THE DETECTION OF MALINGERING AND DECEPTION WITH A SHORT FORM OF THE MMPI-2 BASED ON THE L, F, AND K SCALES

JEFFREY E. CASSISI AND DON E. WORKMAN

Illinois Institute of Technology

This project examined the use of a short inventory based on the validity subscales from the MMPI-2. In Study 1, 69 subjects responded to the extracted L, F, and K items and to those items when embedded in the complete MMPI-2. Correlations between administrations were equivalent to the original test-retest reliabilities for these scales. In Study 2, 60 subjects responded to the short-form with instructions to be honest, fake good, or fake bad. The correct classification rate was 77% using the standard MMPI-2 rules of L > 70T for faking good and F > 70T for faking bad. Taken together, these findings suggest that the short-form may be useful alone or when combined with other questionnaires to identify potential problems in self-report.

Psychological assessment relies extensively on questionnaires and self-report measures. For these methods to yield maximally accurate information, it is important to have a notion of the style with which the respondent approached the task. Ideally, after reading and considering each item, the respondent gives a direct and honest answer (Graham, 1990, p. 22). However, in some situations individuals deviate from this ideal, and in extreme situations this deviation invalidates the resultant information. Psychological questionnaires have been criticized for being susceptible to distortion and for not providing an index of test-taking attitude (Wiggins, 1973).

One objective of this research is to suggest a short inventory that can be combined with other questionnaires in order to identify potential response problems. Such a device would increase the veracity of otherwise accepted instruments that are lacking such an index. A common clinical practice is to generalizethe patient's response style on the MMPI validity scales to other questionnaires given within the assessment battery. For example, a patient may be suspected of overreporting depression symptoms on an instrument like the Beck Depression Inventory (Steer & Beck, 1988) if at the same time he or she obtains an extremely high score on the MMPI-2 F scale. This research examines this practice more closely.

Butcher and Hostetler (1990) reviewed the most popular short-forms for the MMPI and discussed the problems with these techniques and their inapplicability to the MMPI-2. They did not discuss the extraction of particular scales for special applications. In particular, a short-form based on the validity scales would be useful in making assumptions with regard to a patient's approach to other assessment devices.

The second objective of this research is to examine the effectiveness of the MMPI-2 validity scales in identifying individuals who are trying to mangle and deceive. Many studies have used the original MMPI to detect malingering in forensic and psychiatric settings (e.g., Greene, 1988; Berry, Baer, & Harris, 1991). These studies relied heavily on the validity scales for classification.

Meehl and Hathaway (1946) and Gough (1947) were among the first to describe methods to study the effects of malingering and deception on self-report measures such

All correspondence should be addressed to Jeffrey E. Cassisi, Department of Psychology, Illinois Institute of Technology, IIT Center, Chicago, IL 60616.
as the MMPI. They instructed volunteers to simulate how they would answer if they were intending to mangle or deceive. This method has been used many times since. The problem with this approach is the questionable generalizability of the simulated responses to clinical and "real world" situations due to differing levels and sources of motivation. Other approaches require independent knowledge of faking, which is not always available or completely accurate. Rogers (1988) in his treatise on malingering and deception concluded that progress in this area requires replication of findings under both clinical and simulated conditions.

Grow, McVaugh, and Eno (1980) evaluated the effectiveness of 13 different techniques for the detection of faking on the MMPI. They found that student volunteers instructed to fake good, fake bad, or answer honestly produced reliable group differences. They concluded that the best formula for detecting faking was the F-K index described by Gough (1950). Grow et al. (1980) then cross validated their findings through review of MMPI scores from patients who had been categorized retrospectively as having faked good, faked bad, or answered honestly. According to Rogers (1988), this study is important because their instructions to volunteers take "into account that subjects present a believable disorder or condition" (p. 310). Therefore, subjects were not simply told to appear as mentally ill as they could because that would lead to extreme and irrelevant results. Rather, subjects were instructed to be as believable as possible within a particular scenario.

Graham, Watts, and Timbrook (1991) reported a study in which 105 undergraduate volunteers were asked to take the MMPI-2 with instructions to fake. Specifically, they were told to present themselves as though they had serious psychological and emotional problems or to try and present a very positive impression, as if they were being evaluated for a job that they really wanted. The fake-good instruction set produced average profiles with L scale T-scores of approximately 73 to 75, and the fake-bad instruction set produced average profiles with F scale T-scores of approximately 107 to 116. (Also see Graham, 1990, pp. 44-48.) Graham et al. (1991) cross-validated their findings prospectively with 50 psychiatric patients with a variety of diagnoses. They found that these clinical populations required an upward adjustment to the cut-off score on the F scale to prevent an unacceptable high false-positive rate for faking bad.

**METHOD AND RESULTS**

**STUDY 1**

In Study 1, 69 volunteers from undergraduate psychology classes were asked to answer the complete MMPI-2 and the short-form including only the 102 items from the validity scales in return for extra credit. The sample was 74% male, 26% female, 69% Caucasian, 8% African-American, 13% Asian, 10% Hispanic, and had an average age of 20 years and a mean of 14 years of education. Subjects' scores on the extracted L, F, and K subscales were correlated with their scores on the complete MMPI-2. Administrations were counterbalanced with a 1-week delay between administrations.

The Pearson product-moment correlation coefficients between the two administrations of the L, F, and K scales were .78, .87, and .83, respectively. All three correlations were significant at \( p < .001 \). These correlations were roughly equivalent to the test-retest reliabilities of .77, .78, and .84, for the 82 male community adults, originally published by the MMPI Restandardization Committee (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989, p. 88). The delay period was also equivalent in this study.

**STUDY 2**

The MMPI-2 validity scales were evaluated by comparing group scores on the short form and several other instruments under simulated conditions. Sixty new volunteers
were solicited from undergraduate psychology classes and offered extra credit. They were assigned randomly to one of three conditions: group 1, answer honestly (honest); group 2, fake in such a way as to deny any psychological symptoms (fake good); or group 3, fake in such a way as to exaggerate any psychological symptoms (fake bad). The hypotheses to be tested followed from the Graham et al. (1991) data: Compared to group 1 (honest) subjects, group 2 (fake good) subjects were expected to have significant elevations on the L scale, and group 3 (fake bad) subjects were expected to have significant elevations on the F scale. The other questionnaires were included to make the assessment situation similar to a psychological test battery and were exploratory in nature.

The sample was 58% male, 42% female, 55% Caucasian, 18% African-American, 17% Asian, and 10% Hispanic and had an average age of 22 years and a mean of 15 years of education. No evidence for significant differences in the distribution of demographic characteristics across groups was obtained by chi-square analyses. Each subject completed the short-form of the MMPI-2, a Beck Depression Inventory (BDI; Steer & Beck, 1988), and a State-Trait Anger Expression Inventory (STAXI; Spielberger, 1988).

Group 1 (honest) was given the following instructions:

**BE OPEN**

We want to reassure you once more that your answers will be anonymous. Read the directions for each questionnaire, and then answer each item as sincerely as possible. Your scores will be anonymous. We are really interested in honesty and how someone responds to these questionnaires when they are being frank and open. In short, we want you to answer these questionnaires as they apply to you as accurately as possible.

Group 2 (fake good) subjects were given the following instructions (after Grow et al., 1980):

**FAKE GOOD**

We want you to take this test in an imaginary situation. You are being cast as a person in an adversary relationship against a psychologist. Read the directions for each questionnaire, and then answer each item as if you were someone who opposes the psychologist. Specifically, imagine a situation in which it would be to your advantage to appear as if you were completely normal and without any psychological disturbance at all. Examples of such a situation would be, if you were trying to secure an early release from prison, trying to secure a release from a mental hospital, or applying for a good job. In short, we want you to answer these questionnaires and deliberately try and *fake good*. Try and be as believable as possible.

Group 3 (fake bad) subjects were given the following instructions (after Grow et al., 1980):

**FAKE BAD**

We want you to take this test in an imaginary situation. You are being cast as a person in an adversary relationship against a psychologist. Read the directions for each questionnaire, and then answer each item as if you were someone who opposes the psychologist. Specifically, imagine you are someone for whom it would be to your advantage to appear as if you had a significant psychological illness. Examples of such a situation would be if you were applying for rehabilitation services, trying to qualify for disability benefits, or trying to beat a charge on the grounds of insanity. In short, we want you to answer these questionnaires and deliberately try and *fake bad*. Try and be as believable as possible.
All subjects were asked to answer a post-test question about their task instructions. This was done as a manipulation check. Every subject correctly identified his or her assignment.

Group mean L, F, and K T-scores are displayed in Table 1. MANOVA revealed significant differences on all measures, F(14,100) = 18.92, p < .001. Subsequent testing with Tukey's HSD revealed that group 2 (fake good) was significantly higher than the other two groups on the L scale (p < .05) and that group 3 (fake bad) was significantly higher than the other two groups on the F scale (p < .05). Tukey's HSD also revealed that group 2 (fake good) and group 3 (fake bad) were significantly different from each other on the K scale (p < .05).

Table 1
Mean T-scores and Standard Deviations for the L, F, and K Scales

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (honest)</th>
<th>Group 2 (fake good)</th>
<th>Group 3 (fake bad)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>L^4</td>
<td>50</td>
<td>10</td>
<td>70</td>
</tr>
<tr>
<td>F^5</td>
<td>54</td>
<td>16</td>
<td>49</td>
</tr>
<tr>
<td>K^6</td>
<td>48</td>
<td>9</td>
<td>52</td>
</tr>
</tbody>
</table>

^4Group 2 (fake good) significantly different from Group 1 (honest).
^5Group 3 (fake bad) significantly different from Group 1 (honest).
^6Group 2 (fake good) significantly different from Group 3 (fake bad).

The classification rate for detecting faking bad or faking good then was computed using the standard rules of a T-score 70 or greater on the L or F scales (described in Butcher et al., 1989, pp. 25-26). Using these rules, 77% of the subjects were classified correctly. The results are displayed in Table 2.

Table 2
Classification Results of Using Standard MMPI-2 Rules of L > 70T for Faking Good and F > 70T for Faking Bad

<table>
<thead>
<tr>
<th>Actual group</th>
<th>Predicted group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1 (honest)</td>
</tr>
<tr>
<td>Group 1 (honest, n = 20)</td>
<td>16 (80%)</td>
</tr>
<tr>
<td>Group 2 (fake good, n = 20)</td>
<td>7 (35%)</td>
</tr>
<tr>
<td>Group 3 (fake bad, n = 20)</td>
<td>1 (5%)</td>
</tr>
</tbody>
</table>

Note.—Total correctly identified: 77%.

DISCUSSION

The results of Study 1 support the notion that subjects respond consistently to L, F, and K scale questions whether they are contained in the full MMPI-2 or the short-form. This suggests that these scales can be extracted without altering the meaning of the questions.

The results of Study 2 support the assertions of Butcher et al. (1989) and Graham (1990) concerning the standard interpretation of T scores above 70 on the L and F scales.
The mean T-scores on the L, F, and K scales obtained here on the short-form replicate those found by Graham et al. (1991), who employed the complete MMPI-2.

The K scale was not particularly useful for distinguishing the groups in this study, which was also the case in the Graham et al. (1991) data. Neither group 2 (fake good) nor group 3 (fake bad) was significantly different from group 1 (honest) on this measure. Unlike the Grow et al. (1980) study, the use of the F-K index was of limited value because of an extremely high false positive rate. That is, the F-K index incorrectly identified 55% of group 1 (honest) subjects as faking good or faking bad.

Many observers have noted that the detection of faking good is more difficult than the detection of faking bad (Graham et al., 1991; Greene, 1988; Grow et al., 1980). This was the case in Study 2 with an obtained hit rate of 55% for group 2 (fake good). Nonetheless, the best balance between hits, misses, false positives, and false negatives was struck by using the standard rules for interpreting the MMPI-2.

These two studies provide an initial experimental impetus for development by the University of Minnesota Press of an instrument based on the validity scales. This would be useful when administration of the complete MMPI-2 is not necessary or when it is contraindicated. Such an instrument probably would be used by a wide range of psychologists who otherwise would rely on more subjective means for evaluating the response style of their patients. No single study is sufficient to establish the sensitivity of an assessment procedure to detect malingering and deception. Further investigation with clinical populations is necessary.

REFERENCES


