Review of the Validity Indicator Profile

Richard I. Frederick

U.S. Medical Center for Federal Prisoners

SUMMARY. The Validity Indicator Profile (VIP™; Frederick, 1997) is a two-alternative forced-choice test procedure intended to identify when the results of cognitive and neuropsychological testing may be invalid because of malingering or other problematic response styles. The test consists of 100 problems that assess nonverbal abstraction capacity and 78 word-definition problems. The VIP attempts to establish whether an individual's performance in an assessment battery should be considered representative of his or her true overall capacities (valid or invalid). This paper reviews the development and cross-validation of the VIP, and the paper discusses how demonstrations of the construct validity of the VIP support a conclusion that its use as described above fulfills *Daubert* evidentiary standards. The author discusses the benefits of using the VIP and identifies some potential challenges for using the VIP in making decisions about the response styles of individuals who are having their cognitive abilities assessed. *[Article copies available for a fee from The Haworth Document Delivery Service: 1-800-HAWORTH. E-mail address: <getinfo@haworthpressinc.com> Website: <http://www.HaworthPress.com>* *© 2002 by The Haworth Press, Inc. All rights reserved.]*

Richard I. Frederick is affiliated with the U.S. Medical Center for Federal Prisoners, Springfield, MO.

Address correspondence to: Richard I. Frederick, PhD, Department of Psychology, U.S. Medical Center for Federal Prisoners, Springfield, MO 65807 (E-mail: richardfrederick@yahoo.com).

The conclusions and opinions reported in this paper do not necessarily represent the views of the Federal Bureau of Prisons or the U.S. Department of Justice.

"VIP" is a trademark of NCS Pearson, Inc.

[Haworth co-indexing entry note]: "Review of the Validity Indicator Profile." Frederick, Richard I. Co-published simultaneously in *Journal of Forensic Neuropsychology* (The Haworth Medical Press, an imprint of The Haworth Press, Inc.) Vol. 2, No. 3/4, 2002, pp. 125-145; and: *Detection of Response Bias in Forensic Neuropsychology* (ed: Jim Hom, and Robert L. Denney) The Haworth Medical Press, an imprint of The Haworth Press, Inc., 2002, pp. 125-145. Single or multiple copies of this article are available for a fee from The Haworth Document Delivery Service [1-800-HAWORTH, 9:00 a.m. - 5:00 p.m. (EST). E-mail address: getinfo@haworthpressinc.com].

© 2002 by The Haworth Press, Inc. All rights reserved.

KEYWORDS. Malingering, neuropsychological assessment, psychological evidence, effort

OVERVIEW

The Validity Indicator Profile (VIP™; Frederick, 1997) is a measure of response validity which is intended to be administered concurrently within a battery of cognitive tests. The VIP consists of two subtests; each can be administered and scored separately. The VIP nonverbal subtest (VIP-NV) presents 100 picture-matrix problems that require simple matching, complex matching, analogous decision making, progression, addition, subtraction, and abstraction. The VIP verbal subtest (VIP-V) consists of 78 word definition problems. Test-takers are presented with a stimulus word (e.g., *carpet*) and are asked to choose one of two possible answers that is more similar in meaning to the stimulus (e.g., *rug* or *shoe*). For both subtests, the items have a hierarchy of difficulty but are presented randomly with respect to item difficulty. Once administered, the items are scored and then re-ordered by difficulty.

As its fundamental analysis of response validity, the items are re-ordered by difficulty, scored (0 = incorrect, 1 = correct), and plotted to generate a performance curve demonstrating the average performance of the test taker across an increasingly difficult range of test items. The plotted points of the performance curve are *running means*, computed by averaging a set of 10 consecutive scored item responses. Responses to items 1 through 10 are averaged to yield the first plotted running mean, representing the individual's average performance on the 10 easiest items. Responses 2 through 11 are averaged to compute the second plotted running mean. This process continues until the last (most difficult) item has been included in the plot of the last running mean. The performance curve is then a plot of the running mean on the vertical axis against the plot of its serial position (i.e., 1, 2, 3, and so on) on the horizontal axis.

Performance Curves Representing Compliance

For the two-alternative forced-choice format, individuals who exert full effort to respond correctly should generate perfect (i.e., running mean values = 1.0) or near-perfect (e.g., running mean = 0.9 or 0.8) performance within the range of their capacity to answer items correctly, transitioning to random responding (i.e., running mean values approxi-

mate 0.5) once the test taker has reached his/her ceiling of ability (see Figure 1). This means that performance curves for compliant test takers should be fairly similar in shape regardless of differences in ability.

Performance Curves Which Are Unexpected for Compliant Responding

Significant deviations from this expected curve have meaning and allow for some reasonable conclusions about the response style of the individual. For example, an individual who performs at 80% throughout most of the test (see Figure 2, Line A) may intend to respond correctly but may also be expending insufficient effort (*careless responding*). Such an inconsistent performance suggests the test taker could have performed perfectly on much easier items, given that he/she correctly solved 80% of the moderately difficult items. Another deviation from the expected curve might be for an individual who performs at about 50% throughout the entire test (Figure 2, Line B). Such an individual is probably marking answers without regard to item content (*irrelevant responding*). As a final example, consider an individual who demonstrates a consistent increase in correct responding as the test items be-

FIGURE 1. Expected shape of performance curve for compliant performance on the VIP. For the easiest items, perfect performance is expected for most individuals. As test item difficulty increases and the test-taker reaches a ceiling, accuracy drops off, but will remain at about 50%, because guessing on two-alternative items will result in about 50% correct responses.

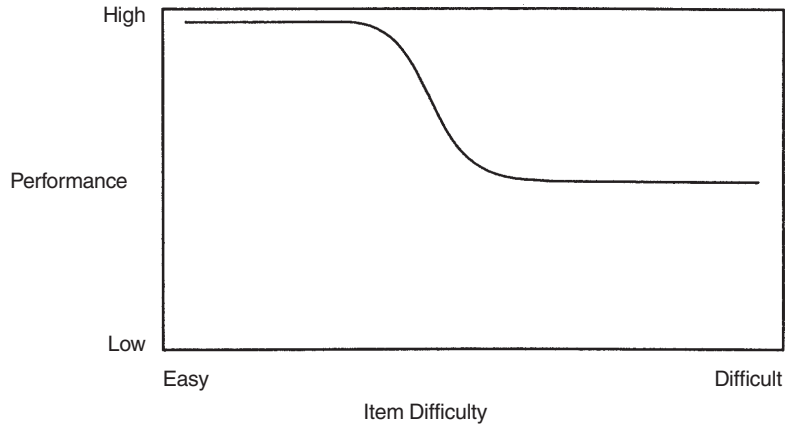
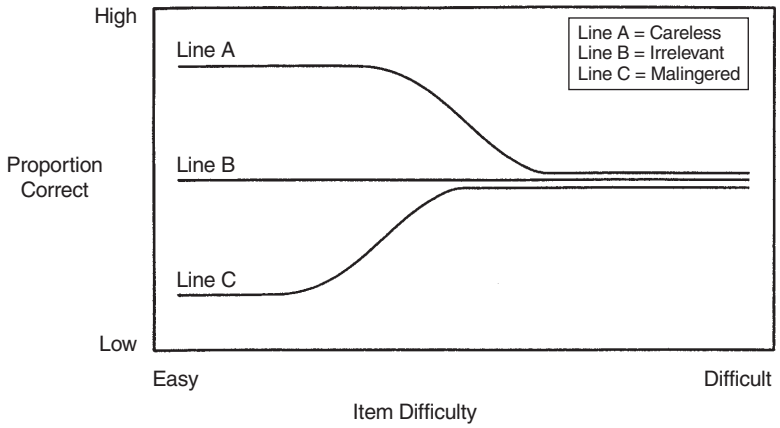


FIGURE 2. Unexpected performances for compliant performance on the VIP. Line A represents carelessness or inconsistent responding. The individual's imperfect responding for the easiest items is inconsistent with a similar level of performance accuracy for items of significantly greater difficulty. Line B represents irrelevant or random responding. Consistent performance accuracy at 50% indicates that responses were irrelevant to item content. Line C represents malingering. The improved performance as items become more difficult is best explained by intentional incorrect responding within the range of capacity to answer correctly and *improved* performance to 50% accuracy when the ceiling of ability is reached.



come more difficult (Figure 2, Line C). Such an individual is most likely intentionally choosing the wrong answer despite knowing the correct solution (*malingering*) and responding randomly only when the correct answer is not known.

In consideration of these four types of observed *performance curves*, Frederick (1997) postulated that the potential validity of performance on cognitive testing is a function of two test-taking characteristics: *motivation* and *effort*. Motivation refers to the intention of the test taker to perform well or poorly. Effort refers to the intensity of application of true ability to perform well or poorly (e.g., low effort or high effort). In this scheme, effort and motivation are independent constructs. Invalid assessments result from motivation to perform poorly or from poor effort when motivated to perform well.

This cross-classification of motivation and effort results in four *response styles*: compliant, careless, malingered, and irrelevant (see Figure 1). *Compliant* responding is characterized by *high effort* and *mo-*

tivation to perform well (alternatively, an *intention to respond correctly*). Compliant test-takers are cooperative with testing procedures, and their performance accurately represents their ability. *Careless,* or inconsistent, test-taking is also characterized by the *motivation to perform well*. However, it differs from compliant responding in that there is *incomplete effort* to respond correctly. Careless test-taking may result from inattention, distraction, or fatigue. *Malingering* is characterized by *high effort when motivated to perform poorly*, in which the test-taker strives to feign cognitive deficits in a convincing manner. Finally, *irrelevant*, or random, responding is characterized by *token effort when motivated to perform poorly* (alternatively, *no intention to respond correctly*). Irrelevant test-takers may be disengaged from the task of responding correctly, perhaps not caring about the outcome of the assessment.

Only Compliant performance curves are considered to represent Valid performances, providing strong evidence of the best effort of the test-taker to respond correctly. Careless, Irrelevant, and Malingering performance curves are considered to represent Invalid performances, each type of curve providing strong evidence of insufficient effort or motivation to perform poorly. The VIP was designed to generate Compliant, Careless, Irrelevant, and Malingering curves to represent, respectively, compliant responding, carelessness, irrelevant responding, and malingering.

DEVELOPMENT PHASE

Participants

The development sample included clinical and nonclinical groups. Participants in the nonclinical group (n = 944) were college students and employees of National Computer Systems, Inc. (NCS). Clinical participants (n = 104) were adults undergoing neuropsychological evaluation. Some were involved in litigation; others were being evaluated as part of ongoing clinical care. Compliant and Noncompliant groups were established based on random assignments for college students and NCS employees. Clinical participants were assigned to Compliant and Noncompliant groups based on a priori clinical judgments about the likelihood they would fake cognitive impairment or based on their performance on collaterally administered malingering tests. Methods of group assignments and participant instruction are reported in Frederick and Fos-

ter (1991), Frederick, Sarfaty, Johnston, and Powel (1994), Frederick (1997), and Frederick and Crosby (2000).

Classifying Valid and Invalid Performances

Performance on the VIP is first classified as Valid or Invalid by decision rules related to the consistency of responding correctly to test items across the test. Consistency is determined in several ways, each way examining in a different manner whether items of equivalent difficulty are answered correctly. Other decision rules evaluate whether the performance curve has a characteristic shape for compliant responding, or whether features of the curve deviate substantially from the shape shown in Figure 1. Frederick (1997) and Frederick and Crosby (2000) explain how these rules work. Decision rules for assignment to Valid or Invalid categories (Primary Validity Indicators [PVI]; Frederick, 1997) were established in the development sample and evaluated with the cross-validation sample. PVI decision rules were defined in the development phase so that the false-positive rate (equal to 1.0 minus specificity) would be approximately 10%. Test statistics were derived from an observation of the rate at which Compliant and Noncompliant participants were assigned to the VIP performance categorizations of Valid and Invalid, respectively. We attempted to maximize the sensitivity of the test while keeping a reasonably high bound on specificity.

CROSS-VALIDATION PHASE

Participants

The cross-validation sample consisted of normal adults, patients with traumatic brain injury, computer-generated cases of random responding, and patients at risk for malingering cognitive deficits. Normal adults were assigned to Compliant or Noncompliant criterion groups on an a priori basis. Compliant participants were instructed to complete the test, giving their best effort. Noncompliant normal participants were asked to feign believable brain impairment in completing tests. They watched a videotaped presentation of the nature of the research and a description of how brain injuries might affect individuals. Noncompliant normal participants were also provided with printed literature designed to help individuals understand the potential impact of mild brain injuries on behavior, mood, and thinking. They were offered meaning-

ful financial incentives for successful feigning. All participants completed the Portland Digit Recognition Test (Binder, 1993) and the Rey malingering tests[1] (Frederick, 1997).

Diagnostic Efficiencies

Classification rates for the VIP-NV and VIP-V are reported in Table 1. The VIP-NV demonstrated an overall classification rate of 79.8%, with 73.5% sensitivity and 85.7% specificity. The VIP-V demonstrated an overall classification rate of 75.5%, with 67.3% sensitivity and 83.1% specificity. Table 2 reports the rate of agreement between the VIP-NV and VIP-V ($n = 310$, kappa = .70, $p < .001$). Breakdowns of Invalid performances subclassified as Careless, Irrelevant, or Malingering are given for the VIP-NV and VIP-V in Table 3.

Most members of the Compliant criterion groups (98.1% or 98.8%) were classified as "motivated to respond correctly" (Compliant or Careless) on the VIP-NV and VIP-V, respectively. About half of the Noncompliant criterion group members (42.0% or 47.7%) were classified as "motivated to respond incorrectly" (Irrelevant or Malingering) on the VIP-V and VIP-NV, respectively. These values are very close to the specificity (99.5%) and sensitivity (48.5%) reported for the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992, p. 24).

Table 4 compares the classification rates of the VIP subtests with the Rey malingering tests and the PDRT. As reported in Frederick and Crosby (2000), two hierarchical logistic regression analyses were performed to compare the incremental validity of the PDRT with the VIP. For both analyses, the Rey tests were entered on the first step (model chi-square = 22.7, df = 3, $p < .0001$) and correctly classified 66.7% of the cases. On the second step, either the PDRT or VIP (VIP-NV and VIP-V) classifications were entered. On the third step, the remaining test was entered. When the VIP classifications were entered as the third step, added to the PDRT, correct classifications improved from 69.8% to 77.3% (improvement chi-square = 41.0, df = 2, $p < .0001$). In contrast, when the PDRT classifications were entered as the third step, added to the VIP, correct classifications improved only from 76.5% to 77.3% (improvement chi-square = 7.7, df = 1, $p < .01$). Hierarchical logistic regression analyses in which the Rey tests were entered as the third step were nonsignificant and demonstrated no improvement over the PDRT and VIP classifications, which together correctly classified 77.3% of the cases.

TABLE 1. Analysis of Test Classification for the Nonverbal and Verbal Subtests of the Validity Indicator Profile Cross-Validation Sample

	Test Classification			
	Nonverbal Subtest			
	Valid		Invalid	
Criterion Group	*n*	%	*n*	%
Compliant Group				
Honest Normals	93	93	7	7
Brain Injured	45	74	16	26
Noncompliant Group				
Coached Normals	13	25	39	75
Suspected Malingerers	27	55	22	45
Random Responding	0	0	50	100
	Verbal Subtest			
	Valid		Invalid	
Criterion Group	*n*	%	*n*	%
Compliant Group				
Honest Normals	95	95	5	5
Brain Injured	38	63	22	37
Noncompliant Group				
Coached Normals	14	27	38	73
Suspected Malingerers	35	73	13	27
Random Responding	0	0	50	100

© 1997 NCS Pearson, Inc. All rights reserved. Reproduced with permission.

CONSTRUCT VALIDATION OF THE VIP

Evidence of the construct validity of these subclassifications was reported in Frederick, Crosby, and Wynkoop (2000). They examined data for a large sample of criminal defendants who were completing court-ordered examinations of competency and criminal responsibility. Classifications of response style by VIP performance curve characteristics were supported by an analysis of concurrently administered Rey malingering tests when matched on essential VIP measures of test perfor-

TABLE 2. Cross-Tabulation of Validity Indicator Profile Verbal and Nonverbal Subtest Classification Rates

	Verbal Subtest	
Nonverbal Subtest	Invalid	Valid
Invalid	157	21
Valid	25	107

Note. $n = 310$. Kappa = .70, $p < .001$.
© 1997 NCS Pearson, Inc. All rights reserved. Reproduced with permission.

mance to control for potential effects from cognitive capacity. Large to very large effect sizes on malingering test performance were seen for individuals classified as "motivated to perform well" (Compliant and Careless classifications) when compared to individuals classified as "motivated to perform poorly" (Irrelevant and Malingering). Moderate to large effect sizes on malingering test performance were seen for differences in motivation even when the effort of subjects was presumed to be low (Irrelevant vs. Careless performance curves). There were zero-order effects on malingering test performance when Compliant test takers were compared with those classified as Careless (i.e., both classes presumed motivated to perform well, but different in effort expended), but defendants classified as Careless had significantly higher scores on MMPI-2 carelessness indicators when compared to defendants classified as Compliant.

Furthermore, Frederick, Crosby, and Wynkoop (2000) hypothesized a mechanism of careless responding and simulated that process by computer for 4000 Compliant curves. The simulation of careless responding resulted in the predicted changes to the performance curve features which are used to describe responding as Careless.

DISCUSSION

Considerations in Light of Daubert

As noted above, reports on the VIP development and cross-validation (including the potential rate for error) and assessment of construct validity have been published in peer-reviewed journals which enjoy general acceptance in the field of neuropsychology and psychological

TABLE 3. Analysis of Response Style Classification of Invalid Performances on the Validity Indicator Profile Cross-Validation Sample

	Response Style Classification					
	Nonverbal Subtest					
Criterion Group	Careless		Irrelevant		Malingered	
	n	%	*n*	%	*n*	%
Compliant Group						
Honest Normals	7	7	0	0	0	0
Brain Injured	14	23	2	3	0	0
Noncompliant Group						
Coached Normal	20	38	16	31	3	6
Suspected Malingerers	18	37	4	8	0	0
Random Responding	1	2	48	96	1	2
	Verbal Subtest					
Criterion Group	Careless		Irrelevant		Malingered	
	n	%	*n*	%	*n*	%
Compliant Group						
Honest Normals	5	5	0	0	0	0
Brain Injured	19	32	3	5	0	0
Noncompliant Group						
Coached Normal	26	50	8	15	4	8
Suspected Malingerers	11	23	2	4	0	0
Random Responding	1	2	48	96	1	2

© 1997 NCS Pearson, Inc. All rights reserved. Reproduced with permission.

assessment. 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; Ross & Adams, 1999). Using performance curve characteristics as a means of examining the potential for response invalidity has received endorsement for cognitive tests other than the VIP (Gudjonsson & Shackleton, 1986; McKinzey, Podd, Krehbeil, & Raven, 1999; Rogers, Harrell, & Liff,

TABLE 4. Comparison of Classification Rates Among Malingering Tests Within Cross-Validation Sample

Test	Sensitivity	Specificity	Overall Classification Rate
VIP Nonverbal Subtest	73.5	85.7	79.8
VIP Verbal Subtest	67.3	83.1	75.5
Portland Digit Recognition Test	17.0	99.4	67.7
Rey 15-Item Memory Test	4.9	97.5	61.4
Rey Word Recognition Test	8.8	100.0	64.6
Rey Dot Counting Test	11.8	97.5	64.1

© 1997 NCS Pearson, Inc. All rights reserved. Reproduced with permission.

1993). Tatsuoka and Tatsuoka (1982, 1983) demonstrated the use of consistency measures used in the PVI decision rules to examine consistency of performance in cognitive tests. The VIP also compares proportion of total correct responses to that expected for no ability as an indicator of test-taking intention. This technique has received wide general acceptance as an indication of response validity for some time (e.g., Brady & Lind, 1961; Pankratz, 1979; Rogers et al., 1993; Theodor & Mandelcorn, 1973). These factors suggest the VIP should easily surpass scrutiny in a *Daubert* hearing concerning its use as a measure of response validity for cognitive assessment.

Criticism of the VIP

Lees-Haley, Dunn, and Betz (1999) criticized the VIP process of subclassification, stating: "This further classification confuses the issues by appearing to exclude irrelevant and careless performances from the definition of malingering, when, in fact, sub-optimal performances are specifically associated with malingering" (p. 12). Actually, the subclassifications of the VIP were never intended to speak definitively about the *ultimate goal* of the test-taker in the examination process. For example, the VIP manual (Frederick, 1997) includes this statement: "It is important to remember that the VIP classification of test performance and response style is made in the absence of any other information about the test taker (p. 37)." Similar statements appear routinely in the interpretive report, for example:

> The descriptions, inferences, and recommendations in this report should be evaluated in the context of other information that is available about the client. Such information may include clinician observations, the client's presenting complaint, scores from other assessment instruments, and an evaluation of situational factors that may influence test performance.

Subclassifications are intended to accurately identify the *intention* of the test taker in completing concurrently administered cognitive tests (i.e., "to respond correctly" or "to respond incorrectly") and to accurately categorize the *level of effort* produced by the examinee to fulfill the intention (i.e., "high effort" or "low effort"). A decision to use the term "Malingering" for only those *performance curves* obviously resulting from a strong effort to answer incorrectly certainly does not exclude suboptimal performance from the definition of a malingering *response style*. The criticism by Lees-Haley and his colleagues does not comport with statements about the process of interpretation in the VIP manual, for example, "An individual who intends to perform poorly and who make a conscious effort at deception might produce results that appear to be careless or irrelevant rather than malingered. This can occur, for example, if a malingerer answers the easiest items correctly" (p. 38). Even Compliant performance curves routinely contain this language in the interpretive report:

> Performances classified as valid are assumed to represent compliance. Individuals classified as compliant demonstrate a predictable decrement in performance as the items increased in difficulty. However, performance that follows such a pattern does not necessarily ensure that the test-taker made his or her best effort. The clinician should make that determination based on all available information.

Again, the use of Malingering as a classificatory label for pronounced positively-sloped performance curves does not indicate at all that malingerers must not instead produce Irrelevant, Careless, or even Compliant curves. For example, the interpretive report for Careless curves routinely contain this statement:

> Careless responding can result from many factors, the most probable of which is an insufficient effort to do well. Occasionally, however, Performance Curves that are classified as careless can result

> from an individual's deliberate attempt to do poorly but not poorly enough to be classified as a malingerer or as an irrelevant responder.

Ostensibly, Lees-Haley and his colleagues were concerned that conclusions that examinees had feigned aspects of their presentation or exaggerated their putative impairment might be foiled by VIP results of Irrelevant or Careless, or even, in some circumstances, a finding of Compliant. No doubt a sharp attorney will highlight seemingly contradictory information to counter a clinical assessment which results in the conclusion an examinee has malingered or exaggerated, but examiners who are familiar with the contents of the VIP manual and interpretive report should have no difficulty explaining that there is no inherent conflict in a finding of malingering in light of the totality of the assessment and the classification of Irrelevant, Careless, or even Compliant on the VIP.

Dichotomous Classifications vs. Four-Fold Classifications

The criticism of the VIP by Lees-Haley et al. (1999) was contained within a review of the Victoria Symptom Validity Test (VSVT; Slick, Hopp, Strauss, Hunter, & Pinch, 1994). Lees-Haley et al. claimed the dichotomous classifications of the VSVT and its ilk were more helpful than the four-fold scheme presented by the VIP. Most classification schemes for cognitive/memory malingering tests result in labels of "malingered" and "not malingered" or "compliant" and "non-compliant" (Binder, 1993; Tombaugh, 1997). Frederick and Crosby (2000) identified four problems with the dichotomous classification scheme.

First, the dichotomous classification puts the clinician in danger of making costly classification errors. A false positive classification means a person who is not faking impairment is classified as "faking." A person who is faking is classified as "compliant." Second, most tests designed to detect malingering that use dichotomous classifications provide no demonstrable support for the classification beyond its statistical or clinical atypicality. As Frederick (2000a) noted, one can always argue that scores worse than cut-scores established for individuals with bona fide impairment mean only that the impairment of the examinee is worse than that of the impaired individuals in the research comparison group. When examinees earn compliant scores on such tests, we cannot conclude that they displayed good effort, but only that they did not display bad effort (Faust & Ackley, 1998). Third, dichotomous malin-

gering classifications limit explanations of positive test scores to noncompliance or classification error. A classification scheme should be able to describe a performance as suboptimal (not representative of the individual's maximal capacity to perform well on cognitive tasks) without mechanically concluding that the invalidity was intentional. Fourth, the dichotomous classification scheme makes research into malingering difficult. It has proven difficult, if not nearly impossible, to develop "pure" groups of definite malingerers and fully compliant test takers, even within analog settings (Arbisi & Ben-Porath, 1995; Frederick, 2000b; Frederick et al., 1994; Greiffenstein, Baker, & Gola, 1994, 1996).

Potential Classification Errors

Most members of the Compliant criterion groups in the cross-validation phase were classified as "motivated to perform well." Classification errors for Compliant participants were generally restricted to the classification of Careless and not as "motivated to perform poorly" (i.e., Malingering or Irrelevant). These findings illustrate that potential false positive classifications are mitigated by the four-fold classification scheme. Excessively careless, inconsistent performance will infrequently result in a classification as Irrelevant, but will not result in a classification of Malingering. Such misclassification will accurately reflect the compromised effort that went into generating correct answers, but will incorrectly raise the hypothesis of motivation to perform poorly. Additionally, irrelevant responding will result in false positive classification as Malingering only about 2% of the time (see Table 3).

About half of the noncompliant criterion group members were classified as "motivated to perform poorly." Potential false negative classifications were mitigated by the four-fold classification scheme as well. Truly irrelevant performances were incorrectly categorized as Careless only about 2% of the time (see Table 3). Incorrect classifications of malingering, in which the test taker is making a concerted effort to answer incorrectly, will generally be classified as Irrelevant, which maintains the hypothesis that the test taker did not intend to answer correctly.

Table 3 indicates that some malingerers may be able develop a strategy of correctly answering some easier items so as to generate a classification as Compliant or Careless, false negative classifications. To do so requires *consistent, above-chance* responding for at least the easiest items. Frederick and Crosby (2000) reviewed the nature of the performances of suspected malingerers and coached normals classified as Compliant. These putatively Noncompliant participants clearly gener-

ated strong effort to respond correctly on the VIP-NV and VIP-V, with average total correct responses at the same level as Compliant coached normals and with performance curves demonstrating sustained effort to respond correctly. Consequently, it is hard to construe these performances as misclassified Invalid performances, and it is more likely these are instances of criterion group contamination and a source of underestimated VIP sensitivity.

When malingerers answer some items correctly as a means of trying not to be obvious, they run the risk of demonstrating too much ability to generate the gain that impairment would produce or generating performance curves which reveal obvious inconsistency or carelessness. Because the VIP can usually generate an estimate of cognitive capacity for Careless and Compliant performance curves (Crosby & Frederick, 1997; Frederick, 1997), the clinician can also compare that estimate with other estimates obtained during the examination. Such a comparison may shed light on the intentions and effort of the test-taker.

Identifying Potential Classification Errors on the VIP

In reviewing VIP performances, it is useful to review which response styles actually generated which performance curve classifications in the cross-validation phase.

Motivation to Perform Poorly

Malingering. The classification of Malingering was primarily restricted to coached normals asked to feign impairment. Two percent of randomly-generated curves were classified as Malingering.

Irrelevant. The classification of Irrelevant responding was primarily restricted to participants in the Noncompliant group. About one-third of coached normals generated this type of curve. Almost all (96%) of randomly-generated curves resulted in the classification of Irrelevant. A small percentage of suspected malingerers (4%) and Compliant participants with bona fide traumatic brain injury (3%) received a classification of Irrelevant for their performances.

These findings highlight the very high specificity of the VIP with respect to distinguishing between "motivation to perform well" and "motivation to perform poorly." When these categorizations result, the clinician can confidently postulate there was no intention to respond correctly or that severe impairment exists.

Motivation to Perform Well

Compliant. Almost all Compliant normal control participants generated Compliant curves. Most Compliant participants with traumatic brain injury were classified as Compliant, but about one-third generated curves classified as Careless. A large number of Noncompliant participants (about one-fourth of coached normals and over one-half of suspected malingerers) generated strong effort to respond correctly on the VIP-NV and VIP-V, with average total correct responses at the same level as Compliant coached normals and with performance curves demonstrating sustained effort to respond correctly. It is unlikely these individuals were trying to feign impairment.

Careless. Very few Compliant coached normals and about one-third of compliant participants with traumatic brain injury were classified as Careless. About one-half of Noncompliant coached normals and a sizable minority of suspected malingerers were classified as Careless.

COMMENTS

I have three comments. The first two concern questions I routinely receive from consumers of the VIP, and the third comment concerns my belief that research on malingering tests, including my own, is frequently plagued by criterion group contamination.

Careless Curves

As I have just noted in my discussions on classification errors, the Careless categorizations of the cross-validation sample included a mix of Compliant and Noncompliant participants. The nature of these results clearly indicate that the potential for false positive errors on the VIP is highest for this categorization, especially for persons with traumatic brain injury. Users who carefully read the test manual and interpretive reports know that the classification of Careless is the default classification of Invalid performances (i.e., failure of two or more Primary Validity Indicators; PVIs) that cannot be classified as Malingered or Irrelevant. The interpretive report notes that sometimes lucky or unlucky guessing of answers for items beyond the test-taker's ceiling can unduly influence the PVIs to generate a classification of Invalid. This means that curves can receive default classification as Careless if they

are classified first as Invalid, *even if the curves have no characteristics associated with careless or inconsistent responding*.

Given that Frederick, Crosby, and Wynkoop (2000) demonstrated that certain features of the performance curve associated with carelessness[2] were associated with other indicators of carelessness, my caveat for inquiries is straightforward. If the curves do not contain either of these features, carelessness or invalidity must be *ruled in* from other sources of information. I urge caution to avoid over-interpreting these curves. If the curves do contain either of these features, however, consider that some sort of inconsistency in responding *is likely*, and should be *ruled out*. Given that a sizeable proportion of individuals with traumatic brain injury generated these sorts of curves, one potential source of inconsistency has nothing to do with motivation to perform poorly, but may reflect unique contributions of certain clinical disorders. Again, I urge clinicians to review the totality of information in making conclusions about VIP performance; I strongly recommend that conclusions about poor effort or motivation to perform poorly reflect a convergence of evidence. Future research on the VIP should attempt to further distinguish among careless categorizations that result from sophisticated malingering, those that result from true carelessness, and those that result from clinical conditions which inhibit concentration or mimic careless responding on the VIP.

Use of the Test with Individuals Who Have Mental Retardation

I am commonly asked whether the test can be used to evaluate individuals *suspected* of mental retardation. My answer is, "Of course." This question primarily arises because of statements on page 4 of the VIP manual:

> The VIP is not recommended for use with individuals with historically demonstrable mental retardation or severe cognitive impairment. Individuals with severe impairment will probably receive a categorization of invalid test performance because of their extremely limited cognitive abilities.

I have evaluated my share of individuals who feign severe mental impairment. Such individuals generally produce Malingered or Irrelevant performance curves. The VIP manual reports a sample of 40 individuals with historically demonstrable mental retardation who were in assisted living in the community. About half of them were fully illiterate. Only

5% of the performances of these individuals resulted in classification as Compliant. Sixty-five percent of the VIP-NV curves and about 50% of the VIP-V curves were classified as Careless, and most of the rest were classified as Irrelevant. One of the 40 individuals produced a VIP-V curve classified as Malingering, consistent with the approximate 2% false positive rate for random responding. The question arises, "How can you distinguish between Irrelevant or Careless curves generated by fakers and those generated by persons with severe impairment?" My answer is, "I can't." I have to make decisions about motivations and effort in light of all the evidence I possess. If I am evaluating an individual with "historically demonstrable mental retardation or [continuing] severe cognitive impairment," I have no use for the VIP or any other instrument which measures response style.

Criterion Group Contamination

In recent research on the VIP (Frederick, Crosby, & Wynkoop, 2000), I have strived to avoid criterion group methodology. Instead, I have focused on comparing performance curve characteristics suggesting different response sets and the accompanying effect on concurrently administered tests of response style. As we noted in that article, it was clear the cross-validation Noncompliant group contained a large proportion of individuals who had completed the VIP compliantly. We had no choice but to include them in our computations of sensitivity, but it was clear to us this resulted in significant underestimations of sensitivity.

In Frederick (2000b), I claimed that traditional criterion group validation of malingering tests is continually seriously flawed by criterion group impurity, resulting in significant underestimations of test diagnostic efficiencies (both sensitivity and specificity). Despite the contention of Rogers (1997) that differential prevalence designs are useless in malingering research, I demonstrated that, with appropriate methodology and analysis, the differential prevalence design overcomes the problem of criterion group impurity and allows for potentially more accurate estimates of sensitivity and specificity. For example, in that paper, I showed that using a differential prevalence design for the Rey 15-Item Test resulted in far superior estimates of sensitivity and specificity than commonly reported in the literature about that procedure. I do not think the problem of criterion group contamination is unique to the VIP, and I would encourage other researchers to attempt validations of validity instruments in ways that do not depend on criterion group designs.

I believe the VIP has made an important contribution in attempting to go beyond simplistic thinking with respect to response style. There are many response styles that contribute to observed performance on cognitive tests. Currently, some response styles cannot be effectively distinguished based solely on test performance; for example, random completion of a test because one refuses to engage in the task cannot currently be distinguished from a random completion because one does not know any answer. The goal for future research will be to identify practical methods for fully distinguishing between cooperative and noncooperative performances on cognitive tests.

NOTES

1. The "Rey malingering tests" refer to the most commonly known of Rey's tools for assessment of malingering, the 15-Item Memory Test, the Word Recognition Test, and the Dot Counting Test (Lezak, 1995; Rey, 1941, 1958).

2. A discussion of these features requires a good deal of explanation beyond the scope of this review. For the interested reader, these characteristics are a large Sector 1 residual and Sector 2 distance greater than or equal to Sector 1 distance (see Frederick, 1997; Frederick & Crosby, 2000; or Frederick, Crosby, & Wynkoop, 2000).

REFERENCES

Arbisi, P. A., & Ben-Porath, Y. S. (1995). An MMPI-2 infrequent response scale for use with psychopathological populations: The Infrequency-Psychopathology Scale, *F(p)*. *Psychological Assessment*, *7*, 424-431.

Binder, L. M. (1993). Assessment of malingering after mild head trauma with the Portland Digit Recognition Test. *Journal of Clinical and Experimental Neuropsychology*, *15*, 170-182.

Brady, P. B., & Lind, D. L. (1961). Experimental analysis of hysterical blindness. *Archives of General Psychiatry*, *4*, 331-339.

Crosby, R., & Frederick, R. I. (1997, August). Predicting IQ ranges from performance curve characteristics. Paper presented at the Annual Convention of the American Psychological Association, Chicago.

Faust, D. & Auckley, M. A. (1998). Did you think it was going to be easy? Some methodological suggestions for the investigation and development of malingering detection techniques (pp. 1-54). In C. R. Reynolds (Ed.). *Detection of Malingering During Head Injury Litigation*. New York: Plenum.

Frederick, R. I. (1997). *Validity Indicator Profile manual*. Minnetonka, MN: NCS Assessments.

Frederick, R. I. (2000a). A personal floor effect strategy to evaluate the validity of performance on memory tests. *Journal of Clinical and Experimental Neuropsychology*, *22*, 720-730.

Frederick, R. I. (2000b). Mixed group validation: A method to address the limitations of criterion group validation in research on malingering detection. *Behavioral Sciences and the Law*, *18*, 693-718.

Frederick, R. I., & Crosby, R. D. (2000). Development and validation of the Validity Indicator Profile. *Law and Human Behavior*, *24*, 59-82.

Frederick, R. I., Crosby, R. D., & Wynkoop, T. F. (2000). Performance curve classification of invalid responding on the Validity Indicator Profile. *Archives of Clinical Neuropsychology*, *15*, 281-300.

Frederick, R. I., & Foster, H. G. (1991). Multiple measures of malingering on a forced-choice test of cognitive ability. *Psychological Assessment*, *3*, 596-602.

Frederick, R. I., Sarfaty, S. D., Johnston, J. D., & Powel, J. (1994). Validation of a detector of response bias on a forced-choice test of nonverbal ability. *Neuropsychology*, *8*, 118-125.

Gamache, M. (1998, Spring). The Validity Indicator Profile: Review of an instrument to assess response style. *American Psychology-Law Society Newsletter*, *18(2)*, 11-43.

Gebart-Eaglemont, J. E. (2001). Review of the Validity Indicator Profile. In B. S. Plake and J. C. Impara (Eds.). *The Fourteenth Mental Measurements Yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1994). Validation of malingered amnesia measures in a large clinical sample. *Psychological Assessment*, *6*, 218-224.

Greiffenstein, M. F., Baker, W. J., & Gola, T. (1996). Comparison of multiple scoring methods for Rey's malingered amnesia measures. *Archives of Clinical Neuropsychology*, *11*, 283-293.

Gudjonsson, G. H., & Shackleton, H. (1986). The pattern of scores on Raven's Matrices during "faking bad" and "non faking" performances. *British Journal of Clinical Psychology*, *25*, 35-41.

Ivens, S. H. (2001). Review of the Validity Indicator Profile. In B. S. Plake and J. C. Impara (Eds.). *The Fourteenth Mental Measurements Yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

Lees-Haley, P. R., Dunn, J. T., & Betz, B. P. (1999, Fall). Test review: The Victoria Symptom Validity Test. *American Psychology-Law Society Newsletter*, *19(3)*, 12-16.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford.

McKinzey, R. K., Podd, M. H., Krehbiel, M. A., & Raven, J. (1999). Detection of malingering on the Raven's Standard Progressive Matrices: A cross-validation. *British Journal of Clinical Psychology*, *38*, 435-439.

Pankratz, L. (1979). Symptom validity testing and symptom retraining: Procedures for the assessment and treatment of functional sensory deficits. *Journal of Consulting and Clinical Psychology*, *47*, 409-410.

Rey, A. (1941). L'examen psychologie dans les cas d'encephalopathie traumatique. *Archives de Psychologie*, *28*, 286-340.

Rey, A. (1958). *L'Examen clinique de psychologie*. Paris: Presses Universitaires de France.

Rogers, R. (Ed., 1997). *Clinical assessment of malingering and deception* (2nd ed.). New York: Guilford Press.

Rogers, R., Bagby, R. M., & Dickens, S. E. (1992). *Structured Interview of Reported Symptoms: Professional Manual*. Odessa, FL: Psychological Assessment Resources.

Rogers, R., Harrell, E. H., & Liff, C. D. (1993). Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. *Clinical Psychology Review*, *13*, 255-274.

Ross, S. R., & Adams, K. M. (1999). Test review: One more test of malingering? *Clinical Neuropsychologist*, *13*, 112-116.

Slick, D., Hopp, G., Strauss, E., Hunter, M., & Pinch, D. (1994). Detecting dissimulation: Profiles of simulated malingerers, traumatic brain-injury patients, and normal controls on a revised version of Hiscock and Hiscock's forced-choice memory test. *Journal of Clinical and Experimental Neuropsychology*, *16*, 472-481.

Tatsuoka, K. M., & Tatsuoka, M. M. (1982). Detection of aberrant response patterns and their effect on dimensionality. *Journal of Educational Statistics*, *7*, 215-231.

Tatsuoka, K. M., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the Individual Consistency Index. *Journal of Educational Measurement*, *20*, 221-230.

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*, 260-268.

Theodor, L. H., & Mandelcorn, M. S. (1973). Hysterical blindness: A case report and study using a modern psychophysical technique. *Journal of Abnormal Psychology*, *82*, 552-553.

Vallabhajosula, B., & van Gorp, W. G. (2001). Post-Daubert admissibility of scientific evidence on malingering of cognitive deficits. *Journal of the American Academy of Psychiatry and the Law*, *29*, 207-215.

Copyright of Journal of Forensic Neuropsychology is the property of Haworth Press, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.