Evading Detection on the MMPI-2: Does Caution Produce More Realistic Patterns of Responding?

Donald J. Viglione, Danielle Mellin Wright, Nicole T. Dizon, Jayne E. Moynihan, Sheila DuPuis and Todd D. Pizitz

Assessment 2001; 8; 237
DOI: 10.1177/107319110100800301

ASSESSMENT
2001, Volume 8, Number 3, 237-250

EVADING DETECTION ON THE MMPI-2: DOES CAUTION PRODUCE MORE REALISTIC PATTERNS OF RESPONDING?

Donald J. Viglione
Danielle Mellin Wright
Nicole T. Dizon
Jayne E. Moynihan
Sheila DuPuis
Todd D. Pizitz
California School of Professional Psychology, San Diego and Alliant University

Studies on MMPI and MMPI-2 malingering indexes often sacrifice generalizability in an attempt to control internal validity. This study improves external validity while still maintaining internal validity by providing graduate student participants with a realistic context for malingering on the MMPI-2 (n = 94) and MMPI (n = 30). Contextual parameters include a realistic life predicament, psychological knowledge, an incentive, the presence versus absence of a specific diagnosis, and a caution to be realistic. This study found that cautioning participants not to overexaggerate their responses significantly improves their ability to evade detection on the MMPI-2 and MMPI. Standard malingering indexes (Infrequency, *F*; Back Side, *F, Fb*; *F*-Correction, *F-K*; and Infrequency-Psychopathology, *F(p)*) were insufficiently sensitive in identifying simulators using common cutoff scores for these cautious simulators.

Keywords: Malingering, MMPI-2, response styles, caution, depression

Donald J. Viglione, Clinical PsyD Program; Danielle Mellin Wright, San Diego, California; Nicole T. Dizon, San Diego, California; Jayne E. Moynihan, San Diego, California; Sheila DuPuis, San Diego, California; Todd D. Pizitz, San Diego, California.

National Computer Systems provided funding to score the MMPI-2 protocols in this study. Thanks to Richard Lewak, Kay DiFrancesca, Peter Briggs, and Ed Bienkowski for allowing their students to participate in this study. Thanks to Shayna Gothard, San Diego, California, for her critique of the manuscript.

Correspondence concerning this article and requests for offprints should be addressed to Donald J. Viglione, Clinical PsyD Program, California School of Professional Psychology, San Diego and Alliant University, 6160 Cornerstone Court East, San Diego, CA 92121.

Malingering is defined as a person's deliberate attempt to fabricate or grossly exaggerate symptoms or impairment for the purposes of personal gain (American Psychiatric Association, 1994). The Minnesota Multiphasic Personality Inventory (MMPI; Hathaway & McKinley, 1943) and its revision (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) are the most frequently used instruments to detect malingering (Rogers, Sewell, & Salekin, 1994; Schretlen, 1988). A standard in most psychological assessment batteries, this multi-scale inventory was the first measure to recognize the influence of participants'

response styles and to attempt to identify and quantify response styles, such as malingering, through the use of validity scales (Rogers, Bagby, & Chakraborty, 1993).

Symptom fabrication or exaggeration provides malingerers the possibility of avoiding responsibility for their circumstances and behaviors and/or obtaining compensation for their malingered disability. Individuals may malinger to avoid criminal and civil sanctions, procure a bed in a hospital or improve living conditions, evade military responsibility, secure prescription medication, or obtain financial compensation for disability. Estimations of the prevalence and cost of malingering vary depending upon the specific context, yet there is agreement that it is widespread in criminal investigations, civil litigation, insurance claims, workman's compensation claims, and other arenas (Resnick, 1997; Rogers, 1997; Rogers & Cavanaugh, 1983; Rogers, Sewell, & Goldstein, 1994).

Contextual Characteristics of Malingering

Malingerers attempt to match a specific disorder and a specific set of events with a reasonable life predicament. They must identify which symptoms are indicative of the disorder they are attempting to feign and which symptoms would be considered excessive or atypical. Malingerers, such as those involved in criminal and civil litigation or those who are without optimal room and board, most likely encounter many opportunities to test out malingering roles and strategies. For these individuals, malingering emerges across time and may become more subtle and successful over a developmental course. Malingerers may even come to believe their feigned roles, as in common childhood experience many of us have become confused about whether we were actually sick in our secret efforts to avoid upsetting demands and challenges.

Studies using actual malingerers are rare because malingerers typically do not identify themselves as malingering. Therefore, malingering studies often rely on simulators, or individuals pretending to malinger[1] in an effort to improve internal validity. Standard simulation studies utilizing the MMPI-2 validity scales indicate that these indexes work reasonably well in detecting malingering, often approaching 90% to 100% accuracy when using cutoff scores that are chosen post hoc (Graham, 1998). In studies comparing malingerers or simulators with psychiatric samples, average effect sizes ranged from 1.08 to 3.33 (Cohen's *d*) depending on the sample and validity indexes used (Rogers, et al., 1994). The generalizability of these studies (Rogers, 1988; Schretlen, Wilkins, Van-Gorp, & Bobholz, 1992) to an actual malingering population is questionable due to potentially lower rates of sensitivity and specificity (Viglione, Fals-Stewart, & Moxham, 1995).

To improve the generalizability of MMPI-2 simulation studies, in this study, conditions were designed to be more representative of malingering in realistic life contexts. We identified and attempted to reproduce five contextual parameters that might differentiate simulators from actual malingerers[2]: diagnostic specificity, a realistic life predicament, psychological knowledge, an incentive, and a caution to be realistic. Realistic, successful, difficult to detect malingerers (those requiring psychological assessment) most likely have assimilated personal experience, observation of others, instructions from those who share their predicament, and response shaping from interactions with evaluators to produce a more accurate and sophisticated symptom presentation.

Coaching or Caution?

Within the past 5 years, several studies have focused on the impact that coaching has on a malingerer's ability to evade detection on the MMPI-2 (Bagby et al., 1997; Storm & Graham, 2000; Walters & Clopton, 2000). Although researchers are concerned that studies on coaching may compromise test validity, the results reveal that coaching helps people successfully malinger and that individuals who would benefit from malingering are frequently given information on how to fake. In a survey assessing attorneys' thoughts regarding coaching clients, 63% of attorneys expressed the belief that they had a duty to provide

[1]In this study, the term *malingering* will be used for faking mental disorder specificity among real-life fakers, whereas the term *simulators* is reserved for role players in experimental studies.

[2]Our definition of context is more multi-dimensional than that used by Rogers & Cruise (1998), Who define a context as a "Screening" p. 275).

their clients with information on validity indexes (Baer, Wetter, & Berry, 1995; Graham, 1993; Lees-Haley, 1997). Almost half of these attorneys felt they should provide their clients with as much information as possible.

Recent studies reveal that individuals coached on the validity indexes for both faking good and malingering on the MMPI/MMPI-2 are significantly better at evading detection than individuals who have not been coached (Baer et al., 1995; Bagby et al., 1997; Graham, Watts, & Timbrook, 1991; Rogers et al., 1993; Storm & Graham, 2000; Walters & Clopton, 2000). Storm & Graham (2000) demonstrated positive predictive power ranging from .83 to .96 with uncoached simulators; with coached simulators, positive predictor power ranged from .54 to .86. When participants are knowledgeable about validity indexes, the sensitivity of the MMPI-2 decreases, and the optimal cutoff scores for identifying malingerers are lower. Rogers, Bagby, and Chakraborty (1993) found that two thirds of coached simulators attempting to fake schizophrenia were able to avoid detection using cutoff scores for the *F*, *Fb*, and *F-K* indexes, and one third of coached simulators were able to escape detection using Wiener and Harmon's, year, Obvious and Subtle Subscales (*O-S*).

How much coaching is necessary to enable malingerers to successfully evade detection on the MMPI-2, or, in other words, to limit the sensitivity of the MMPI-2 in the detection of malingering? Baer, Wetter, and Berry (1995) suggested that minimal coaching was sufficient to allow participants to convincingly fake good, but similar comparisons are lacking in the malingering literature. The current studies attempt to determine whether a simple straightforward caution, as opposed to in-depth coaching on the malingering indexes, will suffice to help simulators evade detection on the MMPI-2 and MMPI.

Experiment 1

Method

Participants

The sample consisted of 94 first and second year clinical graduate students (26 men, 68 women) from the California School of Professional Psychology, San Diego. The mean age was 29.34 years (*SD* = 7.25 years). None of these students had previously taken a course on the MMPI or MMPI-2. However, all of the participants had taken a graduate level psychopathology course, suggesting they were knowledgeable about psychological disorders and may have had some exposure to cursory reviews or descriptions of the MMPI-2.

Measures

Malingering indexes. Malingering indexes for this study include *F* (Infrequency Scale), *Fb* (Back Side *F* scale), *F-K* raw score difference, *O-S* (Wiener and Harmon's, Obvious and Subtle Subscales), and the more recently developed *F*(*p*) (Infrequency-Psychopathology scale; Lim & Butcher, 1996; Rogers et al., 1994; Rogers et al., 1994). *F*, *Fb*, and *F-K* scales were included due to their prevalence throughout the malingering literature. Additionally, the *F* scale has often been identified by researchers as the best index in identifying malingerers (Graham, 1998; Strong, Greene, & Schinka, 2000). Wiener and Harmon's *O-S* subscale was included in this study because research with realistic life contexts (Greene, 1991; Rogers et al., 1994; Viglione et al., 1995) found the *O-S* comparisons to be more effective than other validity indexes.

Researchers believe that the *F*(*p*) scale (Arbisi & Ben-Porath, 1995) has strong potential to identify malingerers more effectively than other validity indexes (Graham, 1998; Rothke et al., 2000). It was developed by choosing items that were infrequently endorsed by both psychiatric populations and normal populations, and it is reported to have low correlations with other indexes used to detect malingerers (Rogers, Sewell, & Ustad, 1995). Graham (1998) found that *F*(*p*) was effective identifying 93% of uncoached fakers and 76% of coached fakers, suggesting that the *F*(*p*) might be the scale that is the least affected by cautioning participants to be realistic.

Cutoff scores. Throughout the malingering literature, researchers have identified a wide range of scale cutoff scores, and attempts to crossvalidate specific, optimal cutoff scores have met with mixed success (Rogers et al., 1995). For the purpose of this study, a priori cutoff scores were derived from Greene's (2000) Validity Profile for Clinical Settings. They

were selected to minimize false positives from the patient sample. Both high and low cutoff scores for men and women are presented to evaluate the sensitivity of each of the indexes at different stringency levels, with high cutoff scores measured at the 99th percentile and low cutoff scores measured at the 95th percentile of clinical populations (see Table 1).

Specificity. To examine the specificity of our results, we compared the known simulators from our study to men and women in a clinical sample. Data for the clinical sample was derived from a subsample of Caldwell's clinical data set, which at the time of this analysis was comprised of 52,543 psychiatric inpatients and outpatients whose protocols were scored by the Caldwell report (Greene, 2000).[3] We set the specificity level at 95% and then examined sensitivity rates.

Procedure

Data were gathered over 4 years and included three separate cohorts of students. In each cohort, the researchers presented the study to the potential graduate student participants in their classes. A $50 bookstore credit was offered as a positive incentive for the best faker in each cohort. Participants were informed in general terms of the study and asked to sign an informed consent to participate in the research. They were apprised of their right to withdraw from the study at any time. After agreeing to participate, participants were given a packet of materials containing a demographic form and MMPI-2 testing materials. They completed the forms in the classroom, library, or in another room on campus.

To improve external validity, all participants were placed in a realistic life predicament in that they were asked to feign a mental disorder to retain disability payments for an on-the-job injury. The independent variables were Caution (caution to avoid detection vs. no caution) and Disorder Specificity (general psychological disturbance vs. depression). Participants were randomly assigned to one of four groups: Depressed/No Caution, General/No Caution, Depressed/Caution, or General/Caution. Entire role instructions for these four groups are included in the Appendix. Role descriptions instructed participants to feign either a depressive or a general psychological disturbance. The Caution instructions advised participants to be cautious ("You realize that if you present too dramatically, it will look fake, and you will lose any chance of continued coverage. Therefore, use caution to maintain a convincing, realistic profile."), whereas the No Caution instructions omitted this passage.

Immediately following the administration of the MMPI-2, participants were given a post-test questionnaire assessing their understanding of the role play, how convincing the role play was to them, their level of motivation to succeed in the role play, the impact of the incentive, and their faking strategy. After completing the study, participants were given a copy of their MMPI-2 profile and role and given a description of the four different roles in the study. After all profiles had been analyzed, the best *faker* in each cohort received a $50 credit for the bookstore.

[3]More complete information on the Caldwell clinical data set can be found in chapter 9 of Green's (2000) *The MMPI-2: An Interpretive Manual*.

Table 1
Cutoff Scores From Caldwell's Clinical Data Set[a]

	Men		Women	
	95th %ile	99th %ile	95th %ile	99th %ile
F	20	27	18	24
Fb	16	21	17	23
O-S	209	274	196	255
F-K	10	20	9	18
F(p)	5	7	5	7

Note. *F* = Infrequency scale; *Fb* = Back Side *F*; *O-S* = Obvious and Subtle subscales; *F-K* = *F*-Correction; *F(p)* = Infrequency–Psychopathology scale.
[a]Greene, 2000.

Results

To control for inconsistent response styles, six protocols with Variable Response Inconsistency (VRIN) or True Response Inconsistency (TRIN) *T* scores greater than 80 were eliminated prior to statistical analysis, reducing the sample size to 88 (Graham, 1993). Means and standard deviations for the malingering indexes are displayed in Table 2. Univariate *F* tests are also reported for each of the malingering indicators for Disorder Specificity, Caution, and the interaction between Disorder Specificity and Caution. Given the large number of comparisons, a Bonferroni correction was applied as a conservative measure to control for experiment-wise Type 1 error, and the significance level was set at $p < .001$ (Keppel, 1991).

Table 3 presents the percentage of participants identified as simulators in the Caution and No Caution groups. To improve generalizability, cutoff scores for men and women separately are based upon Caldwell's clinical data set (Greene, 2000). Within the current sample, however, *T*-score mean differences were inconsistent with those seen in Caldwell's data set, with women scoring significantly higher for raw *F*, raw *F-K*, raw *Fb*, and raw *F(p)*, and with no significant difference based upon gender for the *O-S* scale. These gender differences and overall results for men should be interpreted cautiously due to the small male sample size, $n = 26$.

Using low cutoff scores within the No Caution group, 84% of the women and only 50% of the men were identified by at least one malingering index. Using low cutoff scores within the Caution group, only 38% of the women and 13% of the men were identified by at least one malingering index. Using high cutoff scores, within the No Caution group, 61% of the women and 40% of the men were identified by at least one malingering index; and within the Caution group, only 9% of the women and none of the men were identified by at least one of the malingering indexes.

In both the Caution and No Caution groups and using both low and high cutoff scores for each gender, men were consistently more difficult to identify as simulators than women. Regardless of cutoff scores, within our sample, men scored significantly lower on each of the malingering indexes than women, $p < .05$. Consistent with the literature, we found the *F* scale to be the most effective validity index regardless of gender, and the *O-S* scale to be the least effective validity index for both genders. In fact, for women, the *O-S* scale was the only malingering index that was unable to discriminate between participants in the Caution and No Caution groups.

Using cutoff scores from Caldwell's clinical sample, with specificity levels set at 95% (Greene, 2000), the raw score *F* was found to be the most sensitive validity index, correctly identifying 48% of the simulators in the total sample, followed by the *F-K* index, which correctly identified 38% of the simulators in the total sample. We found the raw score *Fb* and *F(p)* to be the least sensitive indexes, identifying 31% and 30% of the simulators, respectively.

Caution and Disorder Specificity Results

The effects of Caution and Disorder Specificity on the MMPI-2 malingering indexes were analyzed using several 2 x 2 analyses of variance (ANOVAs). Magnitudes of effect for the impact of caution on the validity scales were calculated for significant comparisons using Cohen's *d* (see Table 2).

Participants in the Caution group scored significantly lower than those in the No Caution group for each of the malingering indexes. Within the Caution group, the average effect size (see Table 2) for all of the significant malingering indexes was $d = 0.95$, with effect sizes ranging from a low of $d = 0.85$ for the *F(p)* index to a high of $d = 0.99$ for the *F-K* scale. Only one of the malingering indexes, *F(p)*, showed a significant difference between participants based upon disorder specificity, $F(1, 84) = 14.55$, $p < .001$ ($d = 0.56$), and none of the malingering indexes showed a significant interaction between Caution and Disorder Specificity at $p < .001$.

Sensitivity Curves. Sensitivity curves, which graphically illustrate the difference between the Caution and No Caution groups, are presented in Figures 1 and 2 for the *F* raw scale and the *F(p)* index. The *F* raw scale was chosen because of its prevalence in the literature (Graham, 1998), and the *F(p)* index was chosen because of its strong support in recent research (Arbisi & Ben-Porath, 1995; Graham, 1998; Rothke et al., 2000; Storm & Graham, 2000).

Table 2
MMPI-2 Descriptive Data and ANOVA Results for MMPI-2 Malingering Indices

	No Caution		Caution		Disorder specificity		Caution		Disorder specificity X Caution
Scale	General	Depressed	General	Depressed	$F(1, 90)$	Cohen's d	$F(1, 90)$	Cohen's d	$F(1, 84)$
F									
M	27.28	21.39	11.92	14.09	0.09		33.37***	$d = 1.02$	4.22*
SD	13.90	7.52	8.61	5.82					
Fb									
M	19.56	18.17	9.67	10.22	0.067		30.98***	$d = 1.03$	0.36
SD	10.26	6.96	6.99	5.69					
F-K									
M	17.83	8.57	-.88	-1.13	3.57		31.82***	$d = 0.99$	3.20
SD	17.50	9.20	11.32	8.38					
F(p)									
M	9.50	4.29	3.12	2.59	14.55***	$d = 0.56$	28.70***	$d = 0.85$	9.45**
SD	5.72	3.09	3.15	1.67					
O-S									
M	197.67	162.43	109.54	107.78	1.40		20.74***	$d = 0.86$	1.14
SD	93.44	56.89	85.83	52.03					

Note. $N = 88$. *F* = raw Infrequency scale; *Fb* = raw *F* back side; *F-K* = raw *F* minus raw correction; *F(p)* = raw psychopathology; *O-S* = obvious scales minus subtle scales, total *T*-score difference. Participants with Variable Response Inconsistency (VRIN) or True Response Inconsistency (TRIN) scores > 80 were excluded from analysis.
*$p < .05$. **$p < .01$, included for reference. ***$p < .001$, significant using Bonferroni correction.

Table 3
MMPI-2 Malingering Indexes Percentage Identified as Simulators

Malingering index		Women[a]			Men[b]	
		No Caution	Caution		No Caution	Caution
F	*F* > 18	77.4%	34.4%	*F* > 20	50.0%	13.3%
	F > 24	54.8%	9.4%	*F* > 27	10.0%	None
Fb	*Fb* > 17	50.0%	18.6%	*Fb* > 16	50.0%	None
	Fb > 23	32.3%	None	*Fb* > 21	30.0%	None
O-S	*O-S* > 196	54.8%	18.8%	*O-S* > 209	30.0%	None
	O-S > 255	16.1%	3.1%	*O-S* > 274	10.0%	None
F-K	*F-K* > 9	63.6%	36.4%	*F-K* > 10	40.0%	None
	F-K > 18	35.3%	6.3%	*F-K* > 20	10.0%	None
F(*p*)	*F*(*p*) > 5	50.0%	18.8%	*F*(*p*) > 5	40.0%	6.7%
	F(*p*) > 7	37.5%	6.3%	*F*(*p*) > 7	10.0%	None
Exceeded low cutoff on at least one index		83.9%	37.5%		50.0%	13.3%
Exceeded high cutoff on at least one index		61.3%	9.4%		40.0%	None

Note. High cutoff scores are measured at the 99th percentile; low cutoff scores are measured at the 95th percentile. *F* = raw Infrequency scale; *Fb* = raw *F* back side; *F-K* = raw *F* minus raw Correction; *F*(*p*) = raw psychopathology; *O-S* = obvious scales minus subtle scales. Participants with Variable Response Inconsistency (VRIN) or True Response Inconsistency (TRIN) *T* scores > 80 were excluded from analysis.
[a]$n = 68$. [b]$n = 26$.

These sensitivity curves visually illustrate the poor sensitivity for these measures of malingering. Consistent with the ANOVA findings, they also show the marked difference in detection of simulators between the Caution and No Caution groups.

Validity of the Experimental Intervention. As a confirmation of the internal validity of our study, we looked at several of the clinical scales to determine whether results for Caution and Disorder Specificity were in the expected direction. Table 4 presents the means and standard deviations for the clinical scales within the Caution and Disorder Specificity groups. Means are in the expected direction for the Caution variable, with participants in the Caution group scoring lower on all clinical scales than participants in the No Caution group. Disorder Specificity results were also in the expected direction. Participants assigned to simulate a depressive disturbance scored significantly higher on the Depression scale, $F(1, 84) = 46.34$, $p < .001$, than participants assigned to simulate a general psychological disturbance. There was a large effect size between these groups, $d = 1.18$, indicating that participants were able to feign the specific symptoms of depression when that was their specified goal. Participants in the Caution group, who were advised that if they responded too dramatically they would be detected, scored significantly lower on the Schizophrenia scale, $F(1, 84) = 29.09$, $p < .001$, $d = 0.99$. Because this scale comprises the most severe symptomatology, it is likely that cautious malingerers more easily recognize items from this scale than from other scales and endorse them less frequently.

The results of the post-test questionnaire further supported the validity of the experimental intervention. Seventy-two percent of the participants believed the situation was realistic, endorsing the item, "The written description of the role play was convincing to me," and 85% reported, "I was motivated to succeed in my role play." Fifteen percent of the sample were unmotivated to succeed in the role-play and were excluded prior to any analysis because they obtained TRIN or VRIN *T* scores greater than 80.

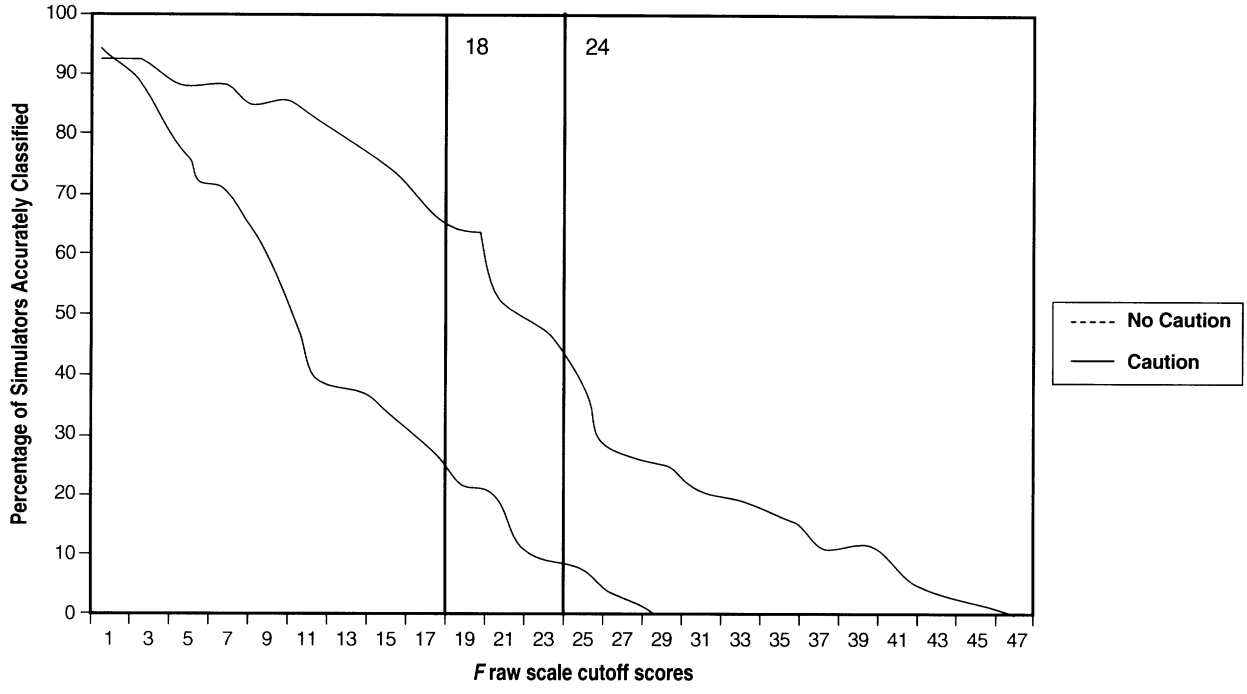

Figure 1. MMPI-2 sensitivity curve for the Infrequency (*F*) raw scale showing the percentage of simulators accurately classified using high and low cutoff scores in the Caution and No Caution groups.

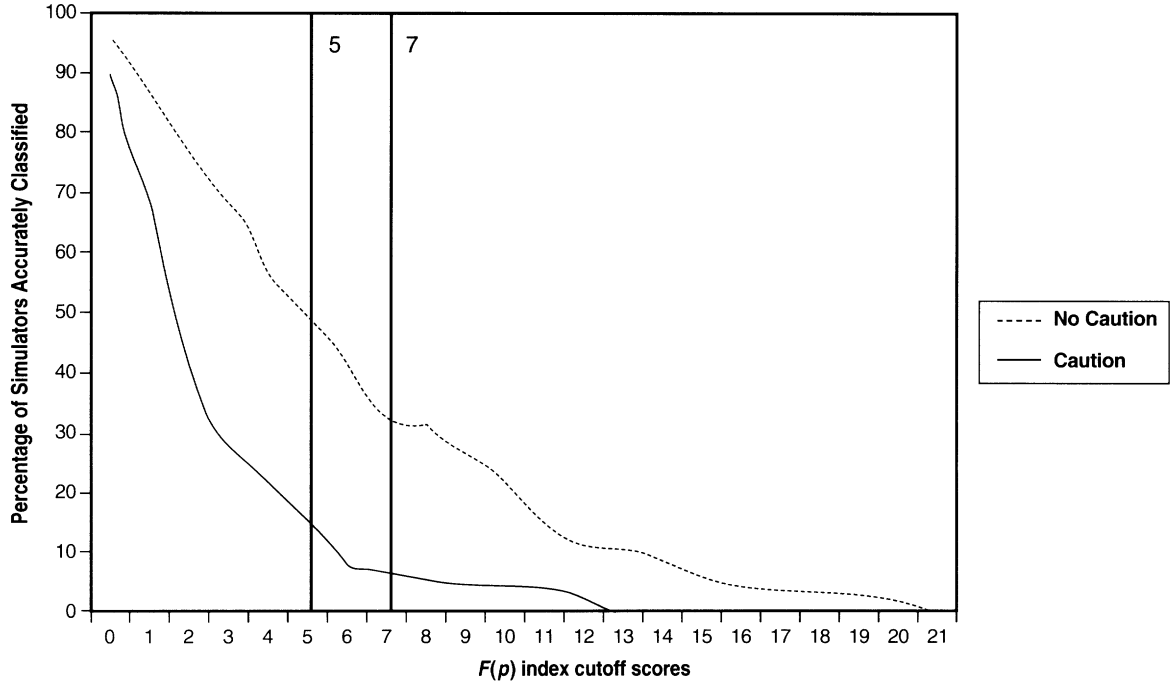

Figure 2. MMPI-2 sensitivity curve for the Infrequency-Psychopathology *F*(*p*) index showing the percentage of simulators accurately classified using high and low cutoff scores in the Caution and No Caution groups.

Results of the post-test questionnaire also suggested the participants in the Caution and No Caution groups approached the test with different mindsets. When provided with a brief caution to be realistic, simulators scored significantly lower on the question "Attempting to fake on the MMPI was confusing for me," $F(1, 120) = 4.35$, $p < .05$. They also reported using a more consistent strategy, $F(1, 120) = 10.82$, $p < .001$. In addition, participants in the Caution group scored significantly lower on the item, "My strategy in faking was to endorse all MMPI items related to my role," $F(1, 120) = 4.04$, $p < .01$, suggesting that cautioned simulators were more conservative in their response strategies. Interestingly, the simulators in the Caution group also reported being significantly less motivated by the incentive than the simulators in the No Caution group, $F(1, 120) = 5.03$, $p < .05$. As there was not a significant difference between the simulators in the Caution and No Caution groups in terms of overall motivation to succeed, it is likely that the two groups had different, although equal, motivation to succeed in the roleplay.

Experiment 2

Method

Participants

A second study was undertaken to increase the confidence in the results from the first study. As a replication of Experiment 1, archival data were analyzed in which the original MMPI was administered to a small sample of graduate students ($n = 30$) using the exact same procedures as those used in Experiment 1. The sample consisted of 30 first and second year clinical graduate students (9 men, 21 women) from the California School of Professional Psychology, San Diego. The mean age was 28.62 years ($SD = 6.37$ years). As in the initial study, none of the students had completed a course on the MMPI or MMPI-2; however, all had taken a graduate level psychopathology course.

Measures

We chose the three available MMPI malingering indexes (*F*, *F-K*, and *O-S*) for the replication. Consistent with Experiment 1, we used Caldwell's

Table 4
Descriptive Data and ANOVA Results for the MMPI-2 Clinical Scales

	No Caution				Caution			
	General		Depressed		General		Depressed	
Scale	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Hs	83.21	14.14	79.23	11.15	76.46	15.93	72.91	7.73
D[a]	83.37	15.15	101.68	7.33	77.50	18.35	97.78	9.32
Hy	77.79	12.95	77.59	12.18	75.63	17.00	75.78	12.28
Pd[b]	71.84	14.85	81.77	9.93	67.17	11.98	71.83	9.00
Pa[b]	90.63	23.60	77.36	12.34	73.83	15.02	71.13	11.34
Pt[a, b]	78.90	14.30	85.27	8.05	67.17	12.71	79.65	11.59
Sc[b]	95.74	23.17	91.32	12.78	70.92	16.76	78.35	11.76
Ma[a]	63.79	18.38	46.00	7.57	52.75	11.02	43.61	9.03
Si	72.53	12.65	79.09	9.47	63.58	14.82	73.61	11.26

Note. All *T* scores are K-corrected. *Hs* = Hypochondriasis; *D* = Depression; *Hy* = Hysteria; *Pd* = Psychopathic Deviate; *Pa* = Paranoia; *Pt* = Psychasthenia; *Sc* = Schizophrenia; *Ma* = Hypomania; *Si* = Social Introversion. Participants with Variable Response Inconsistency (VRIN) or True Response Inconsistency (TRIN) *T* scores > 80 were excluded from analysis.
[a]Main Effect Disorder specificity, $p < .001$. [b]Main Effect Caution, $p < .001$.

cutoff scores for men and women, setting sensitivity at the 95th percentile for low cutoff scores and the 99th percentile for high cutoff scores (Greene, 2000).

Results

Table 5 presents the percentage of participants identified as simulators by both high and low cutoff scores for men and women in the Caution and No Caution groups. Results were consistent with our findings in the original MMPI-2 study in that all three malingering indexes demonstrated insufficient sensitivity within the Caution group. Additionally, participants in the replication study (Experiment 2) who were in the Caution group escaped detection on all of the malingering indexes much more frequently than participants in the No Caution group. As in our original study and consistent with the literature, the *F* raw score proved to be the most effective malingering index, identifying 59% of the participants in the No Caution group and 27% of the participants in the Caution group, using the low cutoff scores.

Table 5
MMPI Malingering Indices Percentage Identified as Simulators by Low and High Cutoff Scores

Scale	No Caution	Caution
Low *F*	58.8%	26.7%
High *F*	29.4%	13.3%
Low *F-K*	52.9%	26.7%
High *F-K*	23.5%	6.7%
Low *O-S*	11.8%	6.7%
High *O-S*	None	None

Note. High cutoff scores are measured at the 99th percentile; low cutoff scores are measured at the 95th percentile using Caldwell's cutoff scores for men and women (Greene, 2000). *F* = Raw Infrequency scale; *F-K* = raw *F* minus raw Correction; *O-S* = *Obvious* scales minus *Subtle* scales.

Discussion

These studies were attempts to place simulating malingering in a context more closely resembling reality. The proposed real-life parameters included a realistic life predicament, psychological knowledge, and an incentive to malinger. We also explored the impact of having participants fake a specific disorder as opposed to a general psychological disorder, and of cautioning participants to be realistic. The results call into question the standard MMPI-2 malingering indexes' sensitivity in detecting malingering when simulators are provided with real-life contexts and cautioned to avoid exaggerating.

The most compelling results in our studies pertain to the caution variable. Merely cautioning our simulators to present a realistic profile and not to exaggerate resulted in unacceptably low sensitivity for the MMPI-2 and MMPI. The sensitivity was especially poor for male simulators, with none of the cautious male simulators being detected using the high cutoff scores and only 13% of the male simulators being detected using the low cutoff scores.

To evaluate the importance of caution within simulation and malingering contexts, we compared the effect sizes for caution in our study to those for simulators in the MMPI-2 research. Rogers, Sewell, and Salekin (1994) reported that simulation and malingering effect sizes vary considerably in malingering studies. Within these studies there are three types of comparisons: simulators versus control participants, simulators versus patients, and within subjects studies in which the same participants completed the MMPI-2 with both standard and fake-bad instructions. Across studies, the effect sizes average about $d = 2.5$ (Rogers et al., 1994). Our effect sizes were approximately half of this average effect size and are quite large considering that the only difference between our groups is whether or not the groups were briefly cautioned to maintain a realistic profile.

The results of this study suggest that a large portion of the variance in simulated faking can be accounted for by caution. Real-life malingerers often have many opportunities to test and refine their malingering roles so they become more subtle and successful over time. It is therefore likely that real-world malingerers are cautious and more closely resemble our cautious simulators than simulators who are not cautioned to be realistic. Thus, consistent with suspicions in the literature, real-life malingering effect sizes may be much smaller than those in simulation studies.

These findings support and expand upon recent concerns raised in the literature on the ability of standard validity scales of the MMPI or MMPI-2 to

detect coached malingerers (Baer et al., 1995; Bagby et al., 1997; Rogers et al., 1993; Storm & Graham, 2000). Further, our study replicates findings by Baer et al. (1995) that extensive and detailed coaching may be unnecessary to see a remarkable increase in ability to evade detection on the MMPI-2. A simple statement regarding caution may be sufficient for simulators to present themselves realistically.

The authors were further interested in determining which indexes would be most effective in identifying simulators when there was a realistic context, a psychologically knowledgeable group, and an incentive to malinger irrespective of caution or disorder specificity. Despite the realistic life context, no index worked particularly well in identifying simulators within our sample, and all of the malingering indexes were highly correlated, providing redundant information. Across both the Caution and No Caution groups, the *F* scale was the most effective malingering index in this study. Examining sensitivity when specificity was set at 95% in Caldwell's clinical data set (Greene, 2000), we found that the *F* scale correctly identified 48% of all simulators. *F-K* was the next most effective index, identifying 38% of the simulators overall. Given the recent cogent reviews of *F(p)* within the literature, we expected that it would stand out as an exceptional scale in identifying malingerers (Arbisi & Ben-Porath, 1995; Graham, 1998; Rothke et al., 2000); however, the *F(p)* index was not an outstanding index within our sample, identifying only 30% of simulators across groups.

In attempting to ensure that our participants possessed adequate psychological knowledge, we chose graduate students who had taken a course in psychopathology but had not yet completed a MMPI-2 course. It is possible that our results are limited because the sample was exclusively graduate students, and caution might have more of an impact on psychologically sophisticated respondents than on actual malingerers who may not have attained this level of education.

As an incentive, we offered the best faker in each cohort a $50 bookstore credit. Based upon our post-test questionnaire, this was not an adequate incentive. Over 48% of the participants did not agree with the statement, "The incentives motivated me to strive to succeed in my role play." It is therefore highly probable that the generalizability of our incentive to other contexts, such as a forensic setting, may be limited, and it is not comparable to an incentive to avoid criminal sanctions or to obtain disability rewards (Storm & Graham, 2000). On the other hand, the participants who carefully and consistently responded to MMPI-2 items may be motivated for other reasons.

The sample in the present study is exclusively simulators. Therefore, to examine specificity and positive and negative predictive power, the results of this research were compared with data from Caldwell's clinical sample (Greene, 2000). Although this provides a comparative framework for our study, conclusions are limited because Caldwell's normative sample includes malingerers and forensic cases, which were not identified or controlled for in the current study. We recognize that the base rates for malingering may therefore differ secondary to the use of more than one data set. Future research might include control groups to draw comparisons between two groups that are similar in composition and follow a similar testing procedure. Conclusions must also be considered tentative when looking at differences between men and women because the sample was disproportionately women. Although the male sample was disproportionately small, this study found that males scored significantly lower on each of the malingering indexes than females, a finding that deserves further scrutiny in the research. Gathering the data over several years likely increased the possibility that there are cohort effects in the data set, but it also increased the generalizability of the results.

The findings of this study permit an increased understanding of the effect of coaching individuals to be cautious when answering MMPI-2 items. More research is needed to crossvalidate these findings with additional populations. Ancillary measures and other data sources should be considered when assessing for malingering, and cutoff scores should be considered preliminary and utilized with the aforementioned considerations.

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.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Baer, R. A., Wetter, M. W., & Berry, D. T. R. (1995). Effects of information about validity scales on underreporting of symptoms on the MMPI-2: An analogue investigation. *Assessment, 2*, 189-200.

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

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.

Graham, J. R. (1993). *MMPI-2: assessing personality and psychopathology* (2nd ed.). New York: Oxford University Press.

Graham, J. R. (1998). *Assessing malingering on the MMPI-2*. Paper presented at the First Annual Assessment Psychology Conference, Orlando, FL.

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

Greene, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual*. Needham Heights, MA: Allyn & Bacon.

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

Hathaway, S. R., & McKinley, J. C. (1943). *The Minnesota Multiphasic Personality Inventory, second edition*. Minneapolis: University of Minnesota Press.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Lees-Haley, P. R. (1997). Attorneys influence expert evidence in forensic psychological and neuropsychological cases. *Psychological Assessment, 4*, 321-324.

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, 27*, 1-25.

Resnick, P. J. (1997). Malingered psychosis. In R. Rogers (Ed.), *Clinical assessment of malingering and deception* (2nd ed., pp. 47-67). New York: Guilford.

Rogers, R. (Ed.). (1988). *Clinical assessment of malingering and deception*. New York: Guilford.

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

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

Rogers, R., & Cavanaugh, J. L. (1983). "Nothing but the truth": A reexamination of malingering. *The Journal of Psychiatry & Law, 11*(4), 443-459.

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

Rogers, R., Sewell, K. W., & Goldstein, A. M. (1995). Explanatory models of malingering: A prototypical analysis. *Law and Human Behavior, 18*(5), 543-552.

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: A systematic examination of fake-bad indicators. *Psychological Assessment, 2*, 81-89.

Rothke. S. E., Friedman, A. F., Jaffe, A. M., Greene, R. L., Wetter, M. W., Cole, P., & Baker, K. (2000). Normative data for the *F*(*p*) scale of the MMPI-2: Implications for clinical and forensic assessment of malingering. *Psychological Assessment, 12*(3), 335-440.

Schretlen, D. J. (1988). The use of psychological tests to identify malingered symptoms of a mental disorder. *Clinical Psychology Review, 8*, 451-476.

Schretlen, D., Wilkins, S., Van-Gorp, W. G., & Bobholz, J. H. (1992). Cross-validation of a psychological test battery to detect faked insanity. *Psychological Assessment, 4*, 77-83.

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

Strong, D. R., Greene, R. L., & Schinka, J. A. (2000). A taxometric analysis of MMPI-2 Infrequency Scales [*F* and *F*(*p*)] in clinical settings. *Psychological Assessment, 12*(2), 166-173.

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*(2), 183-199.

Viglione, D. J., Jr., Fals-Stewart, W., & Moxham, E. (1995). Maximizing internal and external validity in MMPI malingering research: A study of a military population. *Journal of Personality Assessment, 65*, 502-513.

Appendix

Malingering Instructions, Caution versus No Caution and General versus Depressed

All four groups are provided with the following life predicament:

You are an administrator at a small, well-established firm. Your boss has been trying to cut expenses by having the cleaning crew work before regular work hours are over, thus, getting the job done at a cut rate. You have repeatedly informed him that this is not a safe working condition for the employees, but he has not changed the procedure. One day, near the end of the day, you are leaving to do a special errand for your boss. As you cross a freshly mopped floor, you slip and fall, landing hard on your tailbone.

As a result, you have been out of work for 2 weeks on disability and continue to experience a fair amount of pain, particularly when you have to sit for any length of time. The Workers Compensation physician insists he can find nothing to explain the pain and refuses to authorize any more time off or disability payments, stating that you are able to return to work, a job that requires long periods of time sitting at the computer. You are angry at your boss for the injury you have and frustrated at the physicians' apparent collusion with your boss to unreasonably limit your recovery time (thereby cutoff his disability payments).

Before terminating your case, the physician refers you to the staff psychologist for a routine evaluation. You correctly realize that this evaluation is your only opportunity to remain on disability under this employer's obligation. You have no additional coverage and need an income until you are fully recovered. You also feel that your boss is responsible and that the money should come from the company through Workers Compensation.

You know well that Workers Compensation will continue providing benefits to patients who are psychologically disturbed as a result of a work-related accident. This would not be too unusual because you had tried to take measures to avoid the problem, and now you are suffering as a result of your boss' negligence.

Instructions specific to assigned groups:

*General/*Caution

It is obvious that your only choice is to present yourself *as psychologically disturbed on the MMPI-2 (General)* which the psychologist has given you. However, you also realize that if you present too dramatically, it will look fake, and you will lose any chance of continued coverage. Therefore, use caution to maintain a convincing, realistic profile (Caution).

Please attempt to fill out the following MMPI-2 to present yourself as *psychologically disturbed* as a result of your accident, thereby allowing you to remain on disability. Carefully study each item as you answer.

*Depressed/*Caution

It is obvious that your only choice is to present *yourself as having a significant depressive disorder (Depression)* on the MMPI-2 which the psychologist has given you. However, you also realize that if you present too dramatically, it will look fake, and you will lose any chance of continued coverage. Therefore, use caution to maintain a convincing, realistic profile (Caution).

Please attempt to fill out the following MMPI-2 to present yourself *as having a significant depressive disorder* as a result of your accident, thereby allowing you to remain on disability. Carefully study each item as you answer.

*General/*No Caution

It is obvious that your only choice is to present yourself as *psychologically disturbed* on the MMPI-2 which the psychologist has given you.

Please attempt to fill out the following MMPI-2 to present yourself as *psychologically disturbed* as a result of your accident, thereby allowing you to remain on disability. Carefully study each item as you answer.

*Depressed/*No Caution

It is obvious that your only choice is to present yourself as having a *significant depressive disorder* on the MMPI-2 which the psychologist has given you.

Please attempt to fill out the following MMPI-2 to present *yourself as having a significant depressive disorder* as a result of your accident, thereby allowing you to remain on disability. Carefully study each item as you answer.