2003

Detection of feigned mental disorders - A meta-analysis of the MMPI-2 and malingering

R Rogers

KW Sewell

MA Martin

MJ Vitacco

Recommended Citation
Rogers, R; Sewell, KW; Martin, MA; and Vitacco, MJ, "Detection of feigned mental disorders - A meta-analysis of the MMPI-2 and malingering" (2003). *ASSESSMENT*. 124.
10.1177/1073191103252349

Detection of Feigned Mental Disorders

A Meta-Analysis of the MMPI-2 and Malingering

Richard Rogers
Kenneth W. Sewell
Mary A. Martin
University of North Texas
Michael J. Vitacco
University of Massachusetts

The validity of test data from multiscale inventories is dependent on self-reports that may be easily distorted by malingering. In examining the Minnesota Multiphasic Personality Inventory–2's (MMPI-2) role in the assessment of feigning, this review provides a conceptual analysis of the detection strategies underlying the MMPI-2 validity scales. The conceptual analysis is augmented by comprehensive meta-analysis of 65 MMPI-2 feigning studies plus 11 MMPI-2 diagnostic studies. For the rare-symptoms strategy, Fp (Cohen's d *= 2.02) appears especially effective across diagnostic groups; its cut scores evidence greater consistency than most validity indicators. The data supported the F as an effective scale but questioned the routine use of Fb. Among the specialized scales, Ds appeared especially useful because of its sophisticated strategy, consistent cut score, and minimal false-positives. General guidelines are offered for specific MMPI-2 validity scales in the assessment of malingering with specific diagnoses.*

Keywords: malingering; MMPI-2; overreporting; feigning; detection strategies

The Minnesota Multiphasic Personality Inventory–2 (MMPI-2) is the most extensively researched psychological measure of feigned mental disorders. Several dozen investigations have examined the effects of feigning, primarily under analogue conditions, with comparisons of simulators to mentally disordered samples. These studies are heterogeneous, reflecting important differences in feigning indexes, types of feigned disorders, and simulation designs.

Meta-analyses with the MMPI (Berry, Baer, & Harris, 1991) and the MMPI-2 (Rogers, Sewell, & Salekin, 1994) have catalogued the range of available feigning indexes. In many cases, individual investigators have proliferated new indexes with apparently little attention to the underlying detection strategies. The next section reviews MMPI-2 feigning indexes with respect to their implicit detection strategies.

MMPI-2 DETECTION STRATEGIES

Rogers (1997) outlined detection strategies relevant to malingering on the MMPI-2 that were tested with multiple measures across both simulation designs and known-group comparisons. In particular, MMPI-2 feigning indexes use the following strategies: (a) rare symptoms, (b) symptom severity, (c) obvious versus subtle symptoms, and (d) symptom selectivity. Additional strategies have also been implemented, most notably erroneous stereotypes (Gough, 1954; Rogers & Bender, in press).

A robust detection strategy for feigned mental disorders is the use of *rare symptoms*. Rare symptoms refer to symptoms, characteristics, or associated features of impaired functioning that occur very infrequently in genuinely impaired populations. On the MMPI-2, rare symptoms might be defined as "atypical characteristics as-

TABLE 1
Descriptive Data on MMPI-2 Feigning Indexes

Scale	*Items*	*% True*	*Development*	*Detection Strategy*	r *With F* [a]
F	60	68.3	Normative	Rare symptoms	—
Fb	40	92.5	Normative	Rare symptoms	.86/.59
Fp	27	66.7	Discriminant	Rare symptoms	.75/.57
Ds	58	82.8	Discriminant	Erroneous stereotypes	.84/.61
Dsr	32	81.3	Discriminant	Erroneous stereotypes	—
LW	107	72.9	Content	Symptom severity	.84/.67
O-S	253	46.2[b]	Rational	Obvious vs. subtle	.81/.58
FBS	43	41.9	Rational-discriminant	Erroneous stereotypes	—

NOTE: Normative = uncharacteristic responses based on norms; discriminant = empirically derived items that differentiate between feigning and honest responding; content = nominated by clinical psychologists as representing a specific content area of psychological concerns; rational = heuristic division of items (obvious and subtle); rational-discriminant = rational selection of items taking into account differences between criterion groups. F = Infrequency; Fb = Back Infrequency; Fp = Infrequency-Psychopathology; Ds = Dissimulation; Dsr = Dissimulation–Revised; LW = Lachar-Wrobel; O-S = T score difference of Obvious-Subtle; FBS = Fake Bad Scale.
a. Derived from Greene (2000, p. 66): First correlations are based on 50,966 patients (Caldwell, 1998), whereas second correlations are based on the normative sample (Butcher, Dahlstrom, Graham, Telelgen, & Kaemmer, 1989).
b. Obvious items = 61.4% true; subtle items = 25.9% true.

sociated with psychopathology or impairment that are not commonly endorsed by clinical populations." The implicit logic of rare symptoms is that malingerers are unlikely to differentiate very infrequent symptoms from their more common counterparts.

The rare-symptoms strategy is used by the following MMPI-2 feigning indexes: F (Infrequency), Fb (Back Infrequency), and Fp (Infrequency-Psychopathology). As reported in Table 1, Fb is particularly vulnerable to yea-saying with 92.5% "true" responses. Strictly speaking, the development of F and Fb was flawed from a rare-symptoms perspective because their development involved only normative samples of presumably unimpaired participants. Items that are rare in a normative sample may be more common in a clinical population. As a case in point, 15 or more F items are endorsed by 25% of clinical samples (Greene, 1997). The development of Fp (Arbisi & Ben Porath, 1995) sought to remedy this oversight by identifying symptoms rarely endorsed by genuine patients. As a possible complication, Fp includes four infrequent items from Scale L (Lie); whether their inclusion impedes interpretation is worthy of further investigation (see Gass & Luis, 2001).

A second detection strategy examines *symptom severity*. Symptom severity considers the number of potentially disabling symptoms and characteristics endorsed by genuine patients versus malingerers. This strategy is operationalized on the MMPI-2 in the form of "critical items." The implicit strategy is based on the premise that some malingerers will not take into account symptom severity and will endorse an unexpectedly high number of critical items. Most MMPI-2 malingering research is based on the Lachar and Wrobel (1979) critical items (i.e., LW), representing 14 areas of psychological concern.

A third detection strategy involves the comparison of *obvious and subtle symptoms*. Obvious symptoms refer to items clearly indicative of major psychopathology, whereas subtle symptoms refer to those not typically recognized as such by nonprofessionals. The implicit strategy capitalizes on malingerers' tendency to recognize and endorse more obvious than subtle symptoms.[1] Although several methods have been tested (Greene, 2000), current research has focused on the Wiener and Harmon obvious-subtle subscales (Wiener, 1948). A potential limitation of this strategy is the difficulty in selecting subtle symptoms that are relevant to mental disorders but appear to be unrelated. On this point, Bagby, Nicholson, and Buis (1998) marshaled data in support of using obvious symptoms alone. However, most research has continued to focus on the relationship between obvious and subtle symptoms.

Beyond the Rogers (1997) detection strategies for feigned mental disorders, the MMPI and MMPI-2 use an innovative strategy, namely, *erroneous stereotypes*. Gough (1954) identified MMPI items, based on common misperceptions about neuroticism and maladjustment, that were inaccurately perceived by both professionals and nonprofessionals. These items cover a broad content including somatic complaints, dysphoria, discontent about childhood, sexual conflicts, and bizarre ideation. The implicit strategy rests on the inability of malingerers to differentiate erroneous stereotypes from genuine psychopathology. On the MMPI-2, Gough's dissimulation scale (Ds) and an abbreviated version (Ds–Revised or Dsr) employ erroneous stereotypes. Although originally developed to examine feigned neurosis, these scales have utility with a wide range of disorders. Beyond Gough's work, Lees-Haley, English, and Glenn (1991) developed the Fake-Bad Scale (FBS) to assess erroneous stereotypes and

atypical symptoms specifically related to personal injury cases.

Several potential detection strategies have yet to be rigorously tested. For example, Greene (1997) has proposed a bipolarity hypothesis with malingering and defensiveness (i.e., marked underreporting or denial of psychopathology) representing opposite poles. If correct, malingerers could potentially be identified by the *absence of defensiveness*. The implicit strategy is that malingerers will focus on the production of bogus symptoms and remain incognizant of the need to report some characteristics of defensiveness. An early MMPI-2 study by Graham, Watts, and Timbrook (1991) found suppressed scores on K for both male (M = 35.8T) and female (M = 32.7T) simulators. Another potential MMPI-2 strategy, successful with other measures, is *symptom selectivity*. The implicit strategy is based on the notion that some malingerers will indiscriminately endorse items associated with psychopathology. Problems with symptom selectivity are likely to be reflected in extreme profile elevations (Dahlstrom, Welsh, & Dalhstrom, 1972). Recently, Wetter and Deitsch (1996) found that simulators of post-traumatic stress disorder (PTSD) produced extreme profile elevations for both original (M = 84.72) and retest (M = 80.67) administrations. Both absence of defensiveness and symptom selectivity require further investigation as potential MMPI-2 detection strategies.

PREVIOUS META-ANALYSES AND THE CURRENT STUDY

Berry et al. (1991) performed the first malingering meta-analysis that was based on the original MMPI. Their review compiled 28 studies representing a broad array of nonclinical and clinical samples. Unfortunately, more than one third of these studies did not include clinical samples, thereby limiting the relevance of their findings. In general, Berry et al. (1991) found the largest effect sizes for F, Ds, and F-K. The most effective cut scores for MMPI feigning indexes were difficult to establish because studies varied so widely in their proposals.

Fundamental changes between the MMPI and the MMPI-2 necessitated a reevaluation of validity indexes for feigning. Rogers et al. (1994) examined 14 MMPI-2 feigning studies. As a modification of the Berry et al. (1991) design, effect sizes for feigned versus patient samples were calculated separately. Very large ($d \geq 1.75$) effect sizes[2] were found for F, F-K (raw score difference of Infrequency-Correction), and O-S (T score difference of Obvious-Subtle), paralleling Berry et al. for the first two estimates. Insufficient studies reported Ds, but effect sizes for Dsr were large (i.e., mean d = 1.54). Like Berry et al., cut scores were widely scattered across studies. For example, cut scores derived from individual studies for F ranged markedly from 8 to 29.

The current study is designed to update the Rogers et al. (1994) meta-analysis and improve its methodology. In the past 8 years, the number of MMPI-2 malingering studies has more than doubled; clearly, the effect sizes need to be recalculated in light of these new data. Methodologically, past meta-analyses were forced by the paucity of specific studies to combine data across all simulation conditions and clinical groups. A critical issue is whether MMPI-2 fake-bad indexes are equally effective across different diagnostic groups. For example, do cut scores and effect sizes work equally well for patients presenting with PTSD and schizophrenia? In addition, most MMPI-2 feigning studies appear to use samples of convenience. To broaden the generalizability of the current meta-analysis, we augmented the MMPI-2 feigning research with data on validity scales from other recent studies using clinical populations with specific diagnoses.

METHOD

The basic design for this meta-analysis is modeled after Berry et al. (1991) and Rogers et al. (1994). In keeping with Rogers et al. (1994), we separately examined effect sizes for (a) simulators versus presumably healthy controls and (b) simulators versus patient groups. Because differences between simulators and controls may reflect genuine psychopathology, the latter analysis is more relevant. As a further refinement, effect sizes were also calculated on the basis of litigation status and well-represented diagnostic groups.

Compilation of MMPI-2 Studies

We conducted a PsychInfo search from 1989 (i.e., the publication date of the MMPI-2) through September 2002. We reviewed all abstracts for the MMPI-2 related to the following terms: *malingering*, *faking*, *feigning*, *fake-bad*, and *dissimulation*. To provide additional clinical samples for specific disorders, MMPI-2 abstracts related to diagnosis were examined. We also reviewed the most recent issues of major assessment journals (i.e., *Assessment*, *Journal of Personality Assessment*, and *Psychological Assessment*) for studies not yet reported in PsychInfo.

An a priori decision addressed research designs for the classification of MMPI-2 feigning. Feigning groups were included if they were derived from either known-groups comparisons or simulation designs. Some investigations attempted to use the differential prevalence design, hypothesizing that clinical groups might vary according to

the referral question (e.g., forensic vs. nonforensic) in the proportion (i.e., prevalence) of malingering. Because group membership cannot be determined by this design, their data were not included in the calculation of effect sizes for feigning groups. Research studies were also excluded that did not provide the necessary clinical data (i.e., *M*s and *SD*s for validity scales). Logically, group data were also excluded for participants in experimental conditions for other response styles (e.g., defensiveness and random responding).

Calculation of Effect Sizes and Other Estimates

An important objective of the study was the ability to make direct comparisons with earlier meta-analyses. In line with Berry et al. (1991) and Rogers et al. (1994), Rosenthal's (1984) formula was calculated: $d = (M_f - M_h) \div SD_p$. In defining terms, M_f = the mean of feigning group scores, M_h = the mean of groups under honest (standard) instructions, and SD_p = the pooled standard deviation of the two groups.

Effect sizes were calculated individually for each study on all available feigning indexes. To minimize coding errors, a researcher cross-checked the entered data (*M*s and *SD*s) with published tables. To eliminate computational errors, the effect sizes were computed in Excel via the above formula. Effect sizes were also calculated across studies to evaluate the relative usefulness of specific MMPI-2 validity scales for the determination of feigning.

In line with past research, descriptive data on cut scores were assembled. These data include individual cut scores, their hit rates, and the total number of studies and participants used in their development. Because many recent studies do not include cut scores, we also report *M*s and *SD*s by clinical groups with sufficient representation (i.e., *n*s > 100). This information provides psychologists with the option of calculating *z* scores in estimating the likelihood of feigning versus nonfeigning.

RESULTS

A total of 62 MMPI-2 feigning studies were compiled that provided criterion groups with sufficient descriptive data (*n*s, *M*s, and *SD*s) for computing effect sizes. However, 18 studies relied entirely on a differential prevalence design and were used only to calculate diagnostic data and differences due to (a) litigation or (b) group status (e.g., child custody vs. patient). These feigning studies were augmented with 11 MMPI-2 diagnostic studies that were added to increase the patient samples. Table 2 summarizes the 73 studies used in this meta-analysis, including descriptions of the samples, design, and type of instructions.

Effect sizes for individual studies are described in Table 3. Studies vary dramatically regarding which MMPI-2 scales are used and what types of comparisons are conducted. In addition to feigning indexes, a minority of studies reported standard validity scales for defensiveness, namely, Scales L and K. We included these scales in Table 3 in order to examine the *absence of defensiveness* as a potential detection strategy for MMPI-2 feigning.

An important issue is whether specific MMPI-2 validity scales vary substantially when administered to different diagnostic groups. As noted in Table 4, several scales (O-S mean $d = 3.04$; F-K mean $d = 2.44$) had very large effect sizes for different diagnoses. Psychologists must take this variability into account when evaluating response styles for certain diagnostic groups with moderate elevations.

Psychologists are often concerned about the potential effects of litigation on response styles. The differences on MMPI-2 feigning indexes due to litigation are only modest, ranging from .03 to .83 (see Table 4). Surprisingly, the effect sizes are substantially lower for litigation (mean $d = .43$) than the differences found across diagnoses (mean $d = 1.31$).

Comparisons of feigners and presumably healthy control groups yielded very large effect sizes (mean $d = 2.48$) for most MMPI-2 feigning indexes. The only major exception was the Subtle scale (mean $d = .35$). In stark contrast, three scales evidenced extremely large effect sizes: F (mean $d = 4.05$), Obvious (mean $d = 3.57$), and Fb (mean $d = 3.46$) scales. The overall results do not address the crucial issue of evaluating differences between bogus and genuine patients. Instead, they raise important methodological concerns that feigning-control comparisons may provide highly inflated effect sizes.

The paramount comparison for feigning studies is the examination of all simulators versus all genuine patients. Under nearly all circumstances, psychologists have no reliable data regarding which mental disorders a particular person is likely to feign. Many would-be malingerers are poorly informed about diagnostic information and may have only vague objectives when dissimulating (e.g., appear grossly impaired). Moreover, many patients have a complicated diagnostic presentation that is not represented by a single disorder. Given the lack of specific presentations for both feigners and genuine patients, we believe that a heterogeneous sampling of both response styles is likely to provide the best basis for comparison.

Several robust validity scales are related to three detection strategies, namely, *rare symptoms*, *erroneous stereotypes*, and *obvious-subtle symptoms*. For the rare-symptoms strategy, the two scales produced very large effect sizes, namely, F (mean $d = 2.21$) and Fp (mean $d = 1.90$). These results indicate the robustness of the rare-symptoms strategy and support its routine use for the

TABLE 2
Demographic and Methodological Characteristics for 73 Studies Using the MMPI-2 for Malingering and Clinical Diagnoses

Citation	*Sample*	N	*Age*	*% Male*	*Diagnosis/ Response*	*Comparison*
Alexy & Webb (1999)	OP	109	39.4	71.6	11H lit	NA
Arbisi & Ben-Porath (1997)	VIP	73	46.9	100	4H	3
	VIP	70	56.1	100	3H	
	VIP	80	42.7	100	8H	
	VIP	55	44.2	100	2H	
	VIP	30	48.3	100	6H	
Arbisi & Ben-Porath (1998)	VIP	41	43.1	82.9	7H	2
	VIP	33	43.3	90.9	1F	
Archer, Handel, Greene, Baer, & Elkins (2001)	IP	617	34.0	53.3	7H	2
	ST/CV	203	NR	29.6	1F	
Austin (1992)	ST	33	NR	NR	1H	1
	ST	37	NR	NR	1F	
Baer & Sekirnjak (1997)	OP	20	36.0	35.0	7H	NA
Bagby, Rogers, & Buis (1994)	ST	90	22.9[a]	35.3[a]	1H	1
	ST	58			1F	2
	FIP	173	34.1	34.8	7H	
Bagby, Rogers, Buis, & Kalemba (1994)	ST	90	22.0[a]	29.9[a]	1H	1
	ST	58			1F	2
	IP	95	35.7	51.6	7H	
Bagby, Rogers, Buis, et al. (1997)	ST	40	22.1	40.0	1H	1
	ST	20	20.4	17.5[a]	2F	2
	ST	20	21.6		3F	
	IP	40	39.7	47.5	2H	
	IP	40	37.7	62.5	3H	
Bagby, Rogers, Nicholson, et al. (1997)	ST	26	33.1	53.8	2F	2
	ST	28	33.8	32.1	2F	
	ST	24	22.5	50.0	2F	
	OP	51	38.7	60.8	2H	
Bagby, Nicholson, & Buis (1998)	ST	100	23.3	50.0	1H	1
	ST	74	23.6	39.2	1F	2
	OP	100	36.1	50.0	7H	
Bagby, Nicholson, Buis, Radovanovic, & Fidler (1999)	CC	115	37.4	48.7	1H lit	NA
Bagby, Nicholson, Bacchiochi, Ryder, & Bury (2002)	ST	45	22.8	26.7	1H, 1F	4
	IP/OP	75	40.0	44.0	7H	2
Baldrachi, Hilsenroth, Arsenault, Sloan, & Walter (1999)	VOP	36	45.0	100	4H	3
	VOP	13	45.0	100	4H mild	
Barthlow, Ben-Porath, Tellegen, & McNulty (2002)	OP	1,051	33.1	36.1	7H	NA
Bathurst, Gottfried, & Gottfried (1997)	CC	508	37.5	50.8	1H lit	NA
Ben-Porath, Butcher, & Graham (1991)	IP	76	33.7	57.9	2H	3
	IP	84	33.2	51.2	3H	
Berry et al. (1995)	CV	20	33.9	60.0	1H	1
	CV	18	34.3	50.0	10F	2
	OP	31	32.4	64.5	10H	6
	OP	30	38.6	60.0	10H lit	
Berry et al. (1996)	OP	30	33.2	60.0	7H	2
	OP	30	31.6	26.6	1F	
Berry et al. (2001)	OP	31	31.4	30.0	7H	2
	OP	30	32.0	25.0	1F	
	OP	29	30.7	37.9	4F	
Bowler, Hartney, & Ngo (1998)	OP	49	43.9[a]	44.1[a]	10H lit	6
	OP	9			10H	
Brems & Harris (1996)	ST	40	30.8[a]	27.5[a]	1H	1
	ST	40			1F	
Cassissi & Workman (1992)	ST	20	22.0[a]	58.0[a]	1H	1
	ST	20			1F	

(continued)

TABLE 2 (continued)

Citation	*Sample*	N	*Age*	*% Male*	*Diagnosis/ Response*	*Comparison*
Cramer (1995)	ST	31	20.4[a]	NR	1H	1
	ST	62		NR	2F	
	ST	62		NR	3F	
Cumella, Wall, & Kerr-Almeida (2000)	OP	446	27.0	0	13H	NA
Elhai, Gold, Fruch, & Gold (2000)	OP	124	45.7	100	4H	2
	ST	84	29.8	32.1	4F	
Elhai, Gold, Sellers, & Dorfman (2001)	OP	64	31.2	14.1	4H	2
	ST	80	29.7	31.8	4F	
Fox, Gerson, & Lees-Haley (1995)	OP	289	40.9	45.7	12H lit	NA
Frueh, Smith, & Barker (1996)	VOP	44	45.7[a]	100[a]	4H	6
	VOP	98			4H lit	
Gandolfo (1995)	OP	129	42.5	46.3	12H lit	NA
Graham, Watts, & Timbrook (1991)	ST	50	19.0	60.0	1H, 1F	4
	IP	50	28.6	60.0	7H	2
Greiffenstein, Gola, & Baker (1995)	OP	56	32.7	NR	10H	3
	OP	53	34.6	NR	10H mild	
	OP	68	38.3	NR	10H lit	6
Greiffenstein & Baker (2001) (pre/post injury)	OP	23	40.9	35.0	10H lit	NA
Hoffman, Scott, Emick, & Adams (1999)	OP	62	31.9	79.2	10H	6
	OP	50	37.6	78.0	10H lit	
Iverson, Franzen, & Hammond (1995)	PR	27	36.1	100	1H	1
	PR	28	33.7	100	1F	2
	IP	51	36.2	100	7H	
Kirz, Drescher, Klein, Gusman, & Schwartz (2001)	VIP	118	48.4	100	4H	NA
	IP	59	35.9	0	4H	
Klonsky & Bertelson (2000)	OP	30	30.0[a]	18.0[a]	3H	3
	OP	21			3H mild	
Ladd (1998)	VIP[b]	706	47.7	100	7H	3
	IP	180	38.2	75.5	8H	
Lees-Haley (1991)	OP	48	37.7	41.7	12H lit	NA
Lees-Haley (1992)	OP	55	38.9	58.2	4H lit	6
	OP	64	39.1	42.2	12H lit	
Lees-Haley (1997)	OP	492	42.0	46.7	12H lit	NA
LePage & Mogge (2001)	IP	90	29.9	70.0	7H	NA
Lewis et al. (2002)	FIP	31	43.5	100	7H	2
	FIP	24	32.5	100	7F[c]	
Lim & Butcher (1996)	ST	50	23.9[a]	50.0	1H, 1F	1
	IP	50		60.0	7H	2
Lindblad (1994)	FIP	66	32.7	100	7H, 1F	4
McGrath, Sweeney, O'Malley, & Carlton (1998)	OP	125	39.5	53.6	11H	NA
Meyers, Millis, & Vokert (2002)	OP	100	39.6	63.0	11H	2
	OP	100	38.5	42.0	11H lit	6
	EX	30	44.0	26.7	11F	
Mittenberg, Tremont, & Rayls (1996)	OP	88	49.3	53.4	10H	NA
Morrell & Rubin (2001)	OP	58	36.2[a]	0	4H	3
	OP	35			4H mild	
Moskowitz, Lewis, Ito, & Ehrmentraut (1999)	FIP	43	40.84[a]	70.4[a]	7H	NA
Pensa, Dorfman, Gold, & Schneider (1996)	IP	20	30.2	100	9H	2
	CV	20	30.3	100	9F	
Posthuma & Harper (1998)	CC	188	NR	100	1H lit	6
	OP	95	NR	NR	12H lit	
Rodevich & Wanlass (1995)	OP	42	37.4	100	10H	NA
Rogers, Bagby, & Chakraborty (1993)	CV	13	38.1	48.6	1H	1
	CV	59	38.1	49.0	2F	2
	IP	37	32.8	97.3	2H	
Rogers, Sewell, & Ustad (1995)	OP	42	36.8	51.3	7H, 1F	4
Shea, McKee, Craig Shea, & Culley (1996)	FIP	217	31.3	100	7H	NA

(continued)

TABLE 2 (continued)

Citation	*Sample*	N	*Age*	*% Male*	*Diagnosis/ Response*	*Comparison*
Shores & Carstairs (1998)	ST	18	31.4	27.8	1H	1
	ST	18	35.8	27.8	1F	
Siegel (1996)	CC	80	35.9	57.5	1H lit	NA
Sivec, Lynn, & Garske (1994)	ST	58	19.0[a]	37.9	1H	1
	ST	64		40.6	1F	
	ST	57		42.1	9F	
Sivec et al. (1995)	ST	61	19.0	16.4	1H	1
	ST	65	18.8	24.6	5F	2
	ST	61	18.5	16.4	3F	
	OP	40	28.8	12.5	5H	
Storm & Graham (2000)	IP	352	32.0	54.5	7H	2
	ST	440	19.4	36.4	1F	
Strong, Greene, Hoppe, Johnston, & Olesen (1999)	CC	412	38.1	50.0	1H lit	NA
Stukenberg, Brady, & Klinetob (2000)	IP	521	32.0	48.4	7H	NA
Timbrook, Graham, Keiller, & Watts (1993)	ST	47	19.2	53.3	1H, 1F	4
	IP	47	29.9	59.2	7H	2
Tsushima & Tsushima (2001)	OP	208	47.3	53.4	7H	6
	OP	120	41.4	52.5	7H lit	
Viglione et al. (2001)	ST	44	29.3[a]	28.0[a]	3F	5
	ST	44			1F	
Walters & Clopton (2000)	ST	95	19.2[a]	47.4	1H	1
	ST	370		46.2	1F	
Wetter, Baer, Berry, & Reynolds (1994)	CV	36	33.0	30.6	1H	1
	CV	23	31.0	21.7	1F	
	CV	23	31.0	13.0	5F	2
	OP	36	32.0	16.7	5H	
Wetter, Baer, Berry, Robison, & Sumpter (1993)	VIP/VOP	20	38.3	55.0	2H	2
	VIP/VOP	20	39.4	70.0	4H	2
	CV	20	34.8	40.0	4F	
	CV	22	34.0	68.2	2F	
Wetter, Baer, Berry, Smith, & Larsen (1992)	ST	68	24.6	48.5	1H	1
	ST	70	23.3	42.9	1F	
Wetter & Deitsch (1996) (Time 1 only)	ST	32	18.8	43.8	1H	1
	ST	32	19.8	40.6	4F	
	ST	32	19.4	50.0	10F	
Wong, Lerner-Poppen, & Durham (1998)	ST	28	19.3 [a]	21.5[a]	1H	1
	ST	51			10F	
Youngjohn, Davis, & Wolf (1997)	OP	12	33.6	83.3	10H	6
	OP	48	34.3	66.7	10H lit	

NOTE: Samples were the following: IP = inpatient, OP = outpatient, ST = student, CV = community volunteers, CC = child custody, PR = prisoners, FIP = forensic inpatients, VIP = VA inpatients, VOP = VA outpatients, and EX = experts. Responses were the following: H = Honest (i.e., groups under standard instructions) and F = Fake (i.e., groups under feigning instructions). Specifically, 1H = control or nonclinical sample, 1F = faking global impairment (i.e., "fake-bad" instructions). Diagnoses were the following: 2 = schizophrenia, 3 = depression, 4 = post-traumatic stress disorder (PTSD), 5 = borderline personality, 6 = bipolar, 7 = mixed diagnoses, 8 = substance abuse, 9 = psychosis, 10 = cognitive impairment, 11 = chronic pain, 12 = personal injury/workers' comp, 13 = eating disorder. Lit = litigants. Comparison types were 1 = simulators versus normals, 2 = simulators versus patients, 3 = patients versus patients, 4 = repeated measures (same sample with administration under different conditions), 5 = simulators versus other simulators, and 6 = litigants versus patients/other litigants. NR = not reported; NA = not applicable.

a. Overall means and percentages reported before group assignment/identification.

b. Data taken from VA sample by Arbisi and Ben-Porath (1995).

c. Subsample composed of patients classified as probable feigners according to Structured Interview of Reported Symptoms (SIRS) scores.

MMPI-2 assessment of feigning. The slightly larger mean effect size for F versus Fp was surprising, given the refinements in Fp item selection that specifically differentiate genuine patients from feigners.

Erroneous-stereotypes strategy is a sophisticated method for the detection of feigned mental disorders. As summarized in Table 4, the full Ds scale produced a large effect size (mean $d = 1.62$) that appears slightly larger than the briefer Dsr (mean $d = 1.49$). In addition, two MMPI-2 validity indexes, O-S and Obvious, demonstrate the usefulness of the obvious-subtle strategy in evaluating feigned psychological impairment. Clearly, the "Obvious"

TABLE 3
Effect Sizes for Individual Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Studies

Study and Design	*L*	*F*	*K*	*Fb*	*F-K*	*Fp*	*O-S*	*Ds*	*Dsr*	*Obv*	*Subtle*	*FBS*	*LW*
Arbisi & Ben-Porath (1997)													
(3H vs. 2H)		0.20		0.08		0.40							
(3H vs. 8H)		0.26		0.23		0.06							
(3H vs. 4H)		0.48		0.33		0.14							
(3H vs. 6H)		0.06		0.28		0.06							
Arbisi & Ben-Porath (1998)													
(7H vs. 1F)	0.42	2.19	0.20	1.61		3.78							
Archer, Handel, Greene, Baer, & Elkins (2001)													
(7H vs. 1F)	0.57	0.85	0.27	0.41		0.83							
Austin (1992)													
(1H vs. 1F)	0.10	1.98	1.93		4.43		4.53						
Bagby, Rogers, & Buis (1994)													
(1H vs. 1F)	0.16	3.05	1.00	2.44	2.87		2.00		1.95				2.30
(7H vs. 1F)	0.72	1.74	1.05	1.35	2.08		1.42		1.91				1.64
Bagby, Rogers, Buis, & Kalemba (1994)													
(1H vs. 1F)	0.17	2.89	0.99	2.34	2.78		1.98		1.91				2.23
(7H vs. 1F)	0.34	2.29	1.02	1.66	2.40		1.41		1.51				1.66
Bagby, Buis, et al. (1997)													
(3H vs. 3F)		3.07		3.25	2.71	2.03	2.09	2.43		2.04	1.01		
(1H vs. 3F)		3.58		5.47	4.86	1.98	2.79	3.64		3.21	0.33		
(1H vs. 2F)		6.53		4.23	0.72	4.32	2.89	3.89		3.44	0.14		
(2H vs. 2F)		3.92		2.12	3.28	3.70	1.87	2.37		2.10	0.50		
Bagby, Rogers, Nicholson, et al. (1997)													
(2H vs. 2F)		1.86		0.78	1.79	1.39	1.66	1.44		1.77	0.52		
Bagby, Nicholson, & Buis (1998)													
(1H vs. 1F)		3.02					1.25			2.68	0.59		
(7H vs. 1F)		2.50					0.44			1.39	0.83		
Bagby, Nicholson, Bacchiochi, Ryder, & Bury (2002)													
(1H vs. 1F)		3.06		3.44		2.00							
(7H vs. 1F)		1.26		1.19		1.53							
Ben-Porath, Butcher, & Graham (1991)													
(2H vs. 3H)	0.21	0.29	0.05										
Berry et al. (1995)													
(1H vs. 10F)	0.48	1.79	0.94	1.38	1.49	1.27		1.80					
(10H vs. 10H lit)	0.29	0.90	0.84	0.92	1.08	0.34		1.01					
(10H vs. 10F)	0.95	2.48	2.16	2.21	2.31	1.54		2.41					
Berry et al. (1996)													
(7H vs. 1F)	0.46	3.87	1.64	2.86	2.90	2.52		2.71					
Berry et al. (2001)													
(7H vs. 1F)	0.04	0.28		0.14	0.04	0.37		0.01					
(7H vs. 4F)	0.01	0.03		0.31	0.38	0.07		0.28					
Bowler, Hartney, & Ngo (1998)													
(10H vs. 10H lit)	0.14	0.47	0.28										
Brems & Harris (1996)													
(1H vs. 1F)	0.03			2.08									
Cassisi & Workman (1992)													
(1H vs. 1F)	0.45	3.63	0.82										
Cramer (1995)													
(1H vs. 2F)		2.29		2.12	2.16		1.81					1.25	1.88
(1H vs. 3F)		2.11		2.39	1.99		1.93					1.91	1.88
Elhai, Gold, Fruch, & Gold (2000)													
(4H vs. 4F)	0.22	0.93	0.13		1.00	1.01	0.33	0.87				0.09	
Elhai, Gold, Sellers, & Dorfman (2001)													
(4H vs. 4F)		1.10			1.37	1.42	0.86	1.03				0.47	
Frueh, Smith, & Barker (1996)													
(4H vs. 4Hlit)	0.13	0.79	0.63										

(continued)

TABLE 3 (continued)

Study and Design	*L*	*F*	*K*	*Fb*	*F-K*	*Fp*	*O-S*	*Ds*	*Dsr*	*Obv*	*Subtle*	*FBS*	*LW*
Graham, Watts, & Timbrook (1991)													
(1H vs. 1F)	0.18	4.20	1.57	3.05									
(7H vs. 1F)	0.60	1.96	1.08	1.59									
Greiffenstein, Gola, & Baker (1995)													
(10H vs. 10H lit)	0.03	0.39			0.16								
(10H vs. 10H mild)	0.01	0.10			0.18								
(10Hmild vs. 10H lit)	0.02	0.28			0.38								
Hoffman, Scott, Emick, & Adams (1999)													
(10H vs. 10H lit)	0.08	0.11	0.09										
Iverson, Franzen, & Hammond (1995)													
(1H vs. 1F)	0.86	3.21	0.85	2.02	2.73								
(7H vs. 1F)	0.24	3.20	0.58	1.68	2.78								
Ladd (1998)													
(7H vs. 8H)		0.40		0.45		0.41							
Lees-Haley (1992)													
(4H lit vs. 12H lit)		2.04			2.44		3.04					1.72	
Lewis, Simcox, & Berry (2002)													
(7H vs. 7F)		2.90		3.29	2.53	2.60							
Lim & Butcher (1996)													
(1H vs. 1F)	0.18	10.42	1.90	10.38									
(7H vs. 1F)	0.63	2.14	1.03	1.89									
Lindblad (1994)													
(7H vs. 1F)		3.63					2.72						
Meyers, Millis, & Volkert (2002)													
(11H vs. 11H lit)		0.75			0.72	0.37	0.83		0.90			0.62	
(11H vs. 11F)	0.55	2.80	1.28										
Morrell & Rubin (2001) (4H vs. 4H mild)	0.06	0.57	0.65										
Pensa, Dorfman, Gold, & Schneider (1996)													
(9H vs. 9F)		2.40		1.01	2.24		2.00						
Posthuma & Harper (1998)												1.39	1.56
(1H lit vs. 12H lit)	0.04	1.35	0.89	1.02	0.75								
Rogers, Bagby, & Chakraborty (1993)													
(1H vs. 1F)		1.72		1.58	1.61		1.64		1.80				1.66
(7H vs. 1F)		0.90		0.69	0.97		1.53		1.37				0.62
Rogers, Sewell, & Ustad (1995)													
(7H vs. 1F)	0.20	0.85	0.09	1.65	2.02	2.52	0.63	1.36	1.17			0.47	
Shores & Carstairs (1998)													
(1H vs. 1F)	0.03	13.66	2.10	8.14									
Sivec, Lynn, & Garske (1994)													
(1H vs. 1F)	0.20	4.70	1.21		3.41		3.51	3.34					
(1H vs. 9F)	0.46	5.74	1.39		4.32		4.08	3.71					
Sivec, Hilsenroth, & Lynn (1995)													
(1H vs. 3F)	0.17	5.30	1.23										
(1H vs. 5F)	0.59	4.90	1.95										
(5H vs. 5F)	0.60	2.97	1.26										
Storm & Graham (2000)													
(7H vs. 1F)	0.26	1.40	0.81		1.61	1.90			1.49				1.16
Timbrook, Graham, Keiller, & Watts (1993)													
(1H vs. 1F)		5.39					4.38			4.95	0.32		
(7H vs. 1F)		4.42					2.51			2.89	0.55		
Tsushima & Tsushima (2001)													
(7H vs. 7H lit)		0.09		0.11		0.16		0.03				0.60	
Viglione et al. (2001)													
(1F vs. 3F)		0.32		0.05		0.83							
Walters & Clopton (2000)													
(1H vs. 1F)		2.51		2.70	1.94	1.42	2.29	2.21	2.23				
Wetter, Baer, Berry, Smith, & Larsen (1992)													
(1H vs. 1F)	0.14	4.65	1.55	4.39	1.64			3.49					

(continued)

TABLE 3 (continued)

Study and Design	*L*	*F*	*K*	*Fb*	*F-K*	*Fp*	*O-S*	*Ds*	*Dsr*	*Obv*	*Subtle*	*FBS*	*LW*
Wetter, Baer, Berry, Robison, & Sumpter (1993)													
(2H vs. 2F)	0.50	3.21	1.78	3.40	3.58			2.91					
(4H vs. 4F)	0.57	1.52	0.51	1.13	1.64			1.73					
Wetter, Baer, Berry, & Reynolds (1994)													
(1H vs. 1F)	0.26	3.57	0.73	2.87	2.32	2.28		2.56					
(5H vs. 5F)	0.50	1.67	0.91	1.50	1.67	2.13		1.56					
(1H vs. 5F)	.54	4.52	1.10	4.57	2.86	3.73		3.36					
Wetter & Deitsch (1996) (Time 1 only)													
(1H vs. 4F)	0.25	2.11	0.69	2.25	1.78	1.96		2.48					
(1H vs. 10F)	0.34	1.45	0.66	1.42	1.25	1.23		1.51					
Wong, Lerner-Poppen, & Durham (1998)													
(1H vs. 10F)	0.13	1.55	0.40										
Youngjohn, Davis, & Wolf (1997a)													
(10H vs. 10H lit)	0.14	0.14	0.08										

NOTE: H = honest (i.e., groups under standard instructions), F = fake (i.e., groups under feigning instructions). Specifically, 1H = control or nonclinical sample, 1F = faking global impairment (i.e., "fake-bad" instructions). Diagnoses were the following: 2 = schizophrenia, 3 = depression, 4 = post-traumatic stress disorder (PTSD), 5 = borderline personality, 6 = bipolar, 7 = mixed diagnoses, 8 = substance abuse, 9 = psychosis, 10 = cognitive impairment, 11 = chronic pain, and 12 = personal injury/workers' comp. Lit = litigants. L = Lie; F = Infrequency; K = Correction; Fb = Back Infrequency; F-K = raw score difference of Infrequency-Correction; Fp = Infrequency-Psychopathology; O-S = T score difference of Obvious-Subtle; Ds = Dissimulation; Dsr = Dissimulation-Revised; Obv = Obvious; FBS = Fake Bad Scale; LW = Lachar-Wrobel.

TABLE 4
Composite Effect Sizes (d) for Simulators, Nonclinical Controls, and Patient Groups

*Type (*n*)*	*L*	*F*	*K*	*Fb*	*F-K*	*Fp*	*O-S*	*Ds*	*Dsr*	*Obv*	*Subtle*	*FBS*	*LW*
Comparisons of genuine patients from different diagnostic groups													
NA (1,473)	0.21	0.53	0.05	0.28	2.44	0.27	3.04						
Genuine patients: those with versus those without litigation													
NA (1,138)	0.12	0.44	0.38	0.52	0.59	0.29	0.83	0.03	0.09				0.62
All simulators versus nonclinical controls													
NA (2,514)	0.29	4.05	1.22	3.46	2.51	2.24	2.70	2.95	1.97	3.57	0.35	1.58	1.99
Simulators of specific disorders versus genuine patients with same disorders													
Schizophrenia (231)	0.50	3.00	1.78	2.10	2.88	2.34	1.77	2.24		1.94	0.51	0.19	
Post-traumatic stress disorder (392)	0.40	1.18	0.32	1.13	1.34	1.22	0.97	0.95				0.28	
All simulators versus all genuine patients													
NA (4,151)	0.45	2.21	0.89	1.62	1.98	1.90	1.51	1.62	1.49	2.03	0.68	0.32	1.27

NOTE: NA = not applicable. L = Lie; F = Infrequency; K = Correction; Fb = Back Infrequency; F-K = raw score difference of Infrequency-Correction; Fp = Infrequency-Psychopathology; O-S = T score difference of Obvious-Subtle; Ds = Dissimulation; Dsr = Dissimulation-Revised; Obv = Obvious; FBS = Fake Bad Scale; LW = Lachar-Wrobel.

(mean $d = 2.03$) has a much greater effect than the "Subtle" (mean $d = .68$) component of this subtraction. While Obvious appears very promising, its results are concentrated on a few studies from two research programs (see Table 3). Despite lower effect sizes (mean $d = 1.51$), psychologists may wish to continue using O-S because of its extensive research with clinical comparisons for 11 studies and a total of 1,403 participants.

Recent investigations underscore psychologists' concerns that MMPI-2 validity scales may have only limited applicability to certain diagnostic groups. The primary concern is whether specific disorders result in highly elevated feigning indexes. To address this issue, Table 5 reports descriptive data on five diagnostic categories: schizophrenia, depression, PTSD, cognitive impairment, and mixed diagnoses. Using one standard deviation above the mean as a convenient benchmark, patients with genuine schizophrenia may have extreme elevations[3] on F ($M + 1\ SD = 103.30$), Fb ($M + 1\ SD = 103.62$) and marked elevations on Fp ($M + 1\ SD = 86.80$). In addition, patients with genuine depression have the possibility of extreme elevations on F ($M + 1\ SD = 93.27$) and Fb ($M + 1\ SD = 106.14$).

TABLE 5
Descriptive Data (*M* and *SD*) for Specific Diagnoses for Presumptively Genuine Patients

Type	*L*	*F*	*K*	*Fb*	*F-K*	*Fp*	*O-S*	*Ds*	*Dsr*	*Obv*	*Subtle*	*FBS*	*LW*
Schizophrenia													
M	54.82	80.10	55.41	79.36	–0.89	66.69	58.58	65.67		330.22	251.28		
SD	11.51	23.20	12.67	24.26	10.81	20.11	91.62	16.17		60.75	31.66		
Depression													
M	50.23	71.68	44.99	82.02		59.88	79.10	64.40					
SD	9.46	21.59	9.78	24.12		17.43	61.59	15.01					
Post-traumatic stress disorder													
M	52.67	86.31	38.30	92.31	8.70	69.02	182.24	68.40				80.36	
SD	9.31	21.58	7.31	24.55	10.60	21.00	71.79	14.60				14.51	
Cognitive impairment													
M	55.70	61.96	49.55	68.45	–7.11	50.00							
SD	10.54	16.56	9.31	22.44	10.41	8.10							
Child custody litigants													
M	57.13	45.66	59.05	44.63									
SD	11.50	7.65	9.15	5.12									
Forensic groups excluding child custody													
M	56.44	66.46	47.65	63.77	–4.89	54.66	72.29	52.92				77.60	38.60[a]
SD	11.27	20.48	10.50	22.27	10.17	16.52	85.98	14.12				18.63	15.30
Mixed diagnostic group													
M	54.74	75.56	44.79	79.15	–0.58	59.98	73.82	54.75	64.16	335.82	258.72		39.44[a]
SD	12.23	23.72	10.73	24.84	11.74	19.02	91.12	14.22	16.87	72.85	32.42		20.08
All genuine patients													
M	53.92	65.70	48.00	71.34	–3.34	59.77	77.39	61.24	62.42	333.41	255.53	74.96	39.33[a]
SD	10.70	19.03	9.89	22.23	10.36	18.69	86.89	14.20	15.77	67.94	32.10	17.26	19.52
All feigners													
M	49.42	108.09	38.24	107.52	25.49	86.41	200.84	87.49	96.44			80.71	118.50
SD	11.47	23.82	7.90	25.50	20.55	25.22	73.77	15.70	16.81			16.43	46.57

NOTE: L = Lie; F = Infrequency; K = Correction; Fb = Back Infrequency; F-K = raw score difference of Infrequency-Correction; Fp = Infrequency-Psychopathology; O-S = T score difference of Obvious-Subtle; Ds = Dissimulation; Dsr = Dissimulation-Revised; Obv = Obvious; FBS = Fake Bad Scale; LW = Lachar-Wrobel.
a. Raw scores.

Moreover, patients with genuine PTSD produce slightly higher elevations than the other diagnostic groups with the possibility of very extreme elevations on Fb ($M + 1\ SD = 116.86$) and extreme elevations on F ($M + 1\ SD = 107.89$) and lower but extreme elevations on Fp ($M + 1\ SD = 90.02$).

Concerns have been raised about the effects of cognitive impairment on MMPI-2 profile validity (e.g., Mittenberg, Tremont, & Rayls, 1996; Youngjohn, Davis, & Wolf, 1997). As observed in Table 5, only scale Fb produces a moderate likelihood of an extreme elevation ($M + 1\ SD = 90.89$) as a result of cognitive impairment. In contrast, the F scale has the likelihood of a moderate elevation ($M + 1\ SD = 78.52$), whereas Fp falls clearly in the average range ($M + 1\ SD = 58.10$). Although concerns are likely to continue about clinical interpretations with cognitively impaired patients (Gass & Wald, 1997), the Fp scale appears to work especially well with this population.

Many recent studies have omitted cut scores for MMPI-2 feigning indexes. On one hand, these omissions are understandable in light of the highly divergent results reported in past meta-analyses (Berry et al., 1991; Rogers et al., 1994). On the other hand, the absence of optimized cut scores militates against a systematic analysis of feigning indexes across simulation studies. We address cutting scores from two perspectives (see Table 6). First, we summarized cut scores from feigning studies, similar to past meta-analyses. These data include the optimal cut scores, the number of studies, and the overall hit rates. Second, we adopted a normative approach to ensure that few genuine patients were misclassified as feigning. For the normative approach, we calculated the 98th percentile ($z = 2.06$) for the entire patient sample included in the meta-analysis. For purposes of comparison, we also provided Greene's (2000) compilation of patient data from Caldwell (1998) for the 98th percentile. As summarized in Table 6, normatively based cut scores are only useful with very extreme elevations. This observation is especially true for F, Fb, and O-S. For the rare-symptoms strategy, a strong positive finding was for Fp with strongly convergent data for cut scores, spanning both individual studies and normative compilations.

TABLE 6
Cut Scores for Minnesota Multiphasic Personality Inventory-2 (MMPI-2) Feigning Indexes

F	*Fb*	*F-K*	*Fp*	*O-S*	*Dsr*	*Ds*	*LW*
C (#) %	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*
r15 (2) 90	r11 (1) 82	–8 (1) 87	r4 (2) 90	T90 (1) 93	r15 (1) 80	r35 (6) 77	r57 (1) 83
r16 (2) 86	r17 (1) 66	–4 (1) 91	r5 (1) 77	T100 (1) 85	r16 (2) 77	T97 (1) 71	r61 (1) 78
r17 (3) 92	r18 (1) 88	2 (1) 83	r6 (1) 84	T106 (1) 87	r17 (1) 69		r77 (2) 79
r19 (1) 90	r23 (1) 95	6 (1) 85	r8 (1) 90	T150 (1) 88	r18 (1) 80		r82 (1) 69
r20 (1) 74	r25 (3) 76	7 (1) 87	r9 (3) 79	T160 (1) 87	r21 (1) 87		r90 (1) 85
r22 (1) 93	r28 (1) 93	8 (1) 88	T90 (2) 81	T169 (1) 82	r22 (1) 90		
r28 (1) 72	T80 (1) 85	10 (1) 89	T100 (2) 91	T180 (1) 80	r23 (1) 85		
r29 (4) 83	T93 (1) 82	11 (2) 83		T190 (2) 91			
R (30) 96	T98 (1) 73	12 (2) 90		T221 (1) 63			
T62 (1) 94	T104 (1) 76	13 (1) 76					
T65 (1) 89	T105 (1) 85	14 (1) 68					
T70 (1) 88	T106 (1) 93	15 (2) 88					
T80 (2) 86	T108 (1) 91	16 (1) 84					
T96 (1) 78	T120 (1) 77	17 (1) 70					
T98 (1) 76		18 (3) 84					
T100 (1) 88		23 (1) 89					
T104 (2) 93		32 (1) 94					
T107 (1) 90							
T120 (2) 76							

Unweighted Mean Cut Scores for Reported Studies in the Current Meta-Analysis[a]

F	*Fb*	*F-K*	*Fp*	*O-S*	*Dsr*	*Ds*	*LW*
C (#) %	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*	*C (#) %*
20 (29) 86	18 (16) 82	12 (22) 84	7 (12) 84	156 (10) 85	19 (8) 79	35 (7) 76	74 (6) 79

Normative Cut Scores for Current Meta-Analysis[a]

F	*Fb*	*F-K*	*Fp*	*O-S*	*Dsr*	*Ds*	*LW*
T105[b]	T117[c]	r18	T98[d]	T256	T95	T91[e]	r80

Normative Cut Scores Based on Greene's (2000) Summary of Caldwell's Data Set[f]

F	*Fb*	*F-K*	*Fp*	*O-S*	*Dsr*	*Ds*	*LW*
24	20	15	7	240	NA	35	73

NOTE: C = optimal cut score; # = number of simulation studies; % = the overall classification rates. All research and normative cut scores should be considered close approximations because of rounding. F = Infrequency; Fb = Back Infrequency; F-K = raw score difference of Infrequency-Correction; Fp = Infrequency-Psychopathology; O-S = T score difference of Obvious-Subtle; Ds = Dissimulation; Dsr = Dissimulation-Revised; LW = Lachar-Wrobel.
a. Cut scores are approximate because of *T* to raw score transformations.
b. Approximately r22 for men and r20 for women.
c. Approximately r18 for men and r19 for women.
d. Approximately r8 for men and r9 for women.
e. Approximately r32 for men and r34 for women.
f. Caldwell normative data are provided at the 98th percentile (i.e., cut scores at this level would result in ≤ 2% of presumably genuine patients being misclassified as feigning).

The Ds scale is distinguished from all other MMPI-2 validity scales by the remarkable consistency in published cut scores with six studies using Ds > 35 raw and the seventh study using its *T*-score equivalent for men. Although its overall classification rate is somewhat lower (76%), avoiding the marked range in cut scores plainly outweighs this limitation. Equally impressive, Caldwell's normative data yield the same cut score (Ds > 35) that also minimizes false-positives with the current normative data (see Table 6). A slightly higher cut score (Ds > 99T) would reduce further the possibility of the false-positives for problematic diagnoses, such as PTSD and schizophrenia.

DISCUSSION

Effectiveness of Detection Strategies and Scales

Butcher and Williams (1992) advocated the use of two standard MMPI validity scales (i.e., F and Fb) for the evaluation of feigned profiles. As found in the current meta-analysis across all simulators and genuine patients (see Table 4), F has a very large effect size (mean $d = 2.21$) in contrast to Fb (mean $d = 1.62$). The current data suggest a reconsideration of Butcher and Williams's recommendations. Both F and Fb capitalize on the identical scale development (normative item selection) and detection strategy (rare symptoms). Beyond its redundancy with and lower effect sizes than F, Fb appears vulnerable to the misclassification of genuine patients. Employing the earlier benchmark ($M + 1\ SD$), extreme elevations (i.e., > 100T) are anticipated in a substantial minority of genuine patients with schizophrenia, depression, and PTSD. Therefore, the routine use of Fb runs the risk of more false-positives than F but is unlikely to add incremental validity.

An important consideration is whether the MMPI-2 Fp should be selected as the primary *rare symptoms* strategy. In a straightforward comparison of effect sizes, the Fp (mean $d = 1.90$) produces a slightly lower effect size than F. On a conceptual basis, however, the Fp was designed to assess differences between genuine disorders and feigning. In contradistinction, F is a normatively developed scale that simply measures divergence from normality but does not necessarily distinguish genuine from feigned abnormality. This key difference in scale development is likely responsible for the corresponding differences in clinical elevations. For example, patients with PTSD have marked elevations on F ($M = 86.31$) compared with moderate elevations on Fp ($M = 69.02$). The comparative advantages of F and Fp will be revisited with reference to cut scores.

A second detection strategy that warrants close attention is *erroneous stereotypes*. A large effect size was found for Ds in evaluating erroneous stereotypes (mean $d = 1.62$) for all patients versus all feigners (see Table 4). The Ds scale appears particularly effective in minimizing elevations for genuine patients. In particular, the mixed diagnostic group produced only an average score ($M = 54.75$) with marginal elevations for patients with schizophrenia ($M = 65.67$) and PTSD ($M = 68.40$). Based on the normative data (see Table 5), the Ds clearly merits examination in clinical cases where feigning is suspected. In stark contrast to Ds, FBS also tries to capitalize on erroneous stereotypes but was designed for only circumscribed referrals (i.e., personal injury cases). Its general lack of success (mean $d = .32$) is likely attributable to its narrow focus.

Two additional detection strategies are *obvious-subtle*, and *symptom selectivity*. The obvious-subtle strategy as measured by O-S also produced a large effect size (mean $d = 1.51$). Its marked variation (i.e., SDs > 60) for genuine patients both within diagnoses and across diagnostic groups raises questions about the O-S's clinical applicability. Finally, LW as a measure of symptom selectivity yielded a large effect size ($d = 1.27$) that is substantially lower than most other feigning indexes. In addition, the usefulness of LW remains to be investigated with specific diagnostic groups. At present, both the O-S and LW appear to be very limited in their clinical applicability.

The current findings offer partial support for Greene's bipolarity hypothesis. Whereas the effect size for L was modest ($d = .45$), K had a moderate effect ($d = .89$). Based on Table 5, most feigners do not have elevations on K (i.e., ≤ 55T). However, the magnitude of these effect sizes does not suggest that the absence of defensiveness effectively discriminates feigned from genuine profiles. Despite the lack of current clinical applicability, future research may wish to investigate the usefulness of specialized indicators, such as Wsd and Mp that appear more effective than the traditional L and K scales in the assessment of defensiveness (Baer, Wetter, & Berry, 1992).

A major concern for practitioners is whether certain diagnostic groups, such as bona fide patients with schizophrenia and PTSD, are likely to have markedly elevated scores on validity indicators (see Table 5). Such elevations are likely to lead to misclassifications. When simulators of these two disorders are compared with presumably genuine patients with the same disorders, large effect sizes are found on most feigning scales for both diagnoses (see Table 4). Despite these appreciable group differences, practitioners must be concerned about cut scores with different diagnostic groups.

Clinical Applications of Cut Scores

The establishment of accurate and consistent cut scores is the sine qua non of malingering classification. Because the previous MMPI and MMPI-2 meta-analyses yielded such divergent cut scores, many researchers in recent investigations are disinclined to report cut scores. As a result, the meta-analytic data in Table 6 represent only a modest expansion of the Rogers et al. (1994) results. Obviously, the same divergence of cut scores continues to be observed.

We augmented the cut scores with clinical data from the current study and Greene's (2000) tabulation of Caldwell's data set on more than 50,000 patients. In using a normative approach to clinical cut scores, the basic premise is that extreme scores are almost never observed in presumably genuine populations. For this purpose, we

adopted a very stringent standard (98th percentile). The obvious limitation of this approach is that an unknown but presumably small percentage of clinical samples may be undetected cases of malingering. However, their inclusion in these normative estimates likely will decrease the number of false-positives found with these cut scores.

Combining across empirically derived and normative cut scores, the Fp appears to be the most effective scale in the assessment of feigning for three reasons. First, its empirically derived cut scores are more consistent (range from > 4 to > 9) than most feigning scales and yield good classification rates (M = 84.3%). Second, the normative cut scores also have a narrow range (i.e., Caldwell data = 7; current data = 8 [men] and 9 [women]) and are generally aligned with empirically derived cut scores (see Table 6). Third, these cut scores appear to be effective across disorders (see Table 5) and even moderately useful with the problematic diagnosis of PTSD.[4]

The traditional F scale evidenced several important limitations for its cut scores. First and foremost, F exhibited marked variations in cut scores (i.e., raw scores from > 8 to > 30). As previously noted, genuine patients tended to have elevated F's (M = 65.70) with a wide distribution of scores (SD = 19.03). As a result, only extreme scores appear effective for the classification of feigning. Conservatively, the F > 24 derived from the Caldwell data would result in very few false-positives among genuine patients, including those in the current meta-analysis (see Table 5). However, for certain diagnostic groups (patients with PTSD, schizophrenia, and presumably other psychotic disorders), a cut score at the high end of the empirical data (i.e., F > 30) would appear prudent.

Most clinicians routinely evaluated Fb in the assessment of feigning. Because bona fide patients have moderate elevations (overall M = 71.34) and considerable variation (SD = 22.23), this scale appears to be confounded by genuine psychopathology. One hypothesis is that genuine patients' attention begins to falter during the latter portions of an MMPI-2 administration. Obviously, an inspection of MMPI-2 profiles for response consistency is essential with Fb elevations. Because extreme elevations can be observed in a substantial minority of presumably genuine patients, we do not recommend the routine use of Fb cut scores at the present time.

Greene (2000) suggested caution in the use of F-K as a primary indicator of feigning because of its variability of cut scores and less efficiency than F elevations alone. The current review of F-K cut scores questions its routine clinical use. The extraordinary divergence of cut scores from –8 to 32 provides clinicians with little confidence that a consistent cut score could be achieved.

Ds, capitalizing on erroneous stereotypes, demonstrated a high level of consistency across cut scores (i.e., Ds > 35 raw). Based on the normative data, the same cut score is likely to produce very few (i.e., < 2%) false-positives when combining the Caldwell and current data sets. When faced with challenging presentations (i.e., PTSD or psychotic), a slightly higher cut score (e.g., > 36 for men) may be warranted. Outperforming Dsr on effect sizes and consistency of cut scores, the Ds appears to be the premier specialized validity scale with its sophisticated strategy and minimal risk of false-positives.

O-S produced very large effect sizes, although they varied across diagnostic groups (see Table 4). We found marked variations for empirically derived cut scores (90T to 221T) that were markedly lower than normative cut scores (240 and 256). Like other indexes, we found extreme endorsement levels by presumptively genuine patients with PTSD (M = 182.24, SD = 71.79). The most prudent course of action is simply not to use O-S with any patients with PTSD histories. In addition, O-S is unlikely to be clinically useful except in rare cases of extreme endorsement levels.

Conclusions and Future Directions

The assessment of malingering is a multifaceted process bringing together different clinical methods and multiple indicators (Rogers, 1997). Within this context, the MMPI-2 should not be used as the sole or primary measure of feigning. Instead, the MMPI-2 should be viewed as an important clinical method that incorporates several key detection strategies. Of these strategies, *rare symptoms* and *erroneous stereotypes* appear to hold the most promise.

The current meta-analysis suggests that the most effective scales are likely to combine different models of scale development (i.e., discriminant, normative, and rational methods) with specific strategies (e.g., *rare symptoms* and *erroneous stereotypes*). This conclusion is at odds with the more traditional approach to scale development for feigning indexes (i.e., the exclusively normative approach to F and Fb) and its redundant reliance on the same strategy (i.e., *rare symptoms*). A future direction would be an examination of models for scale development and/or strategies that extend beyond the MMPI-2 to other standardized measures of malingering. The theoretical framework for the assessment of malingering could be improved substantially if we knew which detection strategies and which methods of scale development resulted in accurate classifications.

The most important clinical finding from the current meta-analysis involves the usefulness of the Fp across settings and diagnoses. The Fp yielded strong effect sizes and comparatively consistent cut scores that appear useful across settings and diagnostic groups. Despite time-honored traditions, we recommend the Fp as the primary MMPI-2 scale for the assessment of feigning. When feign-

ing is suspected, the Ds scale is strongly recommended because of its consistency of cut scores and low probability of false-positives.

The current findings raise several issues about the context of the evaluation. Clearly, the mere presence of litigation has only modest effects (mean d = .43) on validity indicators. Researchers employing a differential prevalence design have often assumed that the litigation substantially increases the likelihood of feigning. The current data question both the assumption and the use of this design in feigning research. Beyond litigation per se, forensic groups (even with child custody cases removed) have lower scores on validity scales than genuine patients in general (see Table 5). Indirectly, these combined results for litigation and forensic status cast doubt about the *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition*'s (*DSM-IV*) (American Psychiatric Association, 2001) postulation that the mere context of forensic evaluations increases the likelihood of malingering.

As a future direction, we would like to see the current results tested via known-groups comparisons using either expert clinical judgment or standardized methods producing very few false-positives (e.g., *Structured Interview of Reported Symptoms*) (Rogers, Bagby, & Dickens, 1992) to cross-validate findings of MMPI-2 simulation research. Even with simulation studies, the incorporation of independent measures to evaluate feigning would be strongly advisable. At present, the anomalous results for PTSD samples on select feigning indexes are difficult to interpret. Do the marked elevations on O-S and Fb indicate that these scales are confounded by PTSD symptomatology? Conversely, do these marked elevations indicate that a small proportion of these samples may be engaged in feigning, which remains undetected? When using samples of convenience not systematically screened for feigning, researchers cannot confidently rule out either interpretation.

The past four decades of MMPI/MMPI-2 research have seen a steady rise in the sophistication of feigning research. With methodological improvements (Rogers & Cruise, 1998) and the systematic appraisal of detection strategies, MMPI-2 research is likely to make continued advances in the clinical assessment of malingering.

NOTES

1. Interestingly, many simulators endorse only slightly more obvious than subtle Minnesota Multiphasic Personality Inventory–2 (MMPI-2) items. The use of *T*-score transformations dramatically increases the observed differences because obvious items occur less frequently than subtle items in the normative sample.

2. Feigning research typically produces substantial effect sizes. Therefore, we have adopted the following descriptive terms based on Cohen's *d*: ≥ .75 for "moderate," ≥ 1.25 for "large," and ≥ 1.75 for "very large."

3. For descriptive purposes, clinical scale elevations are described as follows: "moderate" ≥ 65, "marked" ≥ 80T, "extreme" ≥ 90T, and "very extreme" ≥ 110T.

4. A cut score > 9 is unlikely to occur in patients with genuine posttraumatic stress disorder with an extrapolated false-positive rate of 3.9% for men (z_{males} = 1.76) and 1.8% for women ($z_{females}$ = 2.09).

REFERENCES

Alexy, W. D., & Webb, P. M. (1999). Utility of the MMPI-2 in work-hardening rehabilitation. *Rehabilitation Psychology*, *44*(3), 266-273.

American Psychiatric Association. (2001). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

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

Arbisi, P. A., & Ben-Porath, Y. S. (1997). Characteristics of the MMPI-2 F(p) scales as a function of diagnosis in an inpatient sample of veterans. *Psychological Assessment*, *9*, 102-105.

Arbisi, P. A., & Ben-Porath, Y. S. (1998). The ability of the Minnesota Multiphasic Personality Inventory–2 validity scales to detect fake-bad responses in psychiatric inpatients.. *Psychological Assessment*, *10*, 221-228.

Archer, R. P., Handel, R. W., Greene, R. L., Baer, R. A., & Elkins, D. E. (2001). An evaluation of the usefulness of the MMPI-2 F(p) scale. *Journal of Personality Assessment*, *76*(2), 282-295.

Austin, J. S. (1992). The detection of fake-good and fake-bad on the MMPI-2. *Educational and Psychological Measurement*, *52*, 669-674.

Baer, R. A., & Sekirnjak, G. (1997). Detection of underreporting on the MMPI-2 in a clinical population: Effects of information about validity scales. *Journal of Personality Assessment*, *69*(3), 557-567.

Baer, R. A., Wetter, M. W., & Berry, D. T. R. (1992). Detection of underreporting of psychopathology on the MMPI: A meta-analysis. *Clinical Psychology Review*, *12*, 509-525.

Bagby, R. M., Nicholson, R. A., Bacchiochi, J. R., Ryder, A. G., & Bury, A. S. (2002). The predictive capacity of the MMPI-2 and PAI validity scales and indexes to detect coached and uncoached feigning. *Journal of Personality Assessment*, *78*(1), 69-86.

Bagby, R. M., Nicholson, R. A., & Buis, T. (1998). Utility of the deceptive-subtle items in the detection of malingering. *Journal of Personality Assessment*, *70*, 405-415.

Bagby, R. M., Nicholson, R. A., Buis, T., Radovanovic, H., & Fidler, B. J. (1999). Defensive responding on the MMPI-2 in family custody and access evaluations. *Psychological Assessment*, *11*(1), 24-28.

Bagby, R. M., Rogers, R., & Buis, T. (1994). Detecting malingered and defensive responding on the MMPI-2 in a forensic inpatient sample. *Journal of Personality Assessment*, *62*, 191-203.

Bagby, R. M., Rogers, R., Buis, T., & Kalemba, V. (1994). Malingered and defensive response styles on the MMPI-2: An examination of validity scales. *Assessment*, *1*, 31-38.

Bagby, R. M., Rogers, R., Buis, T., Nicholson, R. A., Cameron, S. L., Rector, N. A., et al. (1997). Detecting feigned depression and schizophrenia on the MMPI-2. *Journal of Personality Assessment*, *68*, 650-664.

Bagby, R. M., Rogers, R., Nicholson, R., Buis, T., Seeman, M. V., & Rector, N. (1997). Does clinical training facilitate feigning schizophrenia on the MMPI-2? *Psychological Assessment*, *9*, 106-112.

Baldrachi, R., Hilsenroth, M., Arsenault, L., Sloan, P., & Walter, C. (1999). MMPI-2 assessment of varying levels of posttraumatic stress in Vietnam combat veterans. *Journal of Psychopathology and Behavioral Assessment*, *21*, 109-116.

Barthlow, D. L., Ben-Porath, Y. S., Tellegen, A., & McNulty, J. L. (2002). The appropriateness of the MMPI-2 K correction. *Assessment*, *9*(3), 219-229.

Bathurst, K., Gottfried, A. W., & Gottfried, A. E. (1997). Normative data for the MMPI-2 in child custody litigation. *Psychological Assessment*, *9*(3), 205-211.

Ben-Porath, Y. S., Butcher, J. N., & Graham, J. R. (1991). Contribution of the MMPI-2 content scales to the differential diagnosis of schizophrenia and major depression. *Psychological Assessment*, *3*, 634-640.

Berry, D. T. R., Adams, J. J., Clark, C. D., Thacker, S. R., Burger, T. L., Wetter, M. W., et al. (1996). Detection of a cry for help on the MMPI-2: An analog investigation. *Journal of Personality Assessment*, *67*(1), 26-36.

Berry, D. T. R., Baer, R. A., & Harris, M. J. (1991). Detection of malingering on the MMPI: A meta-analytic review. *Clinical Psychology Review*, *11*, 585-598.

Berry, D. T. R., Cimino, C. R., Chong, N. K., LaVelle, S. H., Ivy, K., Morse, T. L., et al. (2001). MMPI-2 fake-bad scales: An attempted cross-validation of proposed cutting scores for outpatients. *Journal of Personality Assessment*, *76*, 296-314.

Berry, D. T. R., Wetter, M. W., Baer, R. A., Youngjohn, J. R., Gass, C. S., Lamb, D. G., et al. (1995). Overreporting of closed-head injury symptoms on the MMPI-2. *Psychological Assessment*, *7*, 517-523.

Bowler, R. M., Hartney, C., & Ngo, L. H. (1998). Amnestic disturbance and posttraumatic stress disorder in the aftermath of a chemical release. *Journal of Clinical Neuropsychology*, *13*, 455-471.

Brems, C., & Harris, K. (1996). Faking the MMPI-2: Utility of the subtle-obvious scales. *Journal of Clinical Psychology*, *52*(5), 525-533.

Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Telelgen, A., & Kaemmer, B. (1989). *MMPI-2 manual*. Minneapolis: University of Minnesota Press.

Butcher, J. N., & Williams, C. L. (1992). *Essentials of MMPI-2 and MMPI-A interpretation*. Minneapolis: University of Minnesota Press.

Caldwell, A. B. (1998). [MMPI-2 data research file for clinical patients]. Unpublished raw data.

Cassisi, J. E., & Workman, D. E. (1992). Detection of malingering and deception with a short form of the MMPI-2 based on the L, F, and K scales. *Journal of Clinical Psychology*, *48*, 54-58.

Cramer, K. M. (1995). The effects of description clarity and disorder type on the MMPI-2 fake-bad indices. *Journal of Clinical Psychology*, *51*, 831-840.

Cumella, E. J., Wall, A. D., & Kerr-Almeida, N. (2000). MMPI-2 in inpatient assessment of women with eating disorders. *Journal of Personality Assessment*, *75*(3), 387-403.

Dahlstrom, W. G., Welsh, G. S., & Dahlstrom, L. E. (1972). *An MMPI handbook. Volume I: Clinical interpretation* (Rev. ed.). Minneapolis: University of Minnesota Press.

Elhai, J. D., Gold, P. B., Fruch, C., & Gold, S. N. (2000). Cross-validation of the MMPI-2 in detecting malingered posttraumatic stress disorder. *Journal of Personality Assessment*, *75*, 449-463.

Elhai, J. D., Gold, S. N., Sellers, A. H., & Dorfman, W. I. (2001). The detection of malingered posttraumatic stress disorder with MMPI-2 fake bad indices. *Assessment*, *8*, 221-236.

Fox, D. D., Gerson, A., & Lees-Haley, P. R. (1995). Interrelationship of MMPI-2 validity scales in personal injury claims. *Journal of Clinical Psychology*, *51*(1), 42-47.

Frueh, B. C., Smith, D. W., & Barker, S. E. (1996). Compensation seeking status and psychometric assessment of combat veterans seeking treatment for PTSD. *Journal of Traumatic Stress*, *9*, 427-439.

Gandolfo, R. (1995). MMPI-2 profiles of worker's compensation claimants who present with claims of harassment. *Journal of Clinical Psychology*, *51*(5), 711-715.

Gass, C. S., & Luis, C. A. (2001). MMPI-2 Scale F(p) and symptom feigning: Scale refinement. *Assessment*, *8*, 425-429.

Gass, C. S., & Wald, H. S. (1997). MMPI-2 interpretation and closed-head trauma: Cross-validation of a correction factor. *Archives of Clinical Neuropsychology*, *12*, 199-205.

Gough, H. G. (1954). Some common misperceptions about neuroticism. *Journal of Consulting Psychology*, *18*, 287-292.

Graham, J. R., Watts, D., & Timbrook, R. (1991). Detecting fake-good and fake-bad MMPI-2 profiles. *Journal of Personality Assessment*, *57*, 264-277.

Greene, R. L. (1997). Assessment of malingering and defensiveness on multiscale inventories. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 169-207). New York: Guilford.

Greene, R. L. (2000). *The MMPI-2: An interpretive manual*. Boston: Allyn & Bacon.

Greiffenstein, M. F., & Baker, W. J. (2001). Comparison of premorbid and postinjury MMPI-2 profiles in late postconcussion claimants. *The Clinical Neuropsychologist*, *15*(2), 162-170.

Greiffenstein, M. F., Gola, T., & Baker, W. J. (1995). MMPI-2 validity scales versus domain specific measures in detection of factitious traumatic brain injury. *The Clinical Neuropsychologist*, *9*(3), 230-240.

Hoffman, R. G., Scott, J. G., Emick, M. A., & Adams, R. L. (1999). The MMPI-2 and closed-head injury: Effects of litigation and head injury severity. *Journal of Forensic Neuropsychology*, *1*(2), 3-13.

Iverson, G. L., Franzen, M. S., & Hammond, J. A. (1995). Examination of inmates' ability to malinger on the MMPI-2. *Psychological Assessment*, *7*, 118-121.

Kirz, J. L., Drescher, K. D., Klein, J. L., Gusman, F. D., & Schwartz, M. F. (2001). MMPI-2 assessment of differential post-traumatic stress disorder patterns in combat veterans and sexual assault victims. *Journal of Interpersonal Violence*, *16*(7), 619-639.

Klonsky, E. D., & Bertelson, A. D. (2000). MMPI-2 clinical scale difference between dysthymia and major depression. *Assessment*, *7*, 143-149.

Lachar, D., & Wrobel, T. A. (1979). Validating clinicians' hunches: Construction of a new MMPI critical item set. *Journal of Consulting and Clinical Psychology*, *47*, 277-284.

Ladd, J. S. (1998). The F(p) infrequency-psychopathology scale with chemically dependent inpatients. *Journal of Clinical Psychology*, *52*, 367-372.

Lees-Haley, P. R. (1991). Ego strength denial on the MMPI-2 as a clue to simulation of personal injury in vocational neuropsychological and emotional distress evaluations. *Perceptual and Motor Skills*, *72*, 815-819.

Lees-Haley, P. R. (1992). Efficacy of MMPI-2 validity scales and MCMI-II modifier scales for detecting spurious PTSD claims: *F*, *F-K*, Fake Bad scale, Ego Strength, subtle-obvious subscales, DIS, and DEB. *Journal of Clinical Psychology*, *48*, 681-688.

Lees-Haley, P. R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology*, *53*(7), 745-755.

Lees-Haley, P. R., English, L. T., & Glenn, W. J. (1991). A fake bad scale on the MMPI-2 for personal injury claimants. *Psychological Reports*, *68*, 203-210.

LePage, J. P., & Mogge, N. L. (2001). Validity rates of the MMPI-2 and PAI in a rural inpatient psychiatric facility. *Assessment*, *8*(1), 67-74.

Lewis, J. L., Simcox, A. M., & Berry, D. T. R. (2002). Screening for feigned psychiatric symptoms in a forensic sample by using the MMPI-2 and the Structured Inventory of Malingered Symptomatology. *Psychological Assessment*, *14*(2), 170-176.

Lim, J., & Butcher, J. N. (1996). Detection of faking on the MMPI-2: Differentiation among faking-bad, denial, and claiming extreme virtue. *Journal of Personality Assessment*, *67*(1), 1-25.

Lindblad, A. D. (1994). Detection of malingered mental illness within a forensic population: An analogue study. *Dissertation Abstracts International*, *54-B*, 4395.

McGrath, R. E., Sweeney, M., O'Malley, W. B., & Carlton, T. K. (1998). Identifying psychological contributions to chronic pain complaints

with the MMPI-2: The role of the K scale. *Journal of Personality Assessment*, *70*, 448-459.

Meyers, J. E., Millis, S. R., & Volkert, K. (2002). A validity index for the MMPI-2. *Archives of Clinical Neuropsychology*, *17*, 157-169.

Mittenberg, W., Tremont, G., & Rayls, K. R. (1996). Impact of cognitive function on MMPI_2 validity in neurologically impaired patients. *Assessment*, *3*, 157-163.

Morrell, J. S., & Rubin, L. J. (2001). The Minnesota Multiphasic Personality Inventory–2, posttraumatic stress disorder, and women domestic violence survivors. *Professional Psychology: Research and Practice*, *32*(2), 151-156.

Moskowitz, J. L., Lewis, R. J., Ito, M. S., & Ehrmentraut, J. (1999). MMPI-2 profiles of NGRI and civil patients. *Journal of Clinical Psychology*, *55*(5), 659-668.

Pensa, R., Dorfman, W. I., Gold, S. N., & Schneider, B. (1996). Detection of malingered psychosis with the MMPI-2. *Psychotherapy in Private Practice*, *14*, 47-63.

Posthuma, A. B., & Harper, J. F. (1998). Comparison of MMPI-2 responses of child custody and personal injury litigants. *Professional Psychology: Research and Practice*, *29*(5), 437-443.

Rodevich, M. A., & Wanlass, R. L. (1995). The moderating effect of spinal cord injury on MMPI-2 profiles: A clinically derived T score correction procedure. *Rehabilitation Psychology*, *40*(3), 181-190.

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

Rogers, R., Bagby, R. M., & Chakraborty, D. (1993). Faking schizophrenic disorders on the MMPI-2: Detection of coached simulators. *Journal of Personality Assessment*, *60*, 215-226.

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

Rogers, R., & Bender, S. D. (in press). Evaluation of malingering and deception. In A. M. Goldstein (Ed.), *Comprehensive handbook of psychology: Forensic psychology* (Vol. 11). New York: John Wiley.

Rogers, R., & Cruise, C. R. (1998). Assessment of malingering with simulation designs: Threats to external validity. *Law and Human Behavior*, *22*, 273-285.

Rogers, R., Sewell, K. W., & Salekin, R. T. (1994). A meta-analysis of malingering on the MMPI-2. *Assessment*, *1*, 227-237.

Rogers, R., Sewell, K. W., & Ustad, K. L. (1995). Feigning among chronic outpatients on the MMPI-2: An analogue study. *Assessment*, *2*, 81-89.

Rosenthal, R. (1984). *Meta-analytic procedures in social research*. Beverly Hills, CA: Sage.

Shea, S. J., McKee, G. R., Craig Shea, M. E., & Culley, D. C. (1996). MMPI-2 profiles of male pre-trial defendants. *Behavioral Sciences and the Law*, *14*(3), 331-338.

Shores, E. A., & Carstairs, J. R. (1998). Accuracy of the MMPI-2 computerized report in identifying fake-good and fake-bad response sets. *The Clinical Neuropsychologist*, *12*(1), 101-106.

Siegel, J. C. (1996). Traditional MMPI-2 validity indicators and initial presentation in custody evaluations. *American Journal of Forensic Psychology*, *14*(3), 55-63.

Sivec, H. J., Hilsenroth, M. J., & Lynn, S. J. (1995). Impact of simulating borderline personality disorder on the MMPI-2: A costs-benefits model employing base rates. *Journal of Personality Assessment*, *64*(2), 295-311.

Sivec, H. J., Lynn, S. J., & Garske, J. P. (1994). The effect of somatoform disorder and paranoid psychotic disorder role-related dissimulations as a response set on the MMPI-2. *Assessment*, *1*, 69-81.

Storm, J., & Graham, J. R. (2000). Detection of coached general malingering on the MMPI-2. *Psychological Assessment*, *12*, 158-165.

Strong, D. R., Greene, R. L., Hoppe, C., Johnston, T., & Olesen, N. (1999). Taxometric analysis of impression management and self-deception on the MMPI-2 in child-custody litigants. *Journal of Personality Assessment*, *73*(1), 1-18.

Stukenberg, K., Brady, C., & Klinetob, N. (2000). Use of the MMPI-2's VRIN scale with severely disturbed populations: Consistent responding may be more problematic than inconsistent responding. *Psychological Reports*, *86*, 3-14.

Timbrook, R. E., Graham, J. R., Keiller, S. W., & Watts, D. (1993). Comparison of the Wiener-Harmon subtle-obvious scales and the standard validity scales in detecting valid and invalid MMPI-2 profiles. *Psychological Assessment*, *5*, 53-61.

Tsushima, W. T., & Tsushima, V. G. (2001). Comparison of the Fake Bad Scale and other MMPI-2 validity scales with personal injury litigants. *Assessment*, *8*, 205-212.

Viglione, D. J., Mellin Wright, D., Dizon, N. T., Moynihan, J. E., DuPuis, S., & Pizitz, T. D. (2001). Evading detection on the MMPI-2: Does caution produce more realistic patterns of responding? *Assessment*, *8*(3), 237-250.

Walters, G. L., & Clopton, J. R. (2000). Effect of symptom information and validity scale information on the malingering of depression on the MMPI-2. *Journal of Personality Assessment*, *75*, 183-199.

Wetter, M. W., Baer, R. A., Berry, D. T. R., & Reynolds, S. K. (1994). The effect of symptom information on faking on the MMPI-2. *Assessment*, *1*, 199-207.

Wetter, M. W., Baer, R. A., Berry, D. T. R., Robison, L. H., & Sumpter, J. (1993). MMPI-2 profiles of motivated fakers given specific symptom information: A comparison to matched patients. *Psychological Assessment*, *5*, 317-323.

Wetter, M. W., Baer, R. A., Berry, D. T., Smith, G. T., & Larsen, L. (1992). Sensitivity of MMPI-2 validity scales to random responding and malingering. *Psychological Assessment*, *4*, 369-374.

Wetter, M. W., & Deitsch, S. E. (1996). Faking specific disorders and temporal response consistency on the MMPI-2. *Psychological Assessment*, *8*, 39-47.

Wiener, D. N. (1948). Subtle and obvious keys for the MMPI. *Journal of Consulting Psychology*, *12*, 164-170.

Wong, J. L., Lerner-Poppen, L., & Durham, J. (1998). Does warning reduce obvious malingering on memory and motor tasks in college samples? *International Journal of Rehabilitation and Health*, *4*(3), 153-165.

Youngjohn, J. R., Davis, D., & Wolf, I. (1997). Head injury and the MMPI-2: Paradoxical effects and the influence of litigation. *Psychological Assessment*, *9*, 177-184.

Richard Rogers, Ph.D., is a professor of psychology at the University of North Texas. He is highly regarded for his contributions to assessment of response styles and clinical forensic issues. His edited book, *Clinical Assessment of Malingering and Deception*, was awarded the Manfred S. Guttmacher Award from the American Psychiatric Association. He also developed the Structured Interview of Reported Symptoms (SIRS), widely regarded as the best-validated measure of feigned mental disorders. For forensic assessments, he authored with Dan Shuman, *Conducting Insanity Evaluations*. His most recent text on assessment is *A Handbook of Diagnostic and Structured Interviewing* (2001; www.guilford.com).

Kenneth W. Sewell, Ph.D., is a professor of psychology and director of clinical training at the University of North Texas. He is a prolific researcher with extensive publications in trauma, clinical assessment, and forensic issues. Most recently, he has coauthored with his colleagues, *Evaluation of Competency to*

Stand Trial–Revised (ECST-R) slated for publication by Psychological Assessment Resources in early 2004.

Mary A. Martin is an advanced doctoral student at the University of North Texas. Her research interests include risk assessments with adolescent offenders and the clinical assessment of malingering.

Michael J. Vitacco, Ph.D., recently received his doctorate in clinical psychology from the University of North Texas. He is currently a postdoctoral fellow in forensic psychology at the University of Massachusetts Medical Center. His research interests include malingering, psychopathy, and forensic evaluations.