One distinguishing characteristic of neuropsychological assessment is its emphasis on the identification and measurement of psychological—cognitive and behavioral—deficits, for it is primarily in deficiencies and dysfunctional alterations of cognition, emotionality, and self-direction and management (i.e., executive functions) that brain disorders are manifested behaviorally. Neuropsychological assessment is also concerned with the documentation and description of preserved functions—the patient's behavioral competencies and strengths. In assessments focused on delineating neuropsychological dysfunction—whether for the purpose of making a diagnostic discrimination, evaluating legal competency or establishing a legal claim, identifying rehabilitation needs, or attempting to understand a patient's aberrant behavior—the examiner still has an obligation to patients and caregivers to identify and report preserved abilities and behavioral potentials.

Yet brain damage always implies behavioral impairment. Even when psychological changes after a brain injury or concomitant with brain disease are viewed as improvement rather than impairment, as when there is a welcome increase in sociability or relief from neuritic anxiety, a careful assessment will probably reveal an underlying loss.

A 47-year-old postal clerk with a bachelor's degree in education boasted of having recently become an "extrovert" after having been painfully shy most of his life. His wife brought him to the neuropsychologist with complaints of deteriorating judgment, childishness, untidiness, and negligent personal hygiene. The patient reported no notable behavioral changes other than his newfound ability to approach and talk with people.

On examination, although many cognitive functions tested at a superior level, in accord with his academic history and his wife's reports of his prior functioning, the patient performed poorly on tests involving immediate memory, new learning, and attention and concentration. The discrepancy between his best and poorest performances suggested that this patient had already sustained cognitive losses. A precociously developing Alzheimer-type dementia was suspected.

In some patients the loss, or deficit, may be subtle, becoming apparent only on complex judgmental tasks or under emotionally charged conditions. In others, behavioral evidence of impairment may be so slight or ill-defined as to be unobservable under ordinary conditions; only patient reports of vague, unaccustomed, frustrations or uneasiness suggest the possibility of an underlying brain disorder.

A 55-year-old dermatologist received a blow to the head when another skier swerved onto him, knocking him to the ground so hard that his helmet was smashed on the left side. Shortly thereafter he sought a neuropsychological consultation to help him decide about continuing to practice as he was exhausted easily, had minor memory lapses, and noticed concentration problems. This highly educated man gave lower than expected performances on tests of verbal abstraction (Similarities), visual judgment (Picture Completion), and verbal recall (story and list learning), and performances were significantly poorer than expected when structuring a drawing (R-O Complex Figure) and on visual recall. Additionally, subtle deficits appeared in word searching hesitations, several instances of loss of instructional set, tracking slips when concentrating on another task, and incidental learning problems which also suggested some slowed processing as delayed recall was considerably better than immediate recall. These lower than expected scores and occasionally bungled responses appeared to reflect mild acquired impairments which together were experienced as memory problems and mental inefficiency.

A year later, he requested a reexamination to confirm his impression that cognitive functioning had improved. He reported an active winter of skiing which validated his feeling that balance and reflexes were normal. However, he had noticed that he missed seeing some close-at-hand objects which—when pointed out—were in plain view and usually on his left side, but he reported no difficulty driving nor did he bump into things. He wondered whether he might have a visual attention problem. On testing, reasoning about visually presented material (Picture Completion) was now in the superior range although he had long response times, and verbal learning had improved to almost normal levels. Visual recall remained defective, but delayed visual recognition was within normal limits. However, on a visual scanning task (Woodcock-Johnson III-Cog [WJ-III Cog], Pair Cancellation), he made eight omission errors on the left side of the page and three on the right (see Fig. 10.1, p. 376). When last year's eight operation errors on printed calculation problems (Fig. 4.1, p. 87) were reviewed, it became apparent that left visuospatial inattention had obscured his awareness of the operation sign on the left of these problems, and that he continued to have a mild form of this problem. It was suspected that he had sustained a mild contra coup in the accident; mild because his acute self-awareness distinguished him from patients with large and/or deep right parietal lesions, contra coup because left visuospatial inattention implicates a right hemisphere lesion in a right-handed man.
Although the effects of brain disorders are rarely confined to a single behavioral dimension or functional system, the assessment of psychological deficit has focused on cognitive impairment for a number of reasons. First, some degree of cognitive impairment accompanies almost all brain dysfunction and is a diagnostically significant feature of many neurological disorders. Moreover, many of the common cognitive defects—aphasias, failures of judgment, lapses of memory, etc.—are likely to be noticed by the casual observer and to interfere most obviously with the patient's capacity to function independently.

In addition, psychologists are better able to measure cognitive activity than any other kind of behavior, except perhaps simple psychophysiological reactions and somatomotor responses. Certainly, cognitive behavior—typically as mental abilities, skills, or knowledge—has been systematically scrutinized more times in more permutations and combinations and with more replications and controls than has any other class of behavior. Out of all these data have evolved numerous reliable and well-standardized techniques for identifying, defining, grading, measuring, and comparing the spectrum of cognitive functioning. Intelligence testing and educational testing provide the neuropsychologist with a ready-made set of operations and a well-defined frame of reference that can be fruitfully applied to deficit measurement (Lezak, 1988c). The deficit measurement paradigm can be used with other behavioral impairments such as personality change, reduced mental efficiency, or defective executive functioning. However, personality measurement, particularly of brain-impaired individuals, has not yet achieved the community of agreement nor the levels of reliability or predictability that are now taken for granted when measuring cognitive functions. Furthermore, in clinical settings impairments in efficiency and executive functions are usually evaluated on the basis of their effect on specific cognitive activities or personality characteristics rather than studied in their own right.

In the following discussion, any mention of a test will refer only to individual tests, not batteries (such as the Wechsler Intelligence Scales [WIS]) or even those test sets, such as Digits Forward and Digits Backward, that custom has led some to think of as a single test. This consideration of individual tests comes from demonstrations of the significant interest variability in patient performances, the strong association of different patterns of test performance with different kinds of brain pathology, the demographic and other factors which contribute to the normal range of intrindividual test score variations, and the specificity of the brain-behavior relationships underlying many cognitive functions (e.g., see Grant and Adams, 1996, *passim*; Mesulam, 2000b; Naugle, Cullum, and Bigler, 1998). This knowledge of intrindividual variations in test performances does not support the popular concept of "intelligence" as a global—or near-global—phenomenon which can be summed up in a single score (Ardila, 1999a), nor does it support summing scores on any two or more tests that measure different functions, such as combining the scores of the WIS for adults (WIS-A) Block Design test which involves abstract visual analysis and visuospatial conceptualization and WIS-A Picture Completion test which not only has

### Calculations

<table>
<thead>
<tr>
<th>25</th>
<th>172</th>
<th>249</th>
<th>750</th>
<th>6712</th>
<th>628</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>65</td>
<td>6148</td>
<td>419</td>
<td>459</td>
<td>413</td>
</tr>
<tr>
<td>+ 42</td>
<td>+ 39</td>
<td>+ 354</td>
<td>419</td>
<td>459</td>
<td>413</td>
</tr>
<tr>
<td>6°5</td>
<td>2.70</td>
<td>7.031</td>
<td>1.164</td>
<td>7.168</td>
<td>27</td>
</tr>
<tr>
<td>= 248</td>
<td>54</td>
<td>1370</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>62</th>
<th>713</th>
<th>472</th>
<th>928</th>
<th>3.56</th>
</tr>
</thead>
<tbody>
<tr>
<td>x 0.5</td>
<td>x 2.4</td>
<td>x 1.6</td>
<td>x 0.53</td>
<td>x 2.8</td>
</tr>
<tr>
<td>67</td>
<td>717</td>
<td>483</td>
<td>581</td>
<td>29.98</td>
</tr>
<tr>
<td>7°2</td>
<td>9.98</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.1** Calculations test errors (circled) made by a 55-year-old dermatologist with a centre coup from striking his head on the left. Note neglect of operation signs on subtraction and multiplication problems.
no visuospatial component and requires no manipulation by the subject but has a considerable verbal weighting and calls on the ability to draw upon acculturated experience, together with a third—also quite different test—Matrix Reasoning (The Psychological Corporation, 1997).

**COMPARISON STANDARDS FOR DEFICIT MEASUREMENT**

The concept of behavioral deficit presupposes some ideal, normal, or prior level of functioning against which the patient's performance may be measured. This level, the *comparison standard,* may be *normative* (derived from an appropriate population) or *individual* (derived from the patient's history or present characteristics), depending on the patient, the behavior being evaluated, and the assessment's purpose(s). Neuropsychological assessment uses both normative and individual comparison standards for measuring deficit, as appropriate for the function or activity being examined and the purpose of the examination. Examiners need to be aware of judgmental biases when estimating premorbid abilities (Kareken, 1997).

**Normative Comparison Standards**

*The population average*

The normative comparison standard may be an *average* or middle (*median*) score. For adults, the normative standard, or "norm," for many measurable psychological functions and characteristics is a score representing the average or median performance of some more or less well-defined population, such as white women or college graduates over 40. For many cognitive functions, variables of age and education or vocational achievement may significantly affect test performance. With test developers' growing sophistication, these variables are increasingly taken into account in establishing test norms for adults. The measurement of children's behavior is concerned with abilities and traits that change with age, so the normative standard may be the average age or grade at which a given trait or function appears or reaches some criterion level of performance (e.g., Binet et Simon, 1908). Because of the differential rate of development for boys and girls, children's norms are best given separately for each sex.

Since so many tests have been constructed for children in education and training programs, normative standards based on either average performance level or average age when performance competence first appears are available for a broad range of cognitive behaviors: from simple visuomotor reaction time or verbal mimicry to the most complex activities involving higher mathematics, visuospatial conceptualization, or sophisticated social judgments (Anastasi and Urbina, 1997; see, e.g., normative tables in Woodcock-Johnson III [Woodcock, McGrew, and Mather, 2001]). Norms based on averages or median scores have also been derived for social behaviors, such as frequency of church attendance or age for participation in team play; for vocational interests, such as medicine or truck driving; or for personality traits, such as assertiveness or hypochondria.

In *neuropsychological assessment,* *population norms* are most useful in evaluating basic cognitive functions that develop throughout childhood. They can be distinguished from complex mental abilities or academic skills when examined as relatively pure functions. Many tests of memory, perception, and attention and those involving motor skills fall into this category (e.g., see Dodrill, 1999; J.M. Williams, 1997). Typically, performances of these capacities do not distribute normally; i.e., the proportions and score ranges of persons receiving scores above and below the mean are not statistically similar as they are in normal distributions (e.g., see Benton, Sivan, Hamsher, et al., 1994; Johnstone, Slaughter, et al., 1997; Stuss, Stethem, and Pelchat, 1988). Moreover, the overall distribution of scores for these capacities tends to be skewed in the substandard direction as a few persons in any randomly selected sample can be expected to perform poorly, while nature has set an upper limit on such aspects of mental activity as processing speed and short-term storage capacity. Functions most suited to evaluation by population norms also tend to be age dependent, particularly from the middle adult years onward, necessitating the use of age-graded norms (Baltes and Graf, 1996; Lezak, 1987); and education too contributes to performance on these tests (e.g., Misrshina, Boone, and D'Elia, 1999, *passim*).

Population norms may be applicable to tests that are relatively pure (and simple) measures of the function of interest (e.g., see Hannay, 1986) as the number of different kinds of variables contributing to a measure increases, the more likely will that measure's distribution approach normality (Siegel, 1956). The distributions of the WISC-A summed IQ scores (for the Verbal Scale [VIQ], the Performance Scale [PIQ], and both scales together, i.e., the Full Scale [FSIQ]) or scores on tests involving a complex of cognitive functions (e.g., Raven's Progressive Matrices) demonstrate this statistical phenomenon.

**Species-wide performance expectations**

The norms for some psychological functions and traits are actually species-wide performance expectations for
adults, although for infants or children they may be age or grade averages. This is the case for all cognitive functions and skills that follow a common course of development, that are usually fully developed long before adulthood, and that are taken for granted as part and parcel of the normal adult behavioral repertory. Speech is a good example. The average two-year-old child speaks in two- and three-word phrases. The ability to communicate verbally most needs and thoughts is expected of four and five year olds. Seventh- and eighth-grade children can utter and comprehend word groupings in all the basic grammatical forms and their elaborations. Subsequent speech development mainly involves more variety, elegance, abstractness, or complexity of verbal expression. Thus, the adult norm for speech is the intact ability to communicate effectively by speech, which all but a few adults can do. Some other skills that almost all neurologically intact adults can perform are counting change, drawing a recognizable person, and using simple construction tools or cooking utensils. Each of these skills is learned, improves with practice, has a common developmental history for most adults, and is sufficiently easy that its mastery or potential mastery is taken for granted. Anything less than an acceptable performance in an adult raises the suspicion of impairment.

Many species-wide capacities, although not apparent at birth, are manifested relatively early and similarly in all intact persons. Their development appears to be essentially maturational and relatively independent of social learning, although training may enhance their expression and aging may dull it. These include capacities for motor and visuomotor control and coordination; basic perceptual discriminations—e.g., of color, pattern, and form; of pitch, tone, and loudness; and of orientation to personal and extrapersonal space. Everyday life rarely calls upon the pure expression of these capacities. Rather, they are integral to the complex behaviors that make up the normal activities of children and adults alike. Thus, in themselves these capacities are usually observed only by deliberate examination.

Other species-wide normative standards involve components of behavior so rudimentary that they are not generally thought of as psychological functions or abilities. Binaural hearing, or the ability to localize a touch on the skin or to discriminate between noxious and pleasant stimuli, are capacities that are an expected part of the endowment of each human organism, present at birth or shortly thereafter. These capacities are not learned in the usual sense, nor, except when impaired by accident or disease, do they change over time and with experience. Some of these species-wide functions, such as fine tactile discrimination, are typically tested in the neurological examination (e.g., Perkin, 1998; Strub, 1996a).

Neuropsychological assessment procedures that test these basic functions possessed by all intact adults usually focus on discrete acts or responses and thus may identify the defective components of impaired cognitive behavior (e.g., A.-L. Christensen, 1979; Luria, 1999). However, examinations limited to discrete components of complex functions and functional systems provide little information about how well the patient can perform the complex behaviors involving component defects. Moreover, when the behavioral concomitants of brain damage are mild or subtle, particularly when associated with widespread or diffuse rather than well-defined lesions, few if any of these rudimentary components of cognitive behavior will be demonstrably impaired on the basis of species-wide norms.

**Customary standards**

A number of assumed normative standards have been arbitrarily set, usually by custom. Probably the most familiar of these is the visual acuity standard: 20/20 vision does not represent an average but an arbitrary ideal, which is met or surpassed by different proportions of the population, depending on age. Among the few customary standards of interest in neuropsychological assessment is verbal response latency—the amount of time a person takes to answer a simple question—which has normative values of one or two seconds for informal conversation in most Western cultures.

**Applications and limitations of normative standards**

Normative comparison standards are useful for most psychological purposes, including the description of cognitive status for both children and adults, for educational and vocational planning, and for personality assessment. In the assessment of persons with known or suspected adult-onset brain pathology, however, normative standards are appropriate only when the function or skill or capacity that is being measured is well within the capability of all intact adults and does not vary greatly with age, sex, education, or general mental ability. Thus, the capacity for meaningful verbal communication will be evaluated on the basis of population norms. In contrast, vocabulary level, which correlates highly with both social class and education (Heaton, Ryan, et al., 1996; Sattler, 2001a; P.E. Vernon, 1978), needs an individual comparison standard.

When it is known or suspected that a patient has suffered a decline in cognitive abilities that are normally distributed in the adult population, a description of that patient's functioning in terms of population norms (i.e., by standard test scores) will, in itself, shed no light on the extent of impairment unless there was documenta-
tion of premorbid cognitive levels (in school achievement tests or army placement examinations, for example). For premorbidly dull patients, low average scores would not indicate a significant drop in the level of examined functions. In contrast, an average score would represent a deficit for a person whose premorbid ability level had been generally superior (see pp. 148–149 for a statistical interpretation of ability categories). Moreover, comparisons with population averages do not add to the information implied in standardized test scores, for standardized test scores are themselves numerical comparisons with population norms. Thus, when examining patients for adult-onset deficits, only by comparing present with prior functioning can the examiner identify real losses.

The first step in measuring cognitive deficit in an adult is to establish—or estimate, when direct information is not available—the patient’s premorbid performance level for all of the functions and abilities being assessed. For those functions with species-wide norms, this task is easy. Adults who can no longer name objects or copy a simple design or who appear unaware of one side of their body have an obvious deficit. For normally distributed functions and abilities for which the normative standard is an average, however, only an individual comparison provides a meaningful basis for assessing deficit. A population average is not an appropriate comparison standard since it will not necessarily apply to the individual patient. By definition, one-half of the population will achieve a score within the average range on any well-constructed psychological test which generates a normal distribution of scores; the remainder perform at many different levels both above and below the average range. Although an average score may be, statistically, the most likely score a person will receive, statistical likelihood is a far cry from the individual case.

Individual Comparison Standards

As a rule, individual comparison standards are called for whenever a psychological trait or function that is normally distributed in the intact adult population is evaluated for change. This rule applies to both deficit measurement and the measurement of behavioral change generally. When dealing with functions for which there are species-wide or customary norms—such as finger-tapping rate or accuracy of auditory discrimination—normative standards are appropriate for deficit measurement. Yet even these kinds of abilities change with age and at some performance levels differ for men and women thus requiring some demographic norming.

The use of individual comparison standards is probably most clearly exemplified in rate of change studies, which depend solely on intraindividual comparisons. Here the same set of tests is administered three times or more at spaced intervals, and the differences between chronologically sequential pairs of test scores are compared. In child psychology the measurement of rate of change is necessary for examining the rate of development. Rate of change procedures also have broad applications in neuropsychology. Knowledge of the rate at which the patient’s performance is deteriorating can contribute to the accuracy of predictions of the course of a degenerative disease (e.g., see B.E. Levin, Tomer, and Rey, 1992; A.J. Thompson, 1998). For purposes of rehabilitation, the rate at which cognitive functions improve following cerebral insult may not only aid in predicting the patient’s ultimate performance levels but also provide information about the effectiveness of rehabilitative efforts (van Balen et al., 2002; Leclercq and Sturm, 2002). Further, rate of change studies contribute to understanding the long-range effects of brain injury on mental abilities.

THE MEASUREMENT OF DEFICIT

For most abilities that distribute normally in the population at large, determination of deficits rests on the comparison between what can be assumed to be the patient’s characteristic premorbid level of cognitive functioning as determined from historical data (including old test scores when available) and the obtained test performances (scores plus qualitative features). Thus, much of clinical neuropsychological assessment involves intraindividual comparisons of the abilities and skills under consideration.

Direct Measurement of Deficit

Deficit can be assessed directly when there are normative comparison standards against which the behavior in question can be compared. The extent of the discrepancy between the level of performance expected for an adult and the level of the patient’s performance (which may be given in terms of the age at which the average child performs in a comparable manner) provides one measure of the amount of deficit the patient has sustained. For example, the average six-year-old will answer 22 to 26 items correctly on the Verbal Comprehension test of the WJ-III Cog. The test performance of an adult who completed high school but can do no better could be reported as being “at the level of a six-year-old” on word knowledge.

Direct deficit measurement using individual comparison standards can be a simple, straightforward operation: The examiner compares premorbid and current
examples of the behavior in question and evaluates the discrepancies. Hooft, Vakil, and Gilboa’s (2000) study of cognitive impairment following brain injuries (mostly due to trauma) illustrates this procedure. They compared the scores that army veterans made on tests taken at the time of their induction into service with scores obtained on the Wechsler Adult Intelligence Scale–Revised (WAIS-R) postinjury approximately 13 years later. The findings of this direct comparison provided unequivocal evidence of cognitive impairment.

The direct method using individual comparison standards requires the availability of premorbid test scores, school grades, or other relevant observational data. In many cases, these will be nonexistent or difficult to obtain. Therefore, more often than not, the examiner must use indirect methods of deficit assessment from which individual comparison standards can be inferred.

**Indirect Measurement of Deficit**

In indirect measurement, the examiner compares the present performance with an *estimate* of the patient’s original ability level. This estimate may be drawn from a variety of sources. It is the examiner’s task to find meaningful and defensible estimates of the pretraumatic or premorbid ability levels to serve as comparison standards for each patient.

**Methods of indirect measurement**

Different methods of inferring the comparison standard for each patient have been applied with varying degrees of success (Axelrod, Vanderploeg, and Schinka, 1999; M.R. Basso, Bornstein, Roper, and McCoy, 2000; Crawford, 1992; Hooft, Vakil, and Gilboa, 2000; Johnstone, Slaughter et al., 1997; also see U.S. Congress, Office of Technology Assessment, 1987, pp. 282–283). Historical and observational data are obvious sources of information from which estimates of premorbid ability may be drawn directly. Estimates based on these sources will be more or less satisfactory depending on how much is known of the patient’s past, and whether what is known or can be observed is sufficiently characteristic to distinguish this patient from other people. For example, if all that an examiner knows about a brain injured, cognitively impaired patient is that he was a logger with a ninth-grade education and his observed vocabulary and interests seem appropriate to his occupation and education, then the examiner can only estimate a barely average ability level as the comparison standard. If the patient had been brighter than most, could reason exceptionally well, could tell stories cleverly, or had been due for a promotion to supervisor, this information would probably not be available to the examiner, who would then have no way of knowing from history and observations alone just how bright this particular logger had been.

Premorbid ability estimates inferred from historical and observational data alone may be spuriously low. Moreover, some patient self-reports may run a little high (Greiffenstein, Baker, and Johnson-Greene, 2002). Yet the need for accurate estimates has increasingly become apparent, especially in evaluating complaints of mental deterioration in older persons (Yuspeh et al., 1998). In response to this need, neuropsychologists have devised a number of distinctive methods for making these estimates. The most commonly used techniques for indirect assessment of premorbid ability rely on cognitive test scores, on extrapolation from current reading ability, on demographic variables, or on some combination of these. In reviewing these methods it is important to appreciate that, without exception, the comparison standard for evaluating them has been the three WIS-A IQ scores or the FSIQ. That the FSIQ as a criterion has its own problems becomes apparent when subjects’ cognitive functioning is not impaired, for then, when the estimate is derived only from the several highest Wechsler test scores, the average of all test scores (i.e., the FSIQ) will of necessity be lower than the derived estimate (excepting, of course, when the test score range covers no more than two points) (see p. 99 for an example). Moreover, the FSIQ will necessarily underrepresent the premorbid level of functioning when patients have cognitive compromise in areas tested by the WIS-A.

**Mental ability test scores for estimating premorbid ability**. A common feature of techniques based on test scores is that the premorbid ability level is estimated from the scores themselves. For many years a popular method for estimating premorbid ability level from test performance used a vocabulary score as the single best indicator of original intellectual endowment (Yates, 1954). This method was based on observations that many cognitively deteriorating patients retained old, well-established verbal skills long after recent memory, reasoning, arithmetic ability, and other cognitive functions were severely compromised. Moreover, of all the Wechsler tests, Vocabulary correlates most highly with education, which also can be a good indicator of premorbid functioning (Heaton, Ryan, et al., 1996; Johnstone, Slaughter et al., 1997; Tremont et al., 1999). A well-known example of this method is the Shipley Institute of Living Scale (SILS) (Shipley and Burlingame, 1941; Zachary, 1986; see pp. 669–670), which contains a multiple-choice (testing recognition rather than recall) vocabulary section and verbal reasoning items. It was expected that mentally deteriorating persons would
show marked discrepancies between their vocabulary and reasoning scores. A recent large-scale study (889 persons 60–94 years old) dispensed with the reasoning items: "WAIS-R Equivalent FSIQ" scores, calculated for the upper range (19–40) of SLS Vocabulary scores alone, are offered as estimates of premorbid ability in dementia examinations (Yuspeh et al., 1998).

D. Wechsler and others used the same principle to devise "deterioration ratios," which mostly compared scores on vocabulary and other verbally weighted scores with performance on tests sensitive to attentional deficits and visuomotor slowing (see pp. 655–656). On the assumption that certain cognitive skills will hold up for most brain damaged persons, McFie (1975)—and, more recently, Krull and his colleagues (1995)—proposed that the sturdiest tests in Wechsler's scales are Vocabulary and Picture Completion, both involving verbal skills. The average of the scores, or the highest score of the two should one of them be markedly depressed, becomes the estimated premorbid IQ score (McFie, 1975) when evaluated with demographic data (Krull et al., 1995, see p. 95; also see Axelrod, Vanderploeg, and Schinka, 1999). Vanderploeg and Schinka (1995) point out the obvious when observing that Verbal scale tests predict VSIQ best and that Performance scale tests predict PSIQ best. Combining the individual WAIS-R tests with demographic variables (age, sex, race, education, occupation) in a series of regression equations, Information and Vocabulary emerged as the best estimates of VSIQ and PSIQ and Block Design, Picture Completion, and Object Assembly gave the best estimates of PSIQ.

Larrabee, Largen, and Levin (1985) found that other Wechsler tests purported to be resilient (e.g., Information and Picture Completion) were as vulnerable to the effects of dementia as those Wechsler regarded as sensitive to mental deterioration. Moreover, the Similarities test, which Wechsler (1958) listed as vulnerable to brain dysfunction, held up best (in both WAIS and WAIS-R versions) when given to neuropsychologically impaired polysubstance abusers (J.A. Sweeney et al., 1988). Vocabulary and related verbal skill scores sometimes do provide the best estimates of the general premorbid ability level. However, vocabulary tests such as Wechsler's, which require oral definitions, tend to be more vulnerable to brain damage than verbal tests that can be answered in a way or two, require only recognition, or call on practical experience. Further, many patients with left hemisphere lesions suffer deterioration of verbal skills which shows up in relatively lower scores on more than one test of verbal function. Aphasic patients have the most obvious verbal disabilities; some are unable to use verbal symbols at all. Some patients with left hemisphere lesions are not technically aphasic, but their verbal fluency is sufficiently depressed that vocabulary scores do not provide good comparison standards.

**Word reading tests for estimating premorbid ability.**

In attempting to improve on vocabulary-based methods of estimating the cognitive deterioration of patients with diffusely dementing conditions, H.E. Nelson (1982; H.E. Nelson and Willison, 1991) and Crawford (with Parker and Besson, 1988; Crawford, Parker, Stewart, et al., 1989; Crawford, Deary, et al., 2001) proposed that scores on the *National Adult Reading Test (NART)* can reliably estimate the comparison standard, i.e., premorbid ability level. The NART requires oral reading of 50 phonetically irregular words, varying in frequency of use (see Table 13.9, p. 524). Of course, this technique can only be used with languages, such as English, in which the spelling of many words is phonetically irregular. In essence, these word reading tests provide an estimate of vocabulary size.

Correlations of NART-generated IQ score estimates with the WAIS and the WAIS-R (British version) FSIQ have run in the range of .72 (H.E. Nelson, 1982) to .81 (Crawford, Parker, Stewart et al., 1989). Correlations with the VSIQ are a little higher, while those with the PSIQ are considerably lower. A revision of the NART (NART-R UK) substituted new items for eight words that had not been scored reliably (Crawford, 1992). The correlation of the IQ score estimate generated by this form of the NART with the WAIS-R FSIQ score was .77. The NART and the British version of the WAIS-R were given to 179 77-year-olds who, at age 11, had taken a "group mental ability test" (presumably paper-and-pencil administration) (Crawford, Deary, et al., 2001). A correlation of NART estimates with the 66-year-old IQ scores ($r = .73$) showed that the late estimates were in the same range as the early test scores.

The *North American Adult Reading Test (NAART)* (Blair and Spreen, 1989; Spreen and Strauss, 1998) was developed for U.S. and Canadian patients. This 61-word list contains 35 of the original NART words (see Table 4.1, p. 93). While the NAART scores correlate reasonably well with the WAIS-R VSIQ ($r = .83$), correlation with the FSIQ ($r = .75$) leaves a great deal of unaccounted variance, and the correlation with the PSIQ ($r = .40$) is too low to be useful as a general indicator of premorbid ability (Spreen and Strauss, 1991). It is of interest that for this verbal skill test the mean number of words correctly pronounced steadily increased from $36.7 \pm 6.7$ at ages 16–29 to approximately $43 \pm 8.4$ at
TABLE 4.1 North American Adult Reading Test (NAART): Word List

<table>
<thead>
<tr>
<th>Deed</th>
<th>Placebo</th>
<th>Detective</th>
<th>Sidereal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debris</td>
<td>Procreate</td>
<td>Impugn</td>
<td>Abstemious</td>
</tr>
<tr>
<td>Alise</td>
<td>Psalm</td>
<td>Capon</td>
<td>Befity</td>
</tr>
<tr>
<td>Reign</td>
<td>Banal</td>
<td>Radii</td>
<td>Gaolied</td>
</tr>
<tr>
<td>Depot</td>
<td>Rarely</td>
<td>Aeon</td>
<td>Demease</td>
</tr>
<tr>
<td>Sincere</td>
<td>Gist</td>
<td>Epitome</td>
<td>Syncope</td>
</tr>
<tr>
<td>Lingerie</td>
<td>Corps</td>
<td>Equivocal</td>
<td>Ennui</td>
</tr>
<tr>
<td>Recipe</td>
<td>Hors d’oeuvre</td>
<td>Reify</td>
<td>Drachim</td>
</tr>
<tr>
<td>Gouge</td>
<td>Sieve</td>
<td>Indices</td>
<td>Gideon</td>
</tr>
<tr>
<td>Heir</td>
<td>Hiatus</td>
<td>Assignee</td>
<td>Epergne</td>
</tr>
<tr>
<td>Subtle</td>
<td>Gauze</td>
<td>Topiary</td>
<td>Vivace</td>
</tr>
<tr>
<td>Catacomb</td>
<td>Zealot</td>
<td>Caveat</td>
<td>Talipes</td>
</tr>
<tr>
<td>Bouquet</td>
<td>Paradigm</td>
<td>Superfluous</td>
<td>Synecdoche</td>
</tr>
<tr>
<td>Gauge</td>
<td>Facade</td>
<td>Leviathan</td>
<td></td>
</tr>
<tr>
<td>Colonel</td>
<td>Cellist</td>
<td>Prelate</td>
<td></td>
</tr>
<tr>
<td>Subpoena</td>
<td>Indict</td>
<td>Quadruped</td>
<td></td>
</tr>
</tbody>
</table>

From Spreen and Strauss [1998]

Ages 40–59 to 45.8 ± 8.4 at 70+ (Spreen and Strauss, 1998, p. 80).

The AMNART is a 45-word “American version” of the NART which proved sensitive to the developing semantic deficits of patients with early Alzheimer-type dementia (Storandt, Stone, and LaBarge, 1995). Mayo norms have been published for 361 healthy persons in 11 age ranges from 56 to 97 (Jvnik, Malek, Smith, et al., 1996). A recent 50-word version of the NART, the American National Reading Test (ANART) was developed to be more appropriate for the ethnically heterogeneous U.S. population (Gladsjo, Heaton, et al., 1999). It shares 28 words with the NAART. The ANART enhanced premorbid estimates for predominantly verbal tests to a limited degree, but made no useful contribution to estimates of either the PSIQ or scores of other tests with relatively few verbal components.

A short form of the NART (Short NART) was recommended for subjects who fail more than five of the first twenty-five items (Beardsall and Brayne, 1990), particularly those who fail many and are thus confronted with repeated failures. For subjects who pronounced between twelve and twenty of the first 25 NART words correctly, a procedure enables the examiner to estimate a full NART score; lower scores are assumed to be no different from what the total score would be if the entire word list had been given. While IQ scores obtained by this method correlated with NART estimates with “virtually equivalent” accuracy, these correlations left a considerable unexplained variance (23%–31%) and produced a small number of very discrepant estimates of ability as defined by the WIS-A IQ scores (Crawford, Allan, Jack, et al., 1991). Baddeley, Emslie, and Nimmo-Smith (1988) simplified even further Nelson’s relatively simple reading pronunciation task in a word recognition test (identifying which of a pair of letter groups is a real word). The level of difficulty at which the patient begins to fail consistently serves as an indicator of premorbid general ability.

A technique developed to be sensitive to anterior lesions, the Homophone Meaning Generation Test (HMGT), has been applied to the problem of estimating the degree to which a patient is impaired (Crawford and Warrington, 2002). The task, which requires a shift in cognitive set, asks the subject for other meanings for a set of eight homophones (e.g., pear—pair, pare; sight—site, cite) (Warrington, 2000, see p. 629). Noting that the HMGT correlated at a statistically significant level with the NART (r = .605), Crawford and Warrington (2002) devised a formula which, by evaluating the discrepancy between the HMGT raw score and the NART estimated premorbid score, will presumably provide a best estimate of “the severity of cognitive deficits.”

The Reading test of the Wide Range Ability Test (WRAT-READ) was developed on the same principle as the NART tests, using more to less frequently appearing words—although not all WRAT-READ words are phonetically irregular—to evaluate reading level (Wilkinson, 1993; see p. 525). Studies of its effectiveness in estimating premorbid mental ability have produced findings similar to those for the NART and its variants (Johnstone and Wilhelm, 1996; Kareken et al., 1995). In comparisons of NART-R and WRAT-READ, Wiens, Bryan, and Crossen (1993) reported that the former test best estimated their cognitively intact subjects whose FSIQ scores were in the 100–109 range while consistently overestimating those whose FSIQ scores fell below 100 and underestimated the rest; WRAT-READ’s estimations were more accurate in predicting lower FSIQ scores but underestimations of average and better FSIQ scores were even greater than for the NART-R. For neurologically impaired patients, a comparison of NAART and WRAT-READ found that while both “are appropriate estimates of premorbid verbal intelligence,” NAART had standardization and range limitations while WRAT-READ provided a better estimate of the lower ranges of the VSIQ making WRAT-READ more applicable to the population “at higher risk for TBI” (Johnstone, Callahan, et al., 1996).

1To use this formula, Crawford invites clinicians to download the computer program (http://www.psyc.ex.ac.uk/benedit/crawford/1HMGT.htm).
Correlations between these word reading tests and the criterion tests (mostly WIS-A IQ scores) tend to be directly related to education level (Crawford, Steward, Garthwaite, et al., 1988; Johnstone, Slaughter, et al., 1997; Maddrey et al., 1996; Storandt, Stone, and LaBarge, 1995; Stebbins, Wilson, et al., 1990). Some studies that took age into account dealt with subjects in the early to middle adult years with insignificant NART/NAART x age correlations resulting (e.g., Blair and Spreen, 1989; Wiens, Bryan, and Crossen, 1993). However, when subjects’ age range extends across several age cohorts into old age, age effects emerge (Crawford, Steward, Garthwaite, et al., 1988; Spreen and Strauss, 1998). Yet, although age effects reached significance ($r = -0.18$) for a wide range subject sample (ages 17–88), when the much stronger correlations for education ($r = -0.51$) and social class ($r = -0.36$) were partialled out, the very small age effects were nullified (Crawford, Steward, Garthwaite, et al., 1988). Kareken and his colleagues (1995) reported significant correlations between race (whites, African Americans) and all three WAIS-R IQ scores and WRAT-READ scores. They questioned whether “quality of education may be a mitigating factor,” but did not consider the pronunciation differences between “Black English” and standard American English (see pp. 313–314).

By large, the findings of studies on this technique have shown that when attempting to predict VSIQ and FSIQ scores of cognitively intact persons from their reading level, these tests are fairly accurate (Crawford, 1992; Crawford, Deary, et al., 2001; J.J. Ryan and Paolo, 1992; Spreen and Strauss, 1998; Wiens, Bryan, and Crossen, 1993). Regardless of which Wechsler edition is used, correlations between NART/NAART or WRAT-READ scores and VSIQ tend to be highest, FSIQ correlations are typically a little lower but still account for a large portion of the variance, while VSIQ correlations are too low for the reading test scores to be predictive of anything. Moreover, the greater the actual IQ score deviation from 100, the more discrepant estimates by the NART or one of its variants: “there is truncation of the spread of predicted IQs on either end of the distribution leading to unreliable estimates for individuals at other than average ability levels” (Spreen and Strauss, 1998, pp. 80–81).

Furthermore, reading test scores do tend to decline when given to dementing patients (Crawford, Millar, and Milne, 2001; Johnstone, Callahan, et al., 1996; Storandt, Stone, and LaBarge, 1995) but typically less than IQ scores (Maddrey et al., 1996). This method has been questioned as underestimating the premorbid ability of dementia patients—the degree of underestimation being fairly directly related to the severity of dementia (Stebbins, Wilson et al., 1990), of mildly demented patients with linguistic deficits (Stebbins, Gilley, et al., 1990), and of those more severely demented (Spreen and Strauss, 1991). For 20 elderly and neurologically impaired patients whose mean education was 8.8 ± 3.07 years, all three WAIS-R IQ scores (77.80 to 82.65) were significantly lower than NART estimates (from 93.05 to 95.25) (J.J. Ryan and Paolo, 1992). Yet, despite “mild” declines in NART-R scores, Maddrey and his colleagues (1996) recommend its use for dementing patients, even those whose deterioration is “more advanced.”

Correlations of the NART with the three Wechsler IQ scores were a little lower for an English speaking South African population than for U.K. subjects. This discrepancy suggests that a language test standardized on one population may not work as well with another in which small differences in language have developed over time (Struben and Tredoux, 1989; see pp. 313–314).

Demographic variables for estimating premorbid ability. One problem with word-reading scores is their vulnerability to brain disorders, especially those involving verbal abilities; one advantage of demographic variables is their independence from the patient’s neuropsychological status at the time of examination. In questioning the use of test score formulas for estimating premorbid ability (specifically, WIS-A FSIQ scores), R.S. Wilson, Rosenbaum, and Brown (1979; also in Rourke, Costa, et al., 1991) devised the first formula using demographic variables (age, sex, race, education, and occupation) to make this estimation. This formula predicted only two-thirds of 491 subjects’ WAIS FSIQ scores within a ten-point error range; most of the larger prediction errors occurred at the high and low ends of their sample, overpredicting high scores and underpredicting low ones (Karzmark, Heaton, et al., 1985; also in Rourke, Costa, et al., 1991). Exaggerated estimations at the distribution extremes were also reported by F.C. Goldstein, Gary, and Levin (1986; also in Rourke, Costa, et al., 1991), although Wilson’s formula provided “an adequate” fit to WAIS data for the 69 neuropsychologically unimpaired subjects.

Recognizing the need for ability estimates geared to the WAIS-R, Barona, Reynolds, and Chastain (1984) elaborated on Wilson’s work by incorporating the variables of geographic region, urban–rural residence, and handedness into the estimation formula. They devised three formulas for predicting each of the WAIS-R IQ scores. These authors did not report the amount and extent of prediction errors produced by their formulas but cautioned that, “where the premorbid Full Scale IQ was above 120 or below 69, utilization of the formula [sic] might result in a serious under- or over-estimation,
respectively” (p. 887). Other studies evaluating both the Wilson and the Barona estimation procedures found that at best they misclassified more than one-half of the patients (Silverstein, 1987), or “both formulas perform[ed] essentially at chance levels” (Sweet, Moberg, and Tovian, et al., 1990). Perez and her colleagues (1996) suggested that this formula—and its 1986 elaboration by Barona and Chastain—was useful for premorbid ability estimates but obtained VSIQ and FSIQ correlations in the .48 to .52 range for the control group. Yet for classifying subjects on the basis of the size of differences between estimated and observed scores, the original Barona procedure identified 76.7% of the neurologically impaired patients and 90% of control subjects; an elaboration of the Barona procedure (Barona and Chastain, 1986) improved classification to 80% and 93% of patients and control subjects, respectively. Helmes (1996) applied the 1984 Barona equations in a truly large-scale study (8,660 randomly selected elderly Canadians—excluding three women in their 100s). The three IQ score means of each 5-year group calculated from this formula were within one point of 100 with few exceptions, suggesting that this formula produced reasonably accurate estimates. Main effects for sex and education were significant. However, another study comparing estimation techniques found that the 1984 Barona method generated the lower correlation of estimated FSIQ with actual FSIQ ($r = .62$) (Axelrod, Vanderploeg, and Schinka, 1999).

A set of equations developed on British demographic data gave even less accurate estimates of WAIS (British) scores, predicting just 50% of the variance for FSIQ and VSIQ scores, and only 30% for the PVIQ score (Crawford, Stewart, Cochrane, et al., 1989). In a later study of the predictive value of demographic variables, Crawford and Allan (1997) found that occupation provided the best estimate of the three WAIS-R IQ scores with correlations of -.65, -.65, and -.50 for FSIQ, VSIQ, and PVIQ, respectively. It is noteworthy that occupation and education correlated relatively highly ($r = .65$). When age and education were added in, the multiple regression results accounted for 53%, 53%, and 32% of the variance for the three IQ scores, respectively. As in most other studies, the contribution of age was negligible. This demographic formula joins word reading tests in not predicting PVIQ effectively.

Demographic variables combined with test scores for estimating premorbid ability. Further efforts to improve estimates of premorbid ability have generated formulas that combine word recognition test scores with demographic variables. In studies of normal subjects, while demographic variables accounted for 30% of the FSIQ score variance (Crawford, Stewart, Cochrane, et al., 1989; Crawford, Stewart, Parker, et al., 1989) and NART scores alone predicted 66% of the variance, the FSIQ score variance based on a combination of these variables was 73% (Crawford, Stewart, Parker, et al., 1989). Strong relationships showed up between scores generated by equations combining NART scores with demographic variables and scores on individual WAIS tests: the greatest factor loadings were on the highly verbal tests (in the .76-.89 range), with almost as strong relationships (.71 and .72) occurring between the equation-generated scores and the Block Design and Arithmetic tests, respectively (Crawford, Cochrane, Besson, et al., 1990). These workers interpreted the findings as indicating that an appropriate combination of the NART score and demographic variables provides a good measure of premorbid general ability. However, another study examining different subject groups (e.g., Korsakoff’s syndrome, Alzheimer’s disease) found that NART (and NART-R) alone correlated better with WIS-A FSIQ than did either of two demographic formulas, nor did combining NART and demographic data enhance NART estimates (Bright et al., 2002).

Other attempts to enhance the accuracy of premorbid estimations from current test performance involve WIS-A tests. Krull and his colleagues (1995) devised the Oklahoma Premorbid Intelligence Estimation (OPIE), using Vocabulary and Picture Completion scores of the WAIS-R standardization population along with age, education, occupation, and race data. They generated formulas for predicting VSIQ, PSIQ, and FSIQ, evaluating the accuracy of their formulas against the (presumably) cognitively intact WAIS-R standardization population on which they were developed. Not surprisingly, the predicted and actual correlations were high ($r = .87, .78, .87$ for V-, P-, and FSIQ scales, respectively). OPIE formulas were then developed to predict FSIQ using raw scores for Vocabulary, Picture Completion, both of these tests, or the raw score for whichever of these two tests had the highest nonage-corrected scaled score (BEST method) for subjects in the authors’ patient database (J.G. Scott et al., 1997). FSIQ predictions made by the BEST method most closely approximated the normative distribution’s mean and standard deviation, a finding interpreted as indicating that the BEST method gave the best estimation. The formula using both Vocabulary and Picture Completion scores produced the least appropriate FSIQ approximations.

Because different patients will make their highest score(s) on different WIS-A tests, Vanderploeg and Schinka (1995) developed tables—which take demographic data of age, sex, race, education, and occupation into account—for estimating the three WAIS-R IQ
scores from the 11 WAIS-R tests. They too used the WAIS-R standardization data to generate their 33 regression formulas, three for each WAIS-R test. Presumably, when the obtained IQ score is significantly below the expected IQ score level for a test, a deficit may be inferred.

**Comparisons between methods for estimating premorbid ability.** With so many estimation procedures to choose from, it is natural to wonder which works best. M.R. Basso, Bornstein, and their colleagues (2000), after testing the Barona, revised Barona, OPIE, and BEST-3, concluded that none of the methods based on regression formulas were satisfactory. They pointed out that the phenomenon of regression to the mean affected all these methods, most significantly the Barona (i.e., purely demographic) methods. Scores at the extremes of the IQ range were most vulnerable to estimation errors. The prediction accuracy of other studies (see below) tends to vary with the demographic characteristics of the samples tested.

For each of the three WIS-A IQ scores, Kareken and his colleagues (1995) compared formulas that included parental education level and race with WRAT-R reading scores to estimations derived from the original Barona (et al., 1984) equation. Using “healthy young adults,” these workers reported that while the average discrepancy between these two estimates was “moderate,” the reading + parental education technique generated both higher scores and a broader range of estimated scores—broader than did Barona estimates or the actual score range. Since correlations between the two methods indicated shared variances of only moderate size (for V-, P-, and FSIQ scores, r = .46, .61, and .55, respectively), the authors concluded that each method “tap[s] different aspects of variance.”

When estimates of impairment due to TBI and derived from WRAT-R reading scores were compared with impairment estimates based on education level, the latter method produced larger impairment estimates for the WAIS-R FSIQ score and also for two noncognitive tests: Grip Strength and Finger Tapping (Johnstone, Slaughter, et al., 1997). The reading score impairment estimations exceeded those predicted by education level on each of the two trials of the Trail Making Test. The authors wisely conclude that “different methods of estimating neuropsychological impairment produce very different results” and suggest that neither of these methods is appropriate for estimating premorbid levels of motor skills.

A comparison of five methods for predicting premorbid ability level used as a criterion how closely the estimated FSIQ of brain impaired patients approximated the actual FSIQ score of matched control sub-

jects (J.G. Scott et al., 1997). Four methods were based on a combination of WIS-A test scores and demographic data: three OPIE variants and a procedure using the OPIE equation that generated the highest score (BEST-3); a fifth was the purely demographically based Barona (et al., 1984) procedure. The demographically based method produced the smallest discrepancy between the clinical sample and the matched control group, and although it had the highest rate of group classification (based on estimated obtained scores), all five methods had “an equal degree of overall classification accuracy.” The Barona score had the lowest correlation by far with the subjects’ actual FSIQ scores (r = .62; all others were in the .84 to .88 range). The authors point out discrepancies between these findings and those of previous studies in concluding that the four methods using OPIE equations were “equally effective,” while expressing puzzlement over the Barona method’s history of good performance in predicting FSIQ scores and in classifying subjects.

Also comparing the Barona and OPIE methods for estimating premorbid intelligence with two reading tests (NAART, WRAT-3), S.L. Griffin and her coworkers (2002) report that the Barona method was least useful, overestimating WAIS-R “below average” and “average” FSIQ scores and underestimating those in the “above average” ranges. OPIE overestimated the “average” FSIQ scores, NAART overestimated “below average” and “average” FSIQ, and the WRAT-R underestimated both “below average” and “above average” FSIQ.

With premorbid ability scores for 54 neurologically impaired patients, Hoofien, Vakil, and Gilboa (2000) evaluated two estimation procedures. BEST-10 utilizes the highest ten predicted test scores generated from Vanderploeg and Schinka’s (1995) 30 prediction equations (the Hebrew WAIS-R has no Vocabulary test) which included age, sex, race, premorbid occupation, and premorbid education. BEST-2, based on which of the Information or Picture Completion test scores is highest, integrates into its score the same demographic variables. For predictive validity, the current WAIS-R FSIQ score correlated significantly with the premorbid transformed IQ score (r = .63); BEST-10’s correlation was virtually identical, and BEST-2’s correlation was a little lower (r = .58); concurrent validity correlations were, as expected, considerably higher (BEST-2 r = .86, BEST-10 r = .85). When comparing paired differences between the transformed premorbid IQ score, the postmorbid WAIS-R FSIQ, and the two BEST methods, the postmorbid FSIQ for these patients was 19.04 points lower than their premorbid scores, BEST-2 estimations were 5.39 points below premorbid IQ scores, but BEST-10 was only 2.07 points lower than the orig-
inal test scores. BEST-10 premorbid estimations had fewer (8) instances of IQ score differences of 15 or greater than did BEST-2 with 13 such instances. These data suggest that BEST-10 may provide a better estimate of premorbid ability than BEST-2. The authors suggest that estimation procedures using best performances have higher predictive accuracy than those constructed on assumptions regarding "hold" tests. However, Hoofien and his colleagues (2000) warn against a purely "mechanical" application of the BEST-10 method, observing that some isolated skills or abilities can lead to an undeservedly high estimate. For such cases they recommend both the inclusion of demographic data for its "balancing effect" and clinical judgment. These authors further note that the large differences between the predictive validity correlations and those for concurrent validity cast serious question on estimation methods validated on current test scores.

Although none of these methods satisfies the clinical need for a reasonably accurate estimate of premorbid ability, all of them show the value of extratest data and the penalties paid for restricting access to any particular kind of information when seeking the most suitable comparison standards for a cognitively impaired patient.

THE BEST PERFORMANCE METHOD

A simpler method utilizes test scores, other observations, and historical data. This is the best performance method, in which the level of the best performance—whether it be the highest score or set of scores, non-scorable behavior not necessarily observed in a formal testing situation, or evidence of premorbid achievement—serves as the best estimate of premorbid ability. Once the highest level of functioning has been identified, it becomes the standard against which all other aspects of the patient’s current performance are compared.

The best performance method rests on a number of assumptions that guide the examiner in its practical applications. Basic to this method is the assumption that, given reasonably normal conditions of physical and mental development, there is one performance level that best represents each person’s cognitive abilities and skills generally. This assumption follows from the well-documented phenomenon of the transitiational consistency of cognitive behavior. According to this assumption, the performance level of most normally developed, healthy persons on most tests of cognitive functioning probably provides a reasonable estimate of their performance level on most other cognitive tasks (see B.D. Bell and Roper, 1998, for a discussion of this phenomenon at the high average ability level; Dodrill, 1999, gives an example at the low average level). This assumption allows the examiner to estimate a cognitively impaired patient’s premorbid general ability level from one or, better yet, several current test scores while also taking into account other indicators such as professional achievement or evidence of a highly developed skill.

Intraindividual differences in ability levels may vary with a person’s experience and interests, perhaps with sex and handedness, and perhaps on the basis of inborn talents and deficiencies. Yet, by and large, persons who perform well in one area perform well in others; and the converse also holds true: a dullard in arithmetic is less likely to spell well than is someone who has mastered calculus. This assumption does not deny its many exceptions, but rather speaks to a general tendency that enables the neuropsychological examiner to use test performances to make as fair an estimate as possible of premorbid ability in neurologically impaired persons with undistinguished school or vocational careers. A corollary assumption is that marked discrepancies between the levels at which a person performs different cognitive functions or skills probably give evidence of disease, developmental anomalies, cultural deprivation, emotional disturbance, or some other condition that has interfered with the full expression of that person’s cognitive potential. An analysis of the WAIS-R normative population into nine average score “core” profiles exemplifies this assumption as only one profile, accounting for 8.2% of this demographically stratified sample, shows a variation of as much as 6 scaled score points, and one that includes 6.2% of the sample shows a 9-point disparity between the average high and low scores (McDermott et al., 1989). The rest of the scatter discrepancies are in the 0–4 point range.

Another assumption is that cognitive potential or capacity of adults can be either realized or reduced by external influences; it is not possible to function at a higher level than biological capacity will permit. Brain injury—or cultural deprivation, poor work habits, or anxiety—can only depress cognitive abilities (A. Rey, 1964). An important corollary to this assumption is that, for cognitively impaired persons, the least depressed abilities may be the best remaining behavioral representatives of the original cognitive potential (Axelrod, Vanderploeg, and Schinka, 1999; Hoofien, Vakil, and Gilboa, 2000; Krull et al., 1995; J.G. Scott et al., 1997).

The phenomenon of overachievement (people performing better than their general ability level would seem to warrant) appears to contradict this assumption; but in fact, overachievers do not exceed their biological limitations. Rather, they expend an inordinate amount of energy and effort on developing one or two special skills, usually to the neglect of others. Academic overachievers generally know their material mostly by rote and cannot handle the complex mental operations
or highly abstract concepts enjoyed by people at superior and very superior ability levels.

A related assumption is that few persons consistently function at their maximum potential, for cognitive effectiveness can be compromised in many ways: by illness, educational deficiencies, impulsivity, test anxiety, disinterest—the list could go on and on. A person's performance of any task may be the best that can be done at that time but still only indicates a floor, not the ceiling, of the level of abilities involved in that task. Running offers an analogy: no matter how fast the runner, the possibility remains that she could have reached the goal even faster, if only by a fraction of a second.

Another related assumption is that within the limits of chance variations, the ability to perform a task is at least as high as a person's highest level of performance of that task. It cannot be less. This assumption may not seem to be so obvious when a psychologist is attempting to estimate a premorbid ability level from remnants of abilities or knowledge. In the face of a generally shabby performance, examiners may be reluctant to extrapolate an estimate of superior premorbid ability from one or two indicators of superiority, such as a demonstration of how to use a complicated machine or the apt use of several abstract or uncommon words, unless they accept the assumption that prerequisite to  knowledge or the development of any skill is the ability to learn or perform it. A patient who names Grant as president of the United States during the Civil War and says that Greece is the capital of Italy but then identifies Einstein and Marie Curie correctly is demonstrating a significantly higher level of prior intellectual achievement than the test score suggests. The poor responses do not negate the good ones; the difference between them suggests the extent to which the patient has suffered cognitive deterioration.

It is also assumed that a patient's premorbid ability level can be reconstructed or estimated from many different kinds of behavioral observations or historical facts. Material on which to base estimates of original cognitive potential may be drawn from interview impressions, reports from family and friends, test scores, prior academic or employment level, school grades, army rating, or an intellectual product such as a letter or an invention. Information that a man had earned a Ph.D. in physics or that a woman had designed a set of complex computer programs is all that is needed to make an estimate of very superior premorbid intelligence, regardless of present mental dilapidation. Except in the most obvious cases of unequivocal high achievement, the estimates should be based on information from as many sources as possible to minimize the likelihood that significant data have been overlooked, resulting in an underestimation of the patient's premorbid ability level.

Verbal fluency can be masked by shyness, or a highly developed graphic design talent can be lost to a motor paralysis. Such achievements might remain unknown without careful testing or painstaking inquiry.

The value of the best performance method depends on the appropriateness of the data on which estimates of premorbid ability are founded. This estimation method places on the examiner the responsibility for making an adequate survey of the patient's accomplishments and residual abilities. This requires sensitive observation with particular attention to qualitative aspects of the patient's test performance; good history taking, including—when possible and potentially relevant—contacting family, friends, and other likely sources of information about the patient such as schools and employers; and enough testing to obtain an overview of the patient's cognitive abilities in each major functional domain.

The best performance method has very practical advantages. Perhaps most important is that a broad range of the patient's abilities is taken into account in identifying a comparison standard for evaluating deficit. By looking at the whole range of cognitive functions and skills for a comparison standard, examiners are less likely to bias their evaluations of any specific group of patients, such as those with depressed verbal functions. Moreover, examiners using this method are not bound to one battery of tests or to tests alone for they can base their estimates on non-test behavior and behavioral reports as well. For patients whose general functioning is too low or too spotty for them to complete a standardized adult test, or who suffer specific sensory or motor defects, children's tests or tests of specific skills or functions used for career counseling or job placement provide opportunities to demonstrate residual cognitive abilities.

In general, the examiner should not rely on a single high test score for estimating premorbid ability unless history or observations provide supporting evidence. The examiner also needs to be alert to overachievers whose highest scores are generally on vocabulary, general information, or arithmetic tests, as these are the skills most commonly inflated by parental or school pressure on an ordinary student. Overachievers frequently have high memory scores too. They do not do as well on tests of reasoning, judgment, original thinking, and problem solving, whether or not words are involved. One or two high scores on memory tests should not be used for estimating the premorbid ability level since, of all the cognitive functions, memory is the least reliable indicator of general cognitive ability. Dull people can have very good memories; some extremely bright people have been notoriously absent-minded.

It is rare to find only one outstandingly high score in a complete neuropsychological examination. Usually even
severely impaired patients produce a cluster of relatively higher scores in their least damaged area of functioning so that the likelihood of overestimating the premorbid ability level from a single, spuriously high score is slight. The examiner is much more likely to err by underestimating the original ability level of the severely brain injured patient who is unable to perform well on any task.

In criticizing this method for systematically producing overestimates of premorbid ability, Mortensen and his colleagues (1991) give some excellent examples of how misuse of the best performance method can result in spurious estimates. Most of their “best performance” estimates were based solely on the highest score obtained by normal control subjects on a WIS-A battery. What they found, of course, was that the highest score among tests contributing to a summation score (i.e., an IQ score) is always higher than the IQ score since the IQ score is essentially a mean of all the scores, both higher and lower. Therefore, in cognitively intact subjects, the highest WIS-A test score is not an acceptable predictor of the WIS-A IQ score. Moreover, in relying solely on the highest score, the Mortensen study violated an important directive for identifying the best performance: that the estimate should take into account as much information as possible about the patient and not rely on test scores alone. In most cases, the best performance estimate will be based on a cluster of highest scores. Thus, developing a comparison standard using this method is not a simple mechanical procedure but calls upon clinical judgment and sensitivity to the many different conditions and variables that can influence a person’s test performances.

THE DEFICIT MEASUREMENT PARADIGM

Once the comparison standard has been determined, whether directly from population norms, premorbid test data, or historical information, or indirectly from current test findings and observation, the examiner may assess deficit. This is done by comparing the level of the patient’s present cognitive performances with the expected level—the comparison standard. Discrepancies between the expected level and present functioning are then evaluated for statistical significance (see pp. 148–149, 153–154). A statistically significant discrepancy between expected and observed performance levels for any cognitive function or activity represents a cognitive deficit.

This comparison is made for each test score. For each comparison where premorbid test scores are not available, the comparison standard is the estimate of original ability. By chance alone, a certain amount of variation (scatter) between test scores can be expected for even the most normal persons. Although these chance variations tend to be small (Cronbach, 1984), they can vary with the test instrument and with different scoring systems (see p. 652 for a discussion of the use of age-graded scores in interpreting WIS-A data). If significant discrepancies occur for more than one test score, a pattern of deficit may emerge. By comparing any given pattern of deficit with patterns known to be associated with specific neurological or psychological conditions, the examiner may be able to identify etiological and remedial possibilities for the patient’s problems. When differences between expected and observed performance levels are not statistically significant, deficit cannot be inferred on the basis of just a few higher or lower scores.

For example, it is statistically unlikely that a person whose premorbid ability level was decidedly better than average cannot solve fourth- or fifth-grade arithmetic problems on paper or name at least 16 animals in one minute. If the performance of a middle-aged patient whose original ability is estimated at the high average level fails to meet these relatively low performance levels, then an assessment of impairment of certain arithmetic and verbal fluency abilities can be made with confidence. If the same patient performs at an average level on tests of verbal reasoning and learning, that discrepancy is not significant even though performance is somewhat lower than expected. These somewhat lowered scores need to be considered in any overall evaluation in which significant impairment has been found in other areas. However, when taken by themselves, average scores obtained by patients of high average mental competence do not indicate impairment, since they may be due to normal score fluctuations. In contrast, just average verbal reasoning and learning scores achieved by persons of estimated original very superior endowment do represent a statistically significant discrepancy, so that in exceptionally bright persons, average scores indicate deficit.

Identifiable patterns of cognitive impairment can be demonstrated by the deficit measurement method. Although the discussion here has focussed on assessment of deficit where a neurological disorder is known or suspected, this method can be used to evaluate the cognitive functioning of psychiatrically disabled or educationally or culturally deprived persons as well because the evaluation is conducted within the context of the patient’s background and experiences, taking into account historical data and the circumstances of the patient’s present situation (Gollin et al., 1989; W.G. Rosen, 1989). The evaluation of children’s cognitive disorders follows the same model (Hynd and Willis, 1987; Sattler, 2001a; E.M. Taylor, 1959). It is of use not only as an aid to neurological or psychiatric diagnosis but also in educational and rehabilitation planning.