CHAPTER 8
THE VIP TEST IN COURT

VIP users may be called upon to present their clinical findings and conclusions, either in person or by report, to a judge, jury, disability board, or agency that makes decisions about remunerative awards. This chapter discusses justifications and evidentiary standards that are pertinent to use of the VIP test in court.

Justifications for VIP Use

There are several justifications for using an objective validity indicator in cognitive testing that may be pertinent to defending VIP use in court. The first is that malingering occurs frequently enough in psycholegal situations to warrant concern. The second is that malingering has strongly undesirable consequences and therefore requires assessment. The third is that malingering can be assessed and that accurate assessment requires objective techniques because subjective clinical judgment is inadequate for detecting malingering. These justifications are elaborated upon below.

Malingering occurs frequently in psycholegal situations. As discussed in Chapter 2, a large number of studies have concluded that malingering routinely occurs in 20% to 30% or more of forensic examinations conducted for personal injury cases and disability determinations and in at least 15% of examinations conducted for

criminal cases. Recently Mittenberg et al. (2002) reviewed more than 30,000 cases in a one-year period that involved individuals in personal injury assessment, disability determinations, criminal defense examinations, and medical examinations. The authors estimated rates of feigned cognitive impairment at 29% for personal injury cases, 30% for disability cases, 19% for criminal cases, 8% for medical cases in general, 35% for fibromyalgia and chronic fatigue cases, 31% for chronic pain cases, and 27% for neurotoxic cases. They also noted that because of sampling issues, these rates probably underestimated the true rates by about 2% to 4%.

Malingering has undesirable consequences. While it seems obvious that malingering has undesirable societal consequences, the case of *U.S. v. Greer* (1998) provides a legal opinion to that effect, specifically, that malingering in a litigated case constitutes an obstruction of justice. In 1994, Charles Greer was arraigned on federal charges of kidnapping and firearms violations. State charges had initially been dismissed after Greer was determined to be incompetent as the result of a mental disorder. The federal prosecutor pursued the case, and Greer was hospitalized at a federal medical center for mental health evaluation. The evaluating psychologist testified that Greer was competent and malingering psychopathology and cognitive impairment.

The judge ruled that he was competent. However, Greer was so disruptive in jail over the next year while awaiting trial that he was reevaluated on an outpatient basis, ruled to be incompetent after another hearing, and committed to a different federal medical center for restoration of competence. After a period of hospitalization, Greer was again evaluated by a psychologist, who concluded that he was malingering and competent. A third competency hearing was held, and the court agreed that Greer was malingering and competent. He was eventually convicted on all counts.

At sentencing, the government asked the court to increase the offense level for purposes of sentencing pursuant to the U.S. Sentencing Guidelines based on the premise that Greer had obstructed justice by pretending to be incompetent. The court granted an enhancement, increasing the sentence from 185 to 210 months. Greer appealed to the U.S. Fifth Circuit Court of Appeals, claiming that the sentence enhancement undercut his right to be tried only if competent. The appellate court affirmed the finding of obstruction of justice by malingering, reasoning that malingering constitutes obstruction of justice because it involves egregiously wrong behavior that requires a significant amount of planning and inherent high risk that justice will indeed be obstructed. The court ruled that feigning incompetency is similar to altering evidence and creating a false record: "A defendant who playacts . . . essentially tries to create a record that includes inaccurate testimony and factual conclusions" (*U.S. v. Greer*, 1998, p. 235). Further, as part of its decision the appellate court recommended that attorneys advise their clients to cooperate during assessment of abilities.

Malingering can be assessed, but assessment requires objective tests and techniques. In Chapter 2, a number of studies were discussed that showed that subjective clinical judgment alone is inadequate to detect cognitive malingering. For example, Faust, Hart, and Guilmette (1988) and Faust, Hart, Guilmette, and Arkes (1988) found that neuropsychologists could do no better than chance in identifying protocols from subjects who were simulating cognitive deficits. Further, the clinicians' accuracy was unrelated to their level of confidence in their classifications.

In response to the inadequacy of subjective clinical judgment, a number of objective tests and procedures for assessing malingering have been published over the past 10 to 15 years. Foremost, Rogers (1990b) has promulgated a process for making determinations about malingering. Rogers emphasized the need to have convincing evidence of malingering; his book (Rogers, 1997) is an edited compendium of practical methods and psychological tests that are useful for detecting malingering for a variety of contexts and feigned presentations. A number of comprehensive reviews of malingering detection methods are available, including Haines and Norris (1995), Iverson and Binder (2000), Nies and Sweet (1994), Rogers and Bender (2003), and Rogers et al. (1993).

In general, objective testing is often superior to subjective evaluations. This is because with objective testing (a) questions are asked in a standardized manner, (b) idiosyncratic contributions of the interviewer are minimized, (c) standardized administration allows comparison of results across test-takers, and (d) clinical predictions, diagnoses, and classifications are governed by the same decision rules (Cronbach, 1990).

Many now feel that an objective validity indicator is a necessary element of judging response validity when testing depends on the good faith of the test-taker. For example, Sweet (1999) recommended the use of tests specifically designed to detect feigned neuropsychological impairment. Similarly, van Gorp et al. (1999) concluded that clinicians "should rely more heavily on neuropsychological measures which have been designed to detect malingering or are clinical measures which have been validated for the detection of malingering in making a determination of malingering versus honest responding" (p. 249). Iverson and Binder (2000) wrote that "a forensic evaluation that does not include careful consideration of possible negative response bias should be considered incomplete" (p. 853).

The VIP Test and the *Daubert* Standards

In legal proceedings, evidence must be scrutinized to determine if it is reliable and valid and has some probative value (i.e., contributes to a decision). Certain standards are applied when evidence is scrutinized. In the case of *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993; see for

example Shuman & Sales, 1999), the Supreme Court articulated standards for the admissibility of scientific evidence in the federal courts. *Daubert* cites four standards for assessing scientific evidence:

1. The underlying theory or technique can be and has been tested.

2. The procedure has been subjected to scrutiny by others in the field through peer review.

3. Error rates and the standards for controlling them are acceptable.

4. The procedure has acceptance within the scientific community.

Although these four standards do not apply in all jurisdictions, some of them are fairly similar to standards (e.g., general acceptability) that apply in other jurisdictions. The remainder of this chapter reviews how the VIP test meets the *Daubert* standards. VIP users may wish to cite this information in defending use of the VIP test to help make decisions about the validity of concurrently administered psychological tests.

The techniques underlying the VIP test can be and have been tested. The VIP response style classification rules (described in Chapter 4) and guidelines for test interpretation (given in Chapter 7) are based on a complementary set of techniques for analyzing cognitive test performance that have been validated and enjoy wide professional acceptance. These techniques or strategies are "comparison to chance," "floor effect," and Performance Curve analysis. Evidence supporting the use of these techniques to identify invalid responding on cognitive tests is presented below.

Comparison to chance. Some aspects of VIP response style classification and interpretation depend on comparisons to what would be expected if the test-taker had merely guessed at the answers. On a two-alternative forced-choice test like the VIP test, an individual who does not know the answer to any given item nevertheless has a 50% chance of answering that item correctly just by guessing. Similarly, for any number of items, probabilities can be computed that represent the chances of each number of correct answers occurring if the respondent merely guessed. When the number of correct answers is so low that it is very unlikely even if the test-taker guessed on every item (e.g., less than 40 out

of 100), there is strong evidence that the individual was deliberately choosing incorrect answers.

This "comparison to chance" approach has long been accepted as providing an indicator of response validity (e.g., Brady & Lind, 1961; Frederick & Denney, 1998; Pankratz, 1979; Rogers et al., 1993; Theodor & Mandelcorn, 1973). Well-recognized and commercially available procedures for the detection of cognitive/ memory malingering, such as the Portland Digit Recognition Test (PDRT; Binder, 1993) and the Test of Memory Malingering (TOMM; Tombaugh, 1997), incorporate "below-chance" responding in interpretation of test results. Specific VIP measures that are compared to values expected for guessing in determining response style classification and/or in interpreting VIP results are the Total Score, the Point of Entry, the Verbal Tail Score, and the various Suppression Sector variables.

When a test-taker scores significantly below chance expectations, it provides very compelling evidence that he or she is malingering; in other words, the specificity of this technique is very high. However, the sensitivity tends to be low because few individuals, even those who are malingering, score significantly below what would be expected for a test-taker who has no ability whatsoever to determine any correct answers. Therefore, while comparison to chance can be a useful element of an overall assessment of response validity, reliance on this strategy alone is not optimal.

Floor effect. The strategy of using a "floor effect" to identify feigned impairment requires observing performance on tasks or problems involving overlearned material (e.g., stating one's name or age, reciting the alphabet) or test items that are easily solved by individuals who are genuinely impaired. An example of a floor effect test is the Rey 15-Item Memory Test (Rey, 1958), which requires the memorization and recall of easily retained information.

The widely recognized PDRT and TOMM, mentioned above, although reporting below-chance performance when it occurs, depend primarily on this floor effect strategy for identifying malingering. Cut-off scores for these tests are well above chance levels and have been derived from studies of the performance of individuals with bona fide impairment.

VIP measures that involve a floor effect strategy include the Point of Entry, the Sector 1 Residual, and comparison of Sector 1 Distance to Sector 2 Distance. The performance of individuals with mental retardation supports a value of 0.8 as a reasonable floor for the Point of Entry (see Chapter 5).

The Sector 1 Residual and the comparison of Sector 1 Distance to Sector 2 Distance involve a sort of twist to the floor effect strategy. Both measures involve comparing an individual's performance on the VIP test to a floor established by the individual's own performance. That is, the measures identify the individual's capacity for extended "better than chance" responding and then identify instances in which his or her performance was worse than expected given his or her level of demonstrated ability (i.e., instances in which he or she performed below his or her own floor). This sort of analysis has been described by Frederick (2000b).

The floor effect strategy, though offering greater sensitivity than the comparison to chance approach, nevertheless still results in a simple dichotomous classification such as "compliant" or "noncompliant." No additional insights are provided to help the clinician understand the nature of the noncompliant performance or to evaluate whether compliant performance represents the maximal ability of the test-taker. "If someone passes . . . all we usually can say is that she was not grossly underrepresenting her abilities on that procedure. It is not sound to conclude that because effort was not bad, it was fine or good" (Faust & Auckley, 1998, p. 47). The third strategy employed by the VIP test, Performance Curve analysis, complements the comparison to chance and floor effect strategies and allows the VIP test to go beyond their limitations.

Performance Curve analysis. Performance Curve analysis consists of examining an individual's performance on test items across a broad range of difficulty. Essentially, the respondent's average performance on test items is compared against average item difficulty, with the expectation that response accuracy will decrease as item difficulty increases (see Gudjonsson & Shackleton, 1986; McKinzey, Podd, Krehbiel, & Raven, 1999). Performance Curve analysis provides a highly detailed view of test performance that can be used to make inferences about the test-taker's intention and effort.

Rey (1941; see Frederick, 2002a) pioneered the use of Performance Curve analysis, comparing performance on easy tests (e.g., word recognition tests) to performance on more difficult tests (e.g., word recall tests). Frederick and Foster (1991) and Frederick et al. (1994) presented large-scale studies demonstrating the effectiveness of a Performance Curve strategy for identifying invalid responding; these studies spurred the initial publication of the VIP test.

Frederick et al. (2000) demonstrated that cross-classifications of effort and intention as identified by Performance Curve characteristics were supported by an analysis of concurrently administered malingering tests. Furthermore, using computer-generated protocols, they demonstrated that Performance Curve features used by the VIP to identify invalid responding were sensitive to manipulations designed to mimic invalid responding. The research described by Frederick et al. (2000) involved more than 700 criminal defendants and 6,000 computer-generated Performance Curves. VIP development and cross-validation were based on close to 1,500 participants and 10,000 randomly generated test protocols (see Chapters 3, 4, and 5).

The VIP test has been subjected to scrutiny through peer review. VIP development and cross-validation (including the potential rate for error) and demonstrations of construct validity have been published in peer-reviewed journals that enjoy general acceptance in the field of neuropsychology and psychological assessment (Frederick, 2002b; Frederick & Crosby, 2000; Frederick et al., 2000; Frederick & Foster, 1991; Frederick et al., 1994). The test has received generally favorable reviews (Gamache, 1998; Ivens, 2001; Ross & Adams, 1999) and comments (Vallabhajosula & van Gorp, 2001), with some accompanying criticisms, concerns, and cautions (Gamache, 1998; Gebart-Eaglemont, 2001; Lees-Haley, Dunn, & Betz, 1999; Ross & Adams, 1999).

VIP error rates and standards for controlling them are acceptable. The VIP classification accuracy rates reported in this manual are superior to those for other tests studied concurrently (see Table 12 in Chapter 5). The specificity rates are quite high, and the impact of false-positive classification is mitigated by the fourfold classification scheme and in-depth analysis of test performance. That is, rather than providing

a simple determination of "compliance" vs. "noncompliance," the VIP test provides a variety of indicators of intention and effort that the clinician can use to develop a more nuanced understanding of the examinee's approach to the assessment.

The VIP test has acceptance within the scientific community. The VIP test is peer-reviewed and incorporates widely accepted techniques for identifying feigning in cognitive assessment. Its error rates are acceptable. The VIP test is distributed worldwide and has been purchased in every state in the U.S., every province of Canada, and many countries outside North America.

In summary, the VIP test should meet whatever evidentiary standards are applicable when its findings are used in decision-making processes.