

# Normative Data for the $F(p)$ Scale of the MMPI-2: Implications for Clinical and Forensic Assessment of Malingering

Steven E. Rothke, Alan F. Friedman, and  
Alan M. Jaffe  
Northwestern University Medical School

Roger L. Greene  
Pacific Graduate School of Psychology

Martha W. Wetter  
University of Kentucky

Patricia Cole and Kathryn Baker  
Northwestern University Medical School

Data from several clinical samples and the Minnesota Multiphasic Personality Inventory—2 standardization group are presented to familiarize the reader with response patterns of different groups on a new validity scale designed to assist in the identification of exaggeration or fabrication of psychological disturbance. Sensitivity-specificity analyses are included along with suggestions for use of the  $F(p)$  Scale with other validity scales. Cautions about setting single cutoff scores are also discussed.

The Infrequency-Psychopathology Scale,  $F(p)$ , was designed by Arbisi and Ben-Porath (1995) for the Minnesota Multiphasic Personality Inventory—2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) as an additional validity measure, particularly for use with patient populations where a high rate of endorsement of psychological disturbance is expected. They suggested that the  $F(p)$  Scale be used to understand elevations on the  $F$  (Infrequency) Scale, which may be elevated for several reasons including random responding, poor reading comprehension of the items, genuine psychopathology, or an attempt to “fake bad” (i.e., attempt to portray oneself as more impaired than is actually the case). The  $F(p)$  Scale contains 27 items of which 15 overlap with the  $F$  Scale. The  $F$  Scale contains items that were endorsed by less than 10% of the normal standardization group of the original MMPI (Hathaway & McKinley, 1943); it was not constructed of items that were endorsed infrequently by psychiatric patients (a common misconception about the scale). Therefore, an elevated  $F$  Scale score cannot automatically be assumed to be an indication of malingering of illness or of an invalid profile (Friedman, Lewak, Nichols, & Webb, in press). The  $F(p)$  Scale was constructed of items endorsed rarely (20% or less in the keyed direction) by two

separate groups of patients with known psychiatric disturbance and the normal standardization sample for the MMPI-2. Arbisi and Ben-Porath (1995) suggested that when the  $F(p)$  Scale is elevated along with the  $F$  Scale, the clinician can more confidently attribute the high scores to a patient's attempt to overreport psychopathology if other validity measures such as Variable Response Inconsistency (VRIN) and True Response Inconsistency (TRIN; Butcher et al., 1989) are not elevated significantly. More recent investigations have found that the  $F(p)$  Scale is less influenced by diagnostic group membership and the confounding effects of distress and psychopathology than is the  $F$  Scale, and it is more effective than the  $F$  Scale in distinguishing groups with genuine psychopathology from those asked to feign psychiatric impairment (Arbisi & Ben-Porath, 1997, 1998). The present article provides information on the  $F(p)$  Scale with additional patient groups often giving rise to questions about the validity of their MMPI-2 profiles and genuineness of their clinical presentations. The Appendix contains all of the items and scoring key for the  $F(p)$  Scale.

## Method

Table 1 contains descriptive data, separately for men and women, for the  $F(p)$  Scale for the MMPI-2 standardization sample and a large variety of clinical samples, some of which are derived from other published studies. For each group, data were excluded for participants whose scores on VRIN were clinically elevated (raw score  $>14$ ) on the basis of guidelines suggested by Greene, Gwin, and Staal (1997). A brief description of the samples follows.

The standardization (normative) sample for the MMPI-2 is described in detail in the test manual (Butcher et al., 1989). The psychiatric inpatient sample (mostly psychotic disorders), the inpatients from the alcohol treatment program at St. Mary's Hospital, San Francisco, California, and the sample of individuals applying for Social Security Disability benefits (either Supplemental Security Income, Social Security Disability Insurance, or both) for reason of psychiatric conditions come from Rothke et al. (1994). The sample sizes in the present investigation are smaller than in Rothke et al. because of the use of the VRIN Scale as an additional measure to eliminate invalid profiles (most notably the sample seeking Social

---

Steven E. Rothke, Alan F. Friedman, Alan M. Jaffe, Patricia Cole, and Kathryn Baker, Department of Psychiatry and Behavioral Sciences, Northwestern University Medical School; Roger L. Greene, Department of Psychology, Pacific Graduate School of Psychology; Martha W. Wetter, Department of Psychology, University of Kentucky.

Martha W. Wetter is now at Bluegrass East Comprehensive Care Center, Lexington, Kentucky. Patricia Cole is now at Center for Pain Studies, Rehabilitation Institute of Chicago. Kathryn Baker is now at Veterans Affairs Puget Sound Health Care System, Seattle, Washington.

We would like to thank Charles E. Metz and Aaron C. Rothke for their assistance with data analysis.

Correspondence concerning this article should be addressed to Steven E. Rothke, 3330 Dundee Road, Suite N-1, Northbrook, Illinois 60062. Electronic mail may be sent to s-rothke@nwu.edu.

Table 1  
 Descriptive Data for the Normative and Clinical Samples and Mean Raw Scores for the  $F(p)$  Scale and Other Validity Indexes

Patient group	n	Age	$F(p)$		F	$F_B$	F-K
			M	Range			
Standardization sample (M)	1,137	41.71 (15.33)	1.16 (1.38)	0-10	4.52 (3.23)	1.86 (2.43)	-10.77 (6.65)
Standardization sample (W)	1,462	40.52 (15.24)	1.03 (1.26)	0-9	3.66 (2.91)	1.94 (2.58)	-11.36 (6.34)
Psychiatric inpatients (M)	282	34.29 (11.20)	3.54 (2.95)	0-14	13.49 (9.08)	11.04 (7.89)	1.55 (12.90)
Psychiatric inpatients (W)	315	36.19 (12.25)	3.72 (3.12)	0-16	12.73 (7.79)	11.01 (7.52)	0.86 (10.93)
St. Mary's alcoholics (M)	166	38.21 (13.87)	1.29 (1.44)	0-7	6.84 (4.13)	5.30 (4.44)	-6.65 (8.06)
St. Mary's alcoholics (W)	96	39.45 (14.69)	1.75 (1.68)	0-7	7.90 (5.12)	7.14 (5.63)	-5.16 (8.75)
Tampa VA alcoholics (M)	1,343	46.88 (12.07)	2.86 (3.01)	0-19	13.81 (9.90)	9.97 (8.24)	2.40 (13.45)
Tampa VA alcoholics (W)	101	38.10 (11.51)	2.47 (2.81)	0-17	9.95 (8.03)	7.35 (7.40)	-3.07 (11.38)
Sequoia alcoholics (M)	49	38.57 (9.22)	1.22 (1.92)	0-12	6.04 (3.82)	3.53 (3.70)	-9.92 (6.58)
Sequoia alcoholics (W)	22	34.18 (7.77)	1.82 (1.87)	0-5	9.05 (5.74)	8.23 (6.58)	-4.55 (8.04)
Soc. Sec. ben. applic. (M)	83	—	5.85 (3.93)	0-15	22.57 (10.99)	17.72 (8.61)	13.97 (12.41)
Soc. Sec. ben. applic. (W)	29	—	5.55 (3.70)	0-13	19.97 (10.00)	17.03 (7.70)	11.17 (12.01)
Nonlitigating TBI (M)	18	33.67 (13.21)	1.61 (1.50)	0-5	5.00 (2.93)	2.00 (2.00)	-11.88 (6.17)
Nonlitigating TBI (W)	9	30.55 (14.94)	0.77 (0.67)	0-2	4.00 (1.58)	1.78 (1.59)	-12.89 (4.76)
Litigating TBI (M)	15	39.00 (11.51)	2.47 (3.85)	0-15	10.67 (7.98)	8.73 (9.00)	0.07 (9.79)
Litigating TBI (W)	14	59.00 (13.19)	1.36 (1.34)	0-5	6.43 (3.86)	3.43 (3.39)	-8.71 (6.87)
Pain program (M)	51	—	1.55 (1.82)	0-11	6.78 (5.43)	3.84 (5.07)	-9.75 (9.28)
Pain program (W)	64	—	1.27 (1.71)	0-10	4.69 (3.92)	2.59 (2.93)	-12.30 (7.02)
Paranoid schiz (M)	10	37.00 (7.01)	2.00 (1.49)	0-5	10.80 (2.70)	5.50 (3.14)	-3.80 (8.22)
Paranoid schiz (W)	9	43.67 (9.89)	4.44 (1.81)	2-7	15.67 (5.68)	11.89 (6.85)	2.44 (8.71)
Sim paranoid schiz (M)	13	31.46 (6.85)	15.53 (4.80)	8-24	42.31 (9.68)	29.15 (6.70)	35.31 (10.30)
Sim paranoid schiz (W)	7	37.86 (7.84)	18.14 (4.78)	10-24	46.43 (7.44)	32.85 (6.79)	41.43 (8.00)
Borderline pers dis (M)	6	31.17 (10.07)	1.50 (2.07)	0-5	9.67 (5.65)	7.00 (5.10)	-2.00 (7.57)
Borderline pers dis (W)	30	31.70 (11.89)	3.56 (2.34)	0-9	12.60 (7.47)	11.03 (6.18)	0.90 (10.30)
Simulated bord pers dis (M)	6	25.67 (3.98)	6.33 (5.89)	0-15	19.00 (12.85)	18.67 (11.81)	8.83 (15.74)
Simulated bord pers dis (W)	36	32.56 (11.30)	6.80 (5.02)	0-18	19.44 (12.49)	17.81 (10.56)	9.55 (14.77)
PTSD (M)	23	41.87 (7.30)	3.13 (3.61)	0-16	16.04 (9.12)	12.00 (7.82)	5.26 (12.15)
PTSD (W)	5	36.00 (2.35)	2.00 (2.35)	0-5	9.60 (2.88)	11.80 (4.27)	0.80 (4.32)
Simulated PTSD (M)	7	34.57 (6.90)	10.71 (5.62)	2-20	35.86 (10.19)	28.71 (7.65)	30.14 (9.54)
Simulated PTSD (W)	13	34.92 (8.11)	10.92 (7.66)	1-24	34.15 (17.11)	25.15 (13.48)	25.77 (18.47)

Note. Numbers in parentheses are standard deviations. Dashes indicate that data were not available.  $F(p)$  = Infrequency-Psychopathology Scale;  $F$  = Infrequency Scale;  $F_B$  = Infrequency-Back Scale;  $F-K$  =  $F-K$  Index. M = men; W = women; VA = Veterans Affairs; Soc. Sec. ben. applic. = Social Security benefits applicants (for psychiatric disability); TBI = traumatic brain injury; Paranoid schiz = paranoid schizophrenia; Sim paranoid schiz = simulated paranoid schizophrenia; Borderline pers dis = borderline personality disorder; Simulated bord pers = simulated borderline personality disorder; PTSD = posttraumatic stress disorder.

Security benefits). The samples from the Tampa Veterans Administration Medical Center and Sequoia Hospital, Redwood City, California, are of inpatients from traditional alcohol treatment programs. The pain management sample consisted mostly of patients with low back pain or myofascial pain disorder who were tested at the beginning of a comprehensive 4-week pain treatment program (mean age was 41.93 years, mean level of education was 13.4 years, and mean interval since the time of injury was 51.6 months). The patients with traumatic brain injury (TBI) either in litigation or nonlitigating came from Berry et al. (1995). The samples of patients with posttraumatic stress disorder or paranoid schizophrenia, and the groups simulating these conditions, were derived from Wetter, Baer, Berry, Robison, and Sumpter (1993). Patients with borderline personality disorder and those simulating the disorder came from a study in progress by Martha W. Wetter.

## Results

Considerable variability is noted across the samples listed in Table 1 with the highest endorsement frequencies noted, not surprisingly, in groups asked to simulate disorders. Also elevated, compared with nonpatients and psychiatric inpatients, are the scores by patients seeking Social Security Disability benefits because of psychiatric disability. This was not unexpected given the

incentive these patients have for appearing impaired (although this is not to suggest that all patients seeking Social Security benefits exaggerate their disability). Most of the other groups obtained scores comparable to the MMPI-2 standardization sample. In light of the fact that the Infrequency-Back Scale,  $F_B$ , a new validity measure (Butcher et al., 1989), with the MMPI-2 is also used to make determinations about the validity of the second half of a participant's MMPI-2 responses, it is included along with the  $F$  Scale and  $F-K$  Index in Table 1 to compare the  $F(p)$  Scale with other major validity indices.

Table 2 lists percentiles corresponding to selected  $F(p)$  Scale raw scores to give the reader a better understanding of the sample distributions and what percentage of each sample exceeded arbitrary cutoff scores.

We performed a Pearson correlation analysis to explore the relationships among the  $F$ ,  $F_B$ , and  $F(p)$  Scales. Because of the small size of many of the samples, only correlations for the normative and psychiatric inpatient groups are reported here. Within the standardization sample, correlations between the  $F$  and  $F(p)$  Scales were .61 and .54, and correlations between the  $F_B$  and

CLINICAL USE OF THE  $F(p)$  SCALE

Table 2  
Cumulative Percentage of Participants Scoring at or Below Selected Raw Score Cutoffs on the  $F(p)$  Scale by Group and Gender

Patient group	<i>n</i>	0	1	2	3	4	5	7	10	12	15
Standardization sample (M)	1,137	41	69	85	93	98	99	99	100		
Standardization sample (W)	1,462	44	74	88	95	98	99	99	100		
Psychiatric inpatients (M)	282	13	30	43	57	70	80	88	98	99	100
Psychiatric inpatients (W)	315	12	24	43	58	68	77	87	97	98	99
St. Mary's alcoholic patients (M)	166	36	69	81	92	97	98	100			
St. Mary's alcoholic patients (W)	96	26	52	78	82	92	96	100			
Tampa VA alcoholic patients (M)	1,343	18	41	59	71	79	84	91	97	98	99
Tampa VA alcoholic patients (W)	101	20	50	66	75	85	87	94	98	99	99
Sequoia Hospital alcoholics (M)	49	41	71	88	94	98	98	98	98	100	
Sequoia Hospital alcoholics (W)	22	32	59	68	73	86	100				
Soc. Sec. ben. applic. (M)	83	1	13	23	33	43	57	70	86	92	100
Soc. Sec. ben. applic. (W)	29	3	14	21	28	52	59	76	90	90	100
Nonlitigating TBI (M)	18	28	56	72	89	94	100				
Nonlitigating TBI (W)	9	33	89	100							
Litigating TBI (M)	15	40	53	60	80	87	93	93	93	93	100
Litigating TBI (W)	14	21	71	86	93	95	100				
Pain management program (M)	51	28	59	84	90	96	98	98	98	100	
Pain management program (W)	64	42	67	84	92	95	98	98	100		
Paranoid schizophrenia (M)	10	10	50	60	90	90	100				
Paranoid schizophrenia (W)	9	0	0	22	33	44	67	100			
Simulated paranoid schizophrenia (M)	13	0	0	0	0	0	0	0	23	38	38
Simulated paranoid schizophrenia (W)	7	0	0	0	0	0	0	0	14	14	29
Borderline personality disorder (M)	6	50	67	67	83	83	100				
Borderline personality disorder (W)	30	3	23	37	57	70	77	93	100		
Simulated borderline personality (M)	6	33	33	33	33	33	50	50	83	83	100
Simulated borderline personality (W)	36	3	14	19	31	39	47	64	78	86	89
PTSD (M)	23	9	44	61	65	78	91	91	96	96	96
PTSD (W)	5	40	60	60	60	80	100				
Simulated PTSD (M)	7	0	0	14	14	14	14	14	57	71	86
Simulated PTSD (W)	13	0	15	15	31	31	31	39	46	46	77

Note.  $F(p)$  = Infrequency-Psychopathology Scale; M = men; W = women; VA = Veterans Affairs; Soc. Sec. ben. applic. = Social Security benefits applicants (for psychiatric disability); TBI = traumatic brain injury; PTSD = posttraumatic stress disorder.

$F(p)$  Scales were .51 and .55 for men and women, respectively ( $p < .001$  for all values). For the inpatient psychiatric sample, correlations between the  $F$  and  $F(p)$  Scales were .74 and .74, and correlations between the  $F_B$  and  $F(p)$  were .67 and .60 for males and females, respectively ( $p < .001$  for all values). These findings indicate that the  $F(p)$  Scale is highly correlated with both the  $F$  and  $F_B$  Scales, although it only shares from 25% to 37% of the variance (in the normative sample) with these two scales. These correlations support the hypothesis of Arbisi and Ben-Porath (1995) that the  $F(p)$  Scale measures something different than the  $F$  and  $F_B$  Scales. The largest correlations were obtained in the simulated samples, as might be expected, because each of these validity measures is sensitive to the effects of overreporting of symptoms.<sup>1</sup> Another possibility is that a floor effect and restriction of range for each of these scales existed in the normative sample, which may have suppressed the correlations among them.

In keeping with the approach used by Arbisi and Ben-Porath (1998), a sensitivity-specificity analysis was run to examine the clinical utility of the  $F(p)$  Scale in distinguishing genuine psychiatric patients from simulators. This analysis was also undertaken to evaluate whether the  $F(p)$  Scale adds any useful information beyond use of the  $F$  Scale alone in discriminating genuine from malingered presentations. In Table 3, analyses are displayed for

two different base rates of malingering. In this set of analyses, groups were collapsed across gender. The malingering group consisted of the simulated posttraumatic stress disorder, borderline personality disorder, and paranoid schizophrenia samples for both base rate analyses. The genuine patient group consisted of the three samples of actual patients with these same disorders for the 50% base rate analysis. For the 11% base rate analysis, the Big Spring Hospital sample was included in the genuine patient group. As can be seen in the table, a slight advantage is obtained by using the  $F(p)$  over the  $F$  Scale, although not to the extent noted by Arbisi and Ben-Porath; the overall values were also less than those obtained by Arbisi and Ben-Porath. The latter findings may be due to the fact that Arbisi and Ben-Porath used patients as their own malingered comparison group (within-subject design); the present analysis was a between-subjects design, which more closely reflects what is actually encountered in clinical practice (i.e., deciding which group a patient belongs to rather than distinguishing which of a patient's two performances are malingered).

<sup>1</sup> The entire set of correlations for each of the samples can be obtained by writing to Steven E. Rothke.

Table 3  
Positive Predictive Power, Negative Predictive Power, Sensitivity, and Specificity for the *F* and *F(p)* Scales Across Different Base Rates of Malingering

Scale and raw cut score	Approximate <i>t</i> cut score	PPP	NPP	Sensitivity	Specificity	Correct classification
Base rate = .50 (82 simulators and 83 patients)						
<i>F</i> $\geq$ 14	$\geq$ 80	.65	.75	.78	.59	.68
<i>F(p)</i> $\geq$ 5	$\geq$ 80	.73	.74	.74	.73	.74
<i>F</i> $\geq$ 17	$\geq$ 90	.75	.76	.72	.76	.74
<i>F(p)</i> $\geq$ 6	$\geq$ 90	.83	.74	.70	.86	.78
<i>F</i> $\geq$ 20	$\geq$ 100	.79	.70	.65	.83	.74
<i>F(p)</i> $\geq$ 8	$\geq$ 100	.93	.71	.61	.95	.78
Base rate = .11 (82 simulators and 680 patients)						
<i>F</i> $\geq$ 14	$\geq$ 80	.19	.96	.78	.59	.68
<i>F(p)</i> $\geq$ 5	$\geq$ 80	.22	.96	.74	.69	.70
<i>F</i> $\geq$ 17	$\geq$ 90	.23	.95	.72	.70	.71
<i>F(p)</i> $\geq$ 6	$\geq$ 90	.29	.96	.70	.79	.78
<i>F</i> $\geq$ 20	$\geq$ 100	.26	.95	.65	.70	.77
<i>F(p)</i> $\geq$ 8	$\geq$ 100	.39	.95	.61	.88	.85

Note. PPP = positive predictive power, or the proportion of participants with test scores at or exceeding the cutoff who were actually from the malingering group; NPP = negative predictive power, or the proportion of participants with test scores below the cutoff who were actually from the genuine patient group; sensitivity = the proportion of malingering participants who obtained scores at or above the cutoff; specificity = the proportion of genuine patients who obtained scores below the cutoff.

A set of receiver operating characteristic (ROC) analyses were run, paralleling the above sensitivity-specificity analyses.<sup>2</sup> Figures 1 and 2 present the ROC curves generated by the *F* and *F(p)* Scales, respectively, when used to discriminate the groups simulating specific

disorders from the groups of actual patients with these diagnoses. Figures 3 and 4 present similar information when the groups simulating specific disorders are contrasted with the larger psychiatric group containing the same three diagnoses as well as inpatients. The *F* and *F(p)* Scales depart significantly from the line of no information in both analyses, with the *F(p)* Scale accounting for a slightly greater area under the curve in each case. However, the differences between the scales did not reach statistical significance. Readers unfamiliar with ROC techniques are referred to Swets (1992) for a detailed description on their use and interpretation.

We ran a set of analyses on the nonlitigating and litigating patients with TBI because the MMPI-2 is being used increasingly to make determinations about possible malingering of disability in these patient populations. A simple *t* test between the means of the litigating and nonlitigating groups (collapsed across gender) was nonsignificant. The apparent difference between men and women in the litigating group was essentially due to the effect of a single extreme score by a male participant. These preliminary findings with the *F(p)* Scale with patients with TBI suggest no differences between litigating and nonlitigating participants or by gender, however, this does not suggest that the *F(p)* Scale lacks use in making distinctions between participants malingering the effects of TBI and those with genuine injury (patients in litigation cannot be assumed to be malingering psychological disturbance).

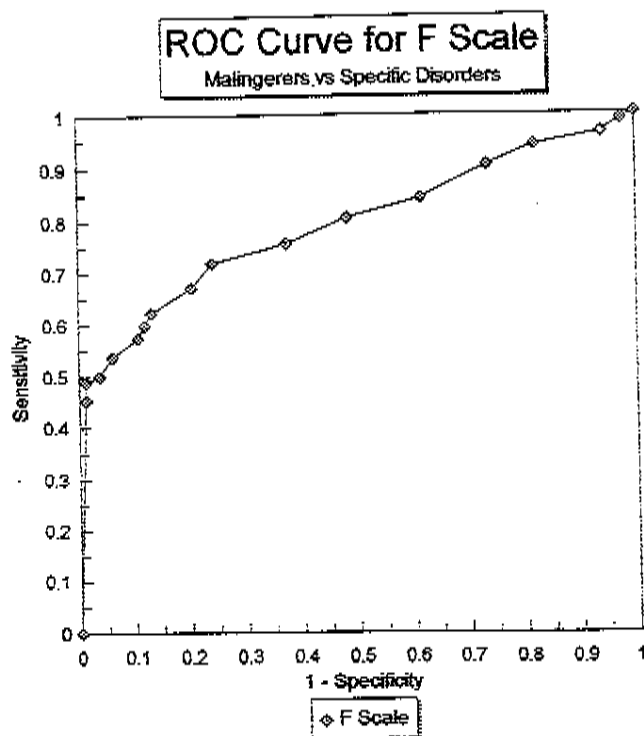


Figure 1. Receiver operating characteristic (ROC) curve for the *F* Scale based on data from Table 3. The abscissa (1 - specificity) is the false positive rate, and the ordinate (sensitivity) is the true positive rate.

<sup>2</sup> ROCFIT (Metz, Shen, Wang, Kronman, & Herman, 1991) and other receiver operator characteristic (ROC) analysis software were obtained from Charles E. Metz, Department of Radiology, University of Chicago Medical Center, 5841 South Maryland Avenue, Chicago, Illinois 60637.

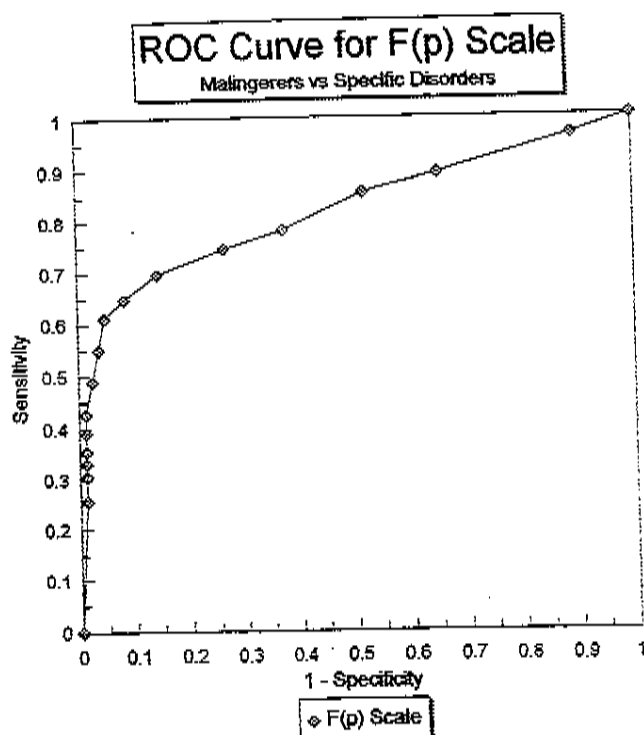


Figure 2. Receiver operating characteristic (ROC) curve for the  $F(p)$  Scale based on data from Table 3.

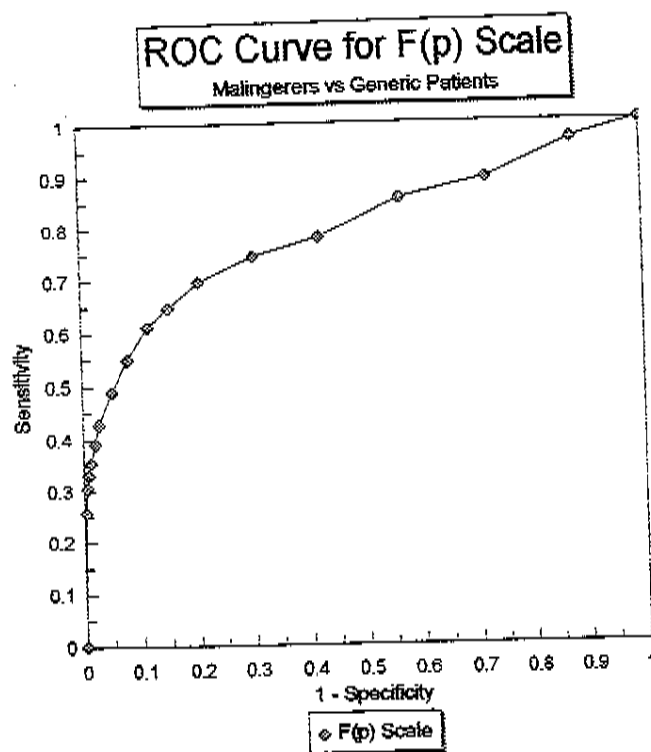


Figure 4. Receiver operating characteristic (ROC) curve for the  $F(p)$  Scale based on data from Table 3.

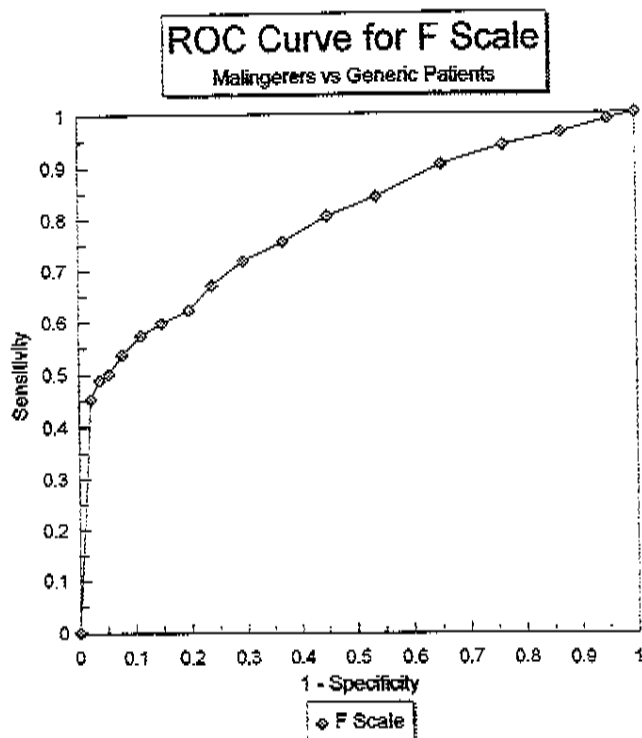


Figure 3. Receiver operating characteristic (ROC) curve for the  $F$  Scale based on data from Table 3.

### Discussion

The present study extends the work of Arbisi and Ben-Porath (1995, 1997, 1998) to patient groups that clinicians are likely to encounter in both clinical and forensic contexts. Elevations on the  $F(p)$  Scale are infrequent in most clinical samples. Elevations are much more common when malingering or exaggeration of psychopathology is expected. Consistent with the findings of Arbisi and Ben-Porath (1997), this study appears to support the finding that the  $F(p)$  Scale is relatively independent of psychiatric diagnoses and is less sensitive than the  $F$  or  $F_B$  Scales to specific psychopathology. These results provide additional support for the use of the  $F(p)$  Scale in determining the validity of MMPI-2 profiles in clinical and forensic practice; however, the sensitivity-specificity and ROC analyses did not indicate a significant advantage of the  $F(p)$  Scale over the  $F$  Scale as was suggested in the investigations by Arbisi and Ben-Porath (none of these earlier studies used the ROC technique).

It is recommended that the  $F(p)$  Scale be studied further, but its applied use should be governed by using a sequence of steps as well as observing cautions about its potential limitations. Arbisi and Ben-Porath (1995) recommended that the following sequence be used in interpreting an elevated  $F$  Scale: (a) rule out random responding and acquiescence by eliminating profiles with elevated VRIN ( $t \geq 100$ ) or TRIN ( $t \geq 80$ ) scores, (b) rule out malingering or exaggeration by considering whether the  $F(p)$  Scale is elevated, and (c) if Steps a and b are negative, then a high  $F$  Scale can be considered consistent with psychopathology. An elevated  $F_B$  Scale score should be interpreted with the same sequence. To this se-

quence we would add the use of another validity measure in Step b, the  $F-K$  Index, for which new guidelines were recently published for the MMPI-2 (Rothke et al., 1994).

As an afterthought, the present investigation included a set of analyses with what was termed the  $F(p)'$  Scale, which consists of the 12 items of the  $F(p)$  Scale that do not overlap with the  $F$  Scale. The purpose was to determine whether removal of item overlap between the  $F(p)$  and  $F$  Scales would result in a superior tool for separating malingering from genuine patients. Little difference was noted in terms of discrimination of samples using the shorter scale than the 27-item  $F(p)$  Scale; therefore, the  $F(p)$  Scale is recommended to avoid the problem of attenuation of reliability. The Appendix contains all of the items and scoring key for the  $F(p)'$  Scale.<sup>3</sup>

Many of the sample sizes in the present study are very small; therefore, cutoff or guideline scores based on these groups can only be advisory. In addition, rigid cutoff scores invariably lead to errors in classification (Dwyer, 1996). We recognize that specific clinical settings reflect different base rates of psychopathology, and it is critical that clinicians understand the nature of their population with regard to frequency and type of psychopathology before interpreting the  $F(p)$  Scale or any other scale score. Arbisi and Ben-Porath (1997) recommended that additional studies be conducted in settings where clear incentives to feign symptoms exist. We suggest that future investigations of samples of psychiatric or injured patients control for motivation to appear impaired (e.g., as might be the case in patients involved in litigation, seeking disability compensation, or attempting to avoid criminal prosecution) by carefully interviewing the patient or by record review, as the  $F(p)$  Scale appears sensitive to efforts to overreport symptoms. In several cases in the present study, participants who were either in the process of litigation and seeking compensation or feigning psychiatric conditions produced elevated  $F(p)$  Scale scores, thereby supporting the construct validity of the scale. This is not to suggest, however, that all litigating participants exaggerate their complaints. When used with these caveats, the  $F(p)$  Scale can be a useful addition to the assessment of MMPI-2 profile validity.

<sup>3</sup> Readers can obtain descriptive data for all the samples for  $F(p)'$  by writing to Steven E. Rothke.

## 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.

- Arbisi, P. A., & Ben-Porath, Y. S. (1997). Characteristics of the MMPI-2  $F(p)$  scale 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.
- Berry, D. T., Wetter, M. W., Baer, R. A., Youngjohn, J. R., Gass, C. S., Lamb, D. G., Franzen, M. D., MacInnes, W. D., & Buchholz, D. (1995). Overreporting of closed-head injury symptoms on the MMPI-2. *Psychological Assessment*, 7, 517-523.
- Butcher, J. N., Dahlstrom, W. G., Graham, J. R., Tellegen, A., & Kaemmer, B. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.
- Dwyer, C. A. (1996). Cut scores and testing: Statistics, judgment, truth, and error. *Psychological Assessment*, 8, 360-362.
- Friedman, A. F., Lewak, R., Nichols, D. S., & Webb, J. T. (in press). *Psychological assessment with the MMPI-2*. Mahwah, NJ: Erlbaum.
- Greene, R. L., Gwin, R., & Staal, M. (1997). Current status of MMPI-2 research: A methodologic overview. *Journal of Personality Assessment*, 68, 20-36.
- Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory Manual*. Minneapolis: University of Minnesota Press.
- Metz, C. E., Shen, J. H., Wang, P. L., Kronman, H. B., & Herman, B. A. (1991). ROCFIT [Computer Software]. Chicago: University of Chicago.
- Rothke, S. E., Friedman, A. F., Dahlstrom, W. G., Greene, R. L., Arredondo, R., & Mann, A. W. (1994). MMPI-2 normative data for the  $F-K$  Index: Implications for clinical, neuropsychological, and forensic practice. *Assessment*, 1, 1-15.
- Swets, J. A. (1992). The science of choosing the right decision threshold in high-stakes diagnostics. *American Psychologist*, 47, 522-532.
- Wetter, M. W., Baer, R. A., Berry, D. T., 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.

## Appendix

### $F(p)$ and $F(p)'$ Items and Scoring Key

A. $F(p)$	True: 66, 114, 162, 193, 216, 228, 252, 270, 282, 291, 294, 322, 323, 336, 371, 387, 478, 555
	False: 51, 77, 90, 93, 102, 126, 192, 276, 501
B. $F(p)'$	True: 193, 291, 322, 523, 371, 387, 478, 555
	False: 51, 77, 93, 501

Note. The  $F(p)$  Scale scoring key copyright 1995 by the Regents of the University of Minnesota.

Received January 15, 1998  
Revision received March 1, 2000  
Accepted March 15, 2000 ■