and Affect (Mueser et al., 1997). A factor analysis for geropsychiatric inpatients came up with a somewhat different factor pattern: Withdrawn Depression, Agitation, Cognitive Dysfunction, Hostile-Suspiciousness, and Psychotic Distortion. This pattern was attributed to the prominence of "conceptual disorganization and disorientation" among these patients (McDonald, 1986). Conceptual Disorganization, Disorientation, and Motor Retardation were the most frequently scored items for severely and moderately TBI patients, while mildly injured patients received ratings within the normal range on these items (H.S. Levin, 1985). These scores differentiated each severity group from the others to a significant degree.