8-1-2006

Detecting Intentional, Realistic Response Distortion on the MMPI-2 Using Multiple Validity Scales

Andrew Minton
Western Kentucky University

Follow this and additional works at: http://digitalcommons.wku.edu/theses

Part of the Industrial and Organizational Psychology Commons

Recommended Citation
http://digitalcommons.wku.edu/theses/251

This Thesis is brought to you for free and open access by TopSCHOLAR®. It has been accepted for inclusion in Masters Theses & Specialist Projects by an authorized administrator of TopSCHOLAR®. For more information, please contact topscholar@wku.edu.
DETECTING INTENTIONAL, REALISTIC RESPONSE DISTORTION ON THE MMPI-2 USING MULTIPLE VALIDITY SCALES

A Thesis
Presented to
The Faculty of the Department of Psychology
Western Kentucky University
Bowling Green, Kentucky

In Partial Fulfillment
Of the Requirements for the Degree
Master of Industrial-Organizational Psychology

By
Andrew James Minton

August 2006
DETECTING INTENTIONAL, REALISTIC RESPONSE DISTORTION ON THE MMPI-2 USING MULTIPLE VALIDITY SCALES

Date Recommended: August 10, 2006

Director of Thesis: [Signature]

Dean, Graduate Studies and Research: [Signature]  Date: August 10, 2006
Acknowledgements

To my mother, Linda. Thank you for your encouragement, patience, and example.
# Table of Contents

**Literature Review** ................................................................. 1  
   The MMPI-2 ................................................................. 4  
   Validity Scales .............................................................. 6  
   Faking Research ............................................................ 10  
   The Present Study ........................................................... 13  

**Method** ............................................................................. 15  
   Participants ................................................................. 15  
   Materials ....................................................................... 15  
   Procedure ................................................................. 16  
   Design and Statistical Analyses ........................................... 16  
   Scale Scoring ............................................................... 17  

**Results** ............................................................................ 20  

**Discussion** ................................................................. 23  

**References** ................................................................. 26  

**Appendix A** ................................................................. 29
The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) is a 567-item questionnaire used by psychologists to help diagnose mental disorders. The effectiveness of the MMPI-2 validity scales (F, Fp, F-K, and Ds-r2) were evaluated by asking college students to respond honestly or fake bad on the MMPI-2. Each scale was scored in the usual manner, but as a group they were evaluated in a noncompensatory fashion. That is, failing even one scale was interpreted as a faked response. MMPI-2 protocols of all participants were analyzed to determine if the validity scales were able to accurately identify which subjects faked and which answered honestly. Participants were randomly assigned to a faking bad or honest condition. The participants in the faking group were given brief instruction regarding the symptoms of schizophrenia and were informed of the detection system built into the test. The four validity scales (F, Fp, F-K, and Ds-r2) were able to correctly identify a significant percentage of the test takers as either answering honestly or faking bad.
Literature Review

Self-report measures are the most commonly used type of instrument for assessing personality and psychopathology in patients. Self-report inventories are attractive to test administrators because they are relatively simple to administer and have clear scoring guidelines. The high degree of structure of these inventories also offers strong intersite reliability. One of the negatives of self-report questionnaires is that the opportunity for fabricated responses, whether it be faking, denying, minimizing, or embellishing problems, may be greater as compared to other methods of measurement (Butcher, 2002). As a result, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989), a self report inventory with a true/false report format, offers ways to identify and potentially reduce faking and is the subject of this research study.

The MMPI is currently the most popular measure of clinical personality disorders and uses a true/false format. The MMPI was initially published in 1943 for use in routine diagnostic assessments (Hathaway & McKinley, 1943). Previous personality inventories had been constructed according to a rational keying approach in which items were selected or generated according to face validity and responses were keyed according to the judgment of the test author (Graham, 1993). Items for the MMPI were chosen by the use of an empirical keying approach, that is, an item analysis procedure in which the data determine not only which items are retained but also how the items are scored. With empirical keying, items that correlate with a criterion variable are retained. The MMPI’s criterion was whether the test taker was a member of the normal group or a group with a specific mental disorder. Thus, for the MMPI item analysis, a good item was answered at
different rates by each group. A bad item was answered in roughly the same manner by
members of both groups. The direction of scoring is determined by identifying which
group endorsed the item at a higher rate; that is, if the group identified as suffering from a
personality disorder endorsed the item *more* often that the group free from disorders, test
takers would be awarded a point for a true response. However, if the group identified as
suffering from a personality disorder endorsed the item *less* often that the group free from
disorders, test takers would be awarded a point for a false response.

The projective format of assessing personality, a more indirect and subjective
method, is based on the idea that “when people attempt to understand an ambiguous
stimulus, their own needs, experiences, thought processes, and psychopathology are
projected” (Butcher, 2002, p. 180). Two of the more popular projective tests are the
Thematic Apperception Test (TAT), in which patients are shown pictures and are
encouraged to envisage their own perceived meaning, and the Rorschach Inkblot Test, in
which patients look at inkblots on cards and verbalize everything they imagine the blot
could resemble.

Hathaway and McKinley, the developers of the MMPI, generated a large pool of
potential inventory items by writing 504 personality type statements based on sources
such as psychological and psychiatric case histories and reports, textbooks, and earlier
published scales of personal and social attitudes. A key to the success of an empirical
keying procedure is the criterion variable. A poorly measured criterion variable will lead
a test developer to retain poor items and discard or miscode good ones. For the MMPI the
criterion variable was the clinical diagnosis of the test taker. One group, called Minnesota
normals, consisted of relatives and visitors of patients in the University of Minnesota
hospitals, high school graduates, and medical patients. The second group was made up of psychiatric patients at the University of Minnesota hospitals, representing all of the major psychiatric categories in clinical use at the time of test construction. These people were divided into discrete groups according to their clinically determined diagnostic labels. The subgroups formed were hypochondriasis (Hs), depression (D), hysteria (Hy), psychopathic deviate (Pd), paranoia (Pa), psychasthenia (Pt), schizophrenia (Sc), and hypomania (Ma). At a later time, two additional clinical scales were constructed: the Masculinity-Femininity (Mf) scale (intended to measure the extent to which test takers fit their sex role stereotype) and the Social Introversion (Si) scale (an extraversion/introversion scale).

In addition to the 10 clinical scales, the MMPI also has four validity scales. The four scales are the Cannot Say scale (symbolized as ?), the L scale, the F scale, and the K scale. These scales are designed to identify responses that do not indicate open, honest test-taker behavior. The Cannot Say scale is a simple count of the number of items not answered. Interpretation of results is not advised if more than 10 (Graham, 1993) or more than 30 items (MMPI-2 manual) are left unanswered. The remaining three validity scales are more sophisticated and will be discussed later.

Raw scores on each of the MMPI’s clinical and validity scales (with the exception of the Cannot Say scale) are converted to normalized T scores. Normalized T scores are set to have a mean of 50 and a standard deviation of 10. Given the purpose of the MMPI (the identification of personality disorders), and given that each scale is scored so that high scores indicate a departure from normal functioning, it is the high T score, as opposed to low, that is of interest.
Over time it became clear that a simple interpretation of the MMPI’s scale scores was not sufficient for a new patient’s valid diagnosis, mainly due to the high level of intercorrelation among the clinical scales. For this reason, among others, the MMPI was used in a way other than originally intended. Graham (1993) found the following:

The modified approach to the MMPI treated each of its scales as an unknown, and … the correlates of each scale were identified. According to this approach, when a person obtained a score on a particular scale, the clinician attributed to that person the characteristics and behaviors that through previous research and experience had been identified and researched for other individuals with similar scores on that scale. (p. 7)

The MMPI-2

Until the MMPI-2 was finalized and published in 1989, the MMPI had not been modified since its 1943 establishment. Furthermore, the standardization sample for the original MMPI came from hospital visitors, mostly from the area surrounding Minneapolis, Minnesota. There were additional concerns that the typical test taker in 1980s no longer matched the population from which the original data had been collected.

There were other issues regarding the item content of the MMPI. Some of the terminology had become outdated. Some items included what would now be regarded as sexist language. Other items included deficient syntax due to poor editing. Finally, there was a neglect of clinical issues that had arisen since its origin (e.g., suicide, addiction to drugs). The revision committee went to great efforts to preserve a strong connection between the original and revised version of the MMPI. This connection allowed the
extensive research collected from the time of the original MMPI’s publication to be considered valid for the revised version.

The most important objective of the MMPI-2 revision was to gather a normative sample that would be more representative of the current United States population than the original sample. Secondary issues were to further enhance the MMPI item pool by altering the language of some items, removing other items considered to be offensive in the modern climate, and creating new test items that would “increase the content dimensions of the item pool” (Graham, 1993, p. 10).

Of the 550 unique items on the original MMPI, 15 were rephrased and 82 were amended to eliminate any gender reference. In other items, outdated terminology was supplanted with present day language, and any intrasocial prejudice was removed. The majority of the item alterations were minor, and all changes were made with the intent of preserving the meaning behind the original items while making it more understandable through use of modern terminology. The second significant modification involved creating new items to be included in the revision. The committee removed over 100 items that were not a part of the clinical or validity scales and replaced these items with 154 original items. These new items primarily concerned areas previously neglected (e.g., suicide, addiction to drugs, Type A behavior, etc.).

The published edition of the revised MMPI (the MMPI-2) has 567 items. Final improvements included a standardization sample that more appropriately reflected the national population (based on the 1980 census), modernized item wording, elimination of offensive items, and new scales. The four MMPI validity scales were maintained on the MMPI-2. Three additional validity scales were developed for the MMPI-2: the Back-
Page Infrequency Scale (Fb), the Variable Response Inconsistency Scale (VRIN), and the True Response Inconsistency Scale (TRIN). The Fb scale is designed to function like the F scale but covers the 154 new items that are not covered by the F scale. As with the L, F, and K scales, the Fb scales will be discussed in greater detail below. The final two validity scales, VRIN and TRIN, are not designed to catch intentional response distortion of the fake good or fake bad variety, but rather noncompliant response patterns. The VRIN scale indicates random or inconsistent responses, while the TRIN scale indicates subjects who have responded by consistently answering arbitrarily with either true (acquiescence) or false responses (nonacquiescence).

Validity Scales

Subjects are sometimes motivated to achieve a test score that does not reflect their true self, either in a falsely positive or falsely negative manner. Faking bad, also called malingering, can be done to avoid criminal prosecution or military service. Conversely, test takers may present an overly positive impression of their personalities by faking good. Clearly, a clinician needs to be aware of any faking on the part of the test taker. As mentioned, the original MMPI had four validity scales (? , L, F, and K), the purposes of which were to detect deviations from open, honest responding. The MMPI-2 added three new validity scales (Fb, VRIN, and TRIN).

The L scale was designed to identify purposeful attempts by the test subjects to portray themselves in an overly positive way. These items were developed in order to gauge a person’s aversion to admit even negligible deficiencies in character (e.g., “I get angry sometimes.”). T scores between 55 and 65 on the L scale imply defensiveness. T
scores above 65 indicate a dishonest test-taking attitude or severe defensiveness. Profiles with T scores exceeding 65 on the L scale should not be interpreted.

The F scale was also designed to distinguish individuals who are consciously answering test items in a manner that is not based on their true feelings or judgments. The F scale’s items were selected by examining the answer frequency of the normal group and using the items endorsed in the scored direction by less than 10 percent of that population. Because so few normals endorsed the items in the nonstandard direction, any person who does endorse an item in that direction is revealing an abnormal response. A large number of such abnormal responses call into question the extent to which the test subject complied with the test instructions while completing the MMPI. Persons with T scores on the F scale over 100 may have used a true or false response bias, malingered, or may suffer from a serious psychopathology. T scores between 80 and 99 also suggest a fake bad attempt or a plea for help. Low T scores on the F scale (less than 50) could also indicate a fake good test taking profile. The difference between those test takers with actual psychopathology and malingerers is determined by looking at a few other key scales. The profiles of malingerers will usually display a high Fb scale score and also very high scores on the clinical scales, particularly paranoia and schizophrenia.

The K scale was designed to identify test-taker defensiveness. A defensive response style is a less direct attempt to put forth a good impression than the traditional fake good strategy targeted by the L scale. Because a defensive response style will result in lower scores on many of the clinical scales, it is important to identify test-taker defensiveness. Furthermore, the K scale allows the experimenter to correct for the defensiveness suggested by the K scale score. T scores between 56 and 65 on the K scale
imply a mild defensiveness, and T scores over 65 indicate a fake good test-taking attitude or a majority of false responses. A high score on this scale calls into question the individual’s responses to all other questions, and the results should be considered invalid. Low T scores (less than 40) may indicate a fake bad test-taking attitude or a majority of true responses.

The Infrequency-Back (Fb) is a 40 item scale designed to function as an F scale for the new items in the second half of the MMPI-2 (Butcher et al., 1989). Ninety-five percent of the Fb items are located after item number 299. Interpretation of Fb scale scores is the same as F scale (i.e., T scores greater than 80 may indicate malingering, among other possibilities). Berry et al. (1991) reported that a cutoff score of Fb greater than or equal to 9, when used in tandem with the F and VRIN scales at the same cutoff score, will offer the best results in identifying malingering. Their experiment resulted in a hit rate of 92.5% when the validity scales were used in this manner.

In addition to the seven validity scales on the MMPI-2, researchers have developed other validity scales based on the MMPI-2’s item pool. Arbisi and Ben-Porath (1995) developed the Infrequency-Psychopathology Scale (Fp) as a contrast to both the F and Fb scales. Fp scale items were selected by distinguishing 27 MMPI-2 items that are answered infrequently (less than 20%) by both patients with significant pathology and those in the normative sample. As a point of comparison, the F and Fb scales are based on items that were answered infrequently by a normative sample alone. The Fp scale is used to differentiate overreported protocols from protocols produced by patients with serious psychopathology. High Fp scores suggest an answering pattern that is abnormal for the normative sample as well as for psychiatric patients. Low scores indicate that
symptoms have not been faked (Gass and Luis, 2001). Archer, Handel, Greene, Baer, and Elkins (2001) further investigated the usefulness of the Fp scale and found numerous studies that provided very strong backing for the scale, including their own. Archer et al. found that the best cutoff score was a T greater than 90 for men and a T greater than 80 for women. Their study indicated that the use of Fp with those cut scores would likely also cause Fp to result in better positive predictive power than both F and Fb.

The F-K scale, computed as the simple difference between F and K scale scores, was proposed by Gough (1950) as a malingering index. Although there is no single cutoff score that can be considered standard in all cases, Gough (1950), Wetzler and Marlowe (1990), and Austin (1992) indicated that a positive difference of either greater than 8 or greater than 10 would produce accurate classification of a malingering profile. Any time the F scale raw score is higher than the K scale raw score, the possibility of malingering should be contemplated; greater differences indicate an increased likelihood of malingering.

Gough (1954) developed a Dissimulation scale (Ds) to identify subjects’ malingering based on shared stereotypes of neuroticism and clinical cases. Item changes in the MMPI-2 necessitated a revision of Dissimulation scale. Although the Dissimulation scale-revised (Ds-r2) is missing 11 of the 40 items from the original scale, Leckart (1994) arrived at nearly identical clinical conclusions as compared to the complete 40-item scale. Using a T-score of 70 as the cutoff, Leckart found that 114 of the 120 individuals tested resulted in a profile that would have matched the decision based on the original scale.
In this section previous research related to faking will be reviewed. It is organized beginning with the studies in which the F scale is the most effective validity scale and moving on to studies in which other validity scales were the most effective indicator.

Cassisi and Workman (1992) used a short inventory of the MMPI-2 based on the F, L, and K scales to assess those scales’ ability to distinguish between honest, fake good, and fake bad profiles. The F scale correctly classified 95% of those faking bad. Use of the F-K index was found to have limited value due to an extremely high false positive rate. With the F-K, 55% of honest subjects were incorrectly identified as faking good or faking bad.

Timbrook, Graham, Keiller, and Watts (1993) instructed respondents to take the MMPI-2 as one would normally, and instructed participants to take the test again with the goal of giving the impression of “being a person who has very serious psychological or emotional problems.” The order of the test taking was counterbalanced. Their results indicated that the F scale had the largest effect size when identifying fake bad profiles, particularly when used with the F-K and Ds scales.

Bagby, Buis, and Nicholson (1995) tested the relative effectiveness of the standard validity scales in detecting Fake Good and Fake Bad responding. It was found that the F scale was the strongest predictor in the detection of malingering, whereas the Ds-r2, which differentiates malingerers from normals, added no incremental value in the prediction of faking bad.

Babgy et al. (1997) studied the comparative effectiveness of MMPI-2 validity scales in detecting feigned depression and schizophrenia. The validity scales F, Fb, and
Fp (in that order) were found to best distinguish participants feigning schizophrenia from patients with schizophrenia, whereas F and Fb best distinguished patients with depression from participants feigning depression. F-K failed to help successfully identify participants faking bad for both disorders.

Graham, Watts, and Timbrook (1991) instructed respondents to take the MMPI-2 normally, then were asked to give the impression of “being a person who has very serious psychological or emotional problems” when taking the test a second time. These subjects were found to overreport symptoms, resulting in a very high F scale score, generally elevated clinical scores (most noticeably on Paranoia and Schizophrenia), and lowered K scale scores compared to the profiles of hospitalized psychiatric patients and normal respondents. The F scale properly classified 96% of the faked profiles. The F-K index was only slightly less successful, and the Fb scale correctly identified 88% of the faked profiles.

Similar to Graham et al. (1991), Lim and Butcher (1996) had respondents take the MMPI-2 twice, honest and faked. The participants tended to overreport symptoms, resulting in elevated F and clinical scale scores and lower K scale scores as compared to true psychiatric patients. The F and Fb scales detected a high percentage of fake bad profiles, and the F-K index was nearly as successful as the F scale in discerning between honest and fake bad profiles.

Wetter, Baer, Berry, Smith, and Larsen (1992) randomly assigned participants to one of four groups: (a) random answering, (b) fake a moderate psychological disturbance, (c) fake a severe psychological disturbance, and (d) standard instructions. Faking psychological disturbances (malingering), in both severe and moderate faking groups,
yielded significant elevations on multiple validity scales: F, Fb, and Ds. The significantly lowered K scale results for malingerers, when combined with the elevated F scale, also produced a significantly high score for the F-K scale. These conditions resulted in the effective detection of intentional response misrepresentations. Although those faking a moderate disturbance had scale score elevations to a lesser degree than those faking a severe disturbance, both levels of faking were detected at a significant level.

Bagby, Nicholson, Buis, and Bacchiochi (2000) enlisted 23 experienced mental health professionals with specific expertise in depression diagnosis to complete the MMPI-2 as if suffering from depression. The responses of these professionals were compared to a sample of patients diagnosed with major depression. Results indicated that the validity scales F, Fb, and Ds were successful at distinguishing between bonafide and feigned depression, with the Fb being most successful.

As with Bagby et al. (2000), Bagby et al. (1997) investigated whether clinical training facilitated feigning schizophrenia on the MMPI-2. The results indicated that both undergraduate psychology students and clinically trained persons with advanced training in psychopathology and psychological testing overendorsed items associated with schizophrenia when directed to feign that disorder. Among the naïve participants, the Fp scale provided the largest effect size of any validity indicator, whereas F and Fb were more powerful indicators for clinically trained participants.

Storm and Graham (2000) assigned participants to three groups, uncoached malingerers, coached malingerers, and hospitalized psychiatric patients. Fp produced the most accurate identification in both coached and uncoached situations. The addition of F led to greater predictive power. The further addition of Ds-r2 also added significantly to
the classification. The predictive power of the scales was lower for those participants who received coaching concerning the validity scales.

Austin (1992) investigated the efficacy of the F, K, L, and F-K indices of the MMPI-2 in detecting fake good, fake bad, and honest profiles. Austin found the F, K, and F-K indices were the most successful at classifying fake bad in college students, with success rates of 95%, 95%, and 100%, respectively. These scales were notably less successful in identifying those subjects instructed to fake good, with success rates of 5%, 30%, and 90% individually.

Rogers, Bagby, and Chakraborty (1993) randomly assigned participants to one of four groups: (a) coached on schizophrenia symptoms, (b) coached on strategies for faking detection by the MMPI-2 validity indices, (c) coached on both symptoms and strategies, and (d) uncoached in any way. The profiles of these participants were compared to the profiles of schizophrenic patients. Participants coached on strategies alone had significantly lower scores on the F and Fb scales than those uncoached or coached on symptoms alone and were basically indistinguishable from actual schizophrenic patients. The Ds-r2, however, correctly identified nearly two-thirds of those coached on strategies alone and approximately 80% of all other groups.

The Present Study

Previous studies examined the efficacy of MMPI-2 validity scales at the detection of malingering using only a subset of the total available validity scales. Furthermore, no study could be found in which all of the validity scales were combined and scored in a noncompensatory model, in which the failing of any one scale is interpreted as malingering. Moreover, faking studies of all kinds often induce faking in an unrealistic
manner; that is, test takers are instructed to fake without any mention of the presence of the validity scales. As demonstrated by Rogers et al. (1993), a manipulation such as this presents an overly positive picture of the efficacy of the validity scales. This study will examine the effectiveness of a simultaneous use of all of the malingering oriented validity scales (F, F-K, Fp, and Ds-r2) combined in a noncompensatory model. Test takers will be instructed to either answer honestly or fake a mental disorder. It is hypothesized that the simultaneous use of the validity scales will accurately identify which subjects are malingering.
Method

Participants

Fifty-three undergraduate students at a large southeastern university served as participants in this study. All participation was voluntary and anonymous. Students were offered extra credit in their psychology courses in exchange for participation. In accordance with ethical principles, all students had access to extra credit regardless of participation. Finally, students were reminded that of the opportunity to withdraw from the study at any time without penalty.

Materials

Participants completed the first 370 items and 10 additional items (needed to complete the Fp and Ds-r² scales) from the 567 item MMPI-2. These 370 MMPI-2 items include all items from the original MMPI that were retained for the MMPI-2. The final 197 items on the MMPI-2 were new items written to address new constructs and have no impact on the scoring of any of the standard scales (aside from the 10 items needed to complete the Fp and Ds-r² scales). The use of the first 370 items needed to score the standard scales is the only acceptable short form of the MMPI-2 (Graham, 1993) and was done to save test-taker time. Additionally, the hypothesis of the study was not related to the final 197 items. Fb was not analyzed for two reasons. First, Fb is designed in the same manner as and replicates the purpose of F. Second, Fb is restricted to the latter 197 items of the test and cannot be scored when the short form of the MMPI-2 is used. As such, Fb would not be available to assist in the identification of malingering for practitioners using the short form.
Procedure

Upon arrival at the session, all participants were given a brief explanation of the study and were asked to sign consent forms. Participants were randomly assigned to one of two groups, honest or malingering. Participants in the honest group were given the usual instructions to the MMPI-2 which emphasize honest, open responding. Participants in the malingering group were read a description of schizophrenia followed by instructions to pretend that they suffered from schizophrenia when completing the test. Participants were additionally told that the test contained scales designed to detect lying. The goal was to appear to be suffering from schizophrenia while not being detected as lying by the faking scales. A complete copy of the experimental script can be found in Appendix A.

Design and Statistical Analyses

Responses were scored according to the MMPI-2 scoring guide. A participant was flagged as faking if he or she failed any one of the validity scales; that is, a noncompensatory approach to malingering detection was used.

Data were analyzed in three ways. First, the hypothesis of this study was examined using a 2 (faking or honest) x 2 (flagged as faking or not flagged as faking) chi-square test of association. The second analysis repeated the first analysis with the exclusion of those participants in the honest group who displayed a schizophrenic score profile. Because this study seeks to determine the efficacy of the MMPI-2 validity scales at detecting malingering among mentally healthy test takers, inclusion of people potentially suffering from schizophrenia obscures results. Finally, the malingering group was examined in isolation. In this analysis the percentage of the successful malingerers
who failed at least one of the validity scales was examined. Ideally, 100% of the malingerers should be identified. Given the usual logical constraints of null hypothesis significance testing and the sample size, the percent of subjects in the malingering group identified as malingering in this sample was tested against a population value of 80%.

**Scale Scoring**

All items were scored by Pearson Assessments. The scale scores for each test taker were coded and dichotomized by the researchers in order to determine whether each subject (a) answered in a manner consistent with schizophrenia and (b) faked his or her responses. All cutoff scores chosen were based on previous faking research. In this section, the reasoning behind the cutoff scores chosen for each of the validity scales used in this research project will be reviewed.

For the F scale, raw scores greater than 18 were classified as faking. This cutoff score was chosen based on research by Graham et al. (1991), Lim and Butcher (1996), and a meta-analysis by Rogers, Sewell, and Salekin (1994). The meta-analysis by Rogers et al. reviewed 15 malingering studies, all of which employed both subjects responding honestly and subjects instructed to feign. Based on this review, Rogers et al. concluded that the optimum cutoff score for the F-scale is 16. The study by Lim and Butcher (1996) found that an F-scale cutoff score of 17 correctly categorized 100% of the profiles produced under fake bad and standard conditions. Graham et al. found that a cutoff score of 18 on the F-scale allowed for the best discrimination between fake-bad and control profiles, correctly categorizing 95% of the profiles in the least successful condition.

For the Fp scale, raw scores greater than seven were classified as faking for the male population and raw scores greater than five were classified as faking for the female
population. These cutoff scores were chosen based on research by Archer et al. (2001) and Storm and Graham (2000). The study by Archer et al. was designed to test the utility of Fp scale using profiles of psychiatric patients taking the MMPI-2 under standard instructions and overreported profiles from studies in which participants were charged with simulating serious psychopathology. It was found that the optimal raw cutoff scores, based on the suggested optimal T-scores, was nine for men and five for women. Storm and Graham (2000) found that the optimal raw cutoff scores were five for men and three for women, respectively. However, they further stated that when dealing with coached malingerers, the cutoff scores should be raised in order to classify more accurately. The suggested raw cutoff scores for such a situation were six for men and five for women.

For the F-K scale, raw scores greater than 11 were classified as faking. This cutoff score was chosen based on research by Austin (1992) and Lim and Butcher (1996), and Wetzler and Marlowe (1990). Austin (1992) tested the efficacy of various MMPI-2 scales to detect fake good, fake bad, and honest profiles. It was found that a raw cutoff score of 11 correctly classified 100% of the fake bad respondents. Lim and Butcher (1996) determined that a raw cutoff score of 12 was optimal. With a cutoff of 12, 92% of the faked male profiles and 100% of the faked female profiles were correctly classified in their study. Wetzler and Marlowe (1990) noted that an F-K cutoff score of 11 is reportedly found in less than 1% of normal subjects and 2.5% of inpatients.

For the Ds-r\textsubscript{2}, raw scores greater than 15 were classified as faking. This cutoff score was chosen based on research by Leckart (1994), Rogers et al. (1993), and Storm and Graham (2000). Leckart (1994), in deriving the revised version of the Ds-r, found that a raw cutoff score of 15 (equitable to a T-score of 70) was the most proper to use and
resulted in almost duplicate clinical determinations as use of the full Ds-r scale. Rogers et al. (1993) also used a cutoff score of 15 and found that the Ds-r2 scale allowed for a very high correct classification percentage of respondents coached on symptoms, strategies, and both symptoms and strategies. Moreover, the Ds-r2 correct classification rates were significantly higher than the rates for the F and F-K scales. Storm and Graham (2000) studied MMPI-2 performance in correctly classifying malingerers that were both coached and not coached and used a Ds-r2 scale cutoff score of 15.

For the Schizophrenia scale, persons with T-scores greater than or equal to 75 were classified as having a schizophrenic profile. This cutoff score was chosen based on Graham’s (1993) book on MMPI-2 assessment and the MMPI-2 manual (Butcher et al., 1989). Graham stated that T-scores greater than or equal to 75 on this scale strongly suggest the possibility of a psychotic disorder. The MMPI-2 manual also states that a cutoff score of T greater than or equal to 75 is appropriate for diagnosing a profile as schizophrenic. Both sources note that a person whose profile results in such a score will likely have many of the symptoms of schizophrenia, such as confused and disorganized thinking, hallucinations and/or delusions, impaired contact with reality, and poor judgment.
Results

Fifty-three undergraduate students enrolled in psychology classes participated in this study. One of the students inadvertently skipped 55 of the 380 questions and was removed from the study. Of the remaining 52 students with valid profiles, 16 were male and 36 were female. Age and race data were not collected to increase feelings of anonymity. Participants were randomly assigned to groups yielding 26 people per group.

An examination of the number of test takers classified as schizophrenic by group indicated that all 26 people in the faking group demonstrated schizophrenic profiles whereas five people in the honest group were identified as schizophrenic (see Table 1 for a complete listing). Although it is desirable that all of the test takers in the faking group were able to produce schizophrenic profiles as directed, the presence of five people in the honest group with schizophrenic profiles is troubling.

Table 1

Schizophrenia Classification of Participants by Group

<table>
<thead>
<tr>
<th>Classified as</th>
<th>Not classified as schizophrenic</th>
</tr>
</thead>
<tbody>
<tr>
<td>schizophrenic</td>
<td></td>
</tr>
<tr>
<td>Honest Group</td>
<td>5</td>
</tr>
<tr>
<td>Faking Group</td>
<td>26</td>
</tr>
</tbody>
</table>

The data for these test takers were checked for coding errors. None were found, leading to two possible conclusions: (a) five of the twenty-six college students in the honest sample (i.e., close to 20% of the honest group) were actually suffering from
schizophrenia at the time they completed the test or (b) five test takers in the honest group were confused as to the directions or were not answering in a systematic, honest fashion. Clearly, the first option is unlikely, leading one to conclude the latter option is the case.

A 2 x 2 chi-square test of association was performed to assess whether the participants were correctly identified by the MMPI-2 (i.e., the participants asked to fake bad did not pass the validity scales and the participants asked to answer honestly were able to pass the validity scales). Although five of the participants in the honest group were classified as schizophrenic (Table 1), the faking scales when scored in a noncompensatory manner were able to correctly identify faking test takers to a significant degree, $X^2(1) = 27.53, p < .05$ (see Table 2 below). As discussed in the analysis section, the five honest group participants that exhibited schizophrenic profiles were removed from the analysis, in order to determine the efficacy of the MMPI-2 validity scales using only those participants that were able to provide profiles for which they were instructed. (Participants in the faking group failing to produce a schizophrenic profile would also have been excluded from this second analysis. However, no such cases occurred.) Once these five honest group participants were removed, the results improved, $X^2(1) = 36.12, p < .05$. See Table 3 (below) for specific results.

The final analysis was a one-tailed significance test, based only on the participants in the faking group. This analysis was run to determine if 80% or more of those faked profiles using the validity scales outlined previously could be successfully captured. All 26 (100%) of the participants in the faking group were identified as faking by at least one validity scale, a significant result, $z = 2.77, p < .05$. 
Table 2

*Validity Scale Classification of All Participants by Group*

<table>
<thead>
<tr>
<th></th>
<th>Failed at least one validity scale</th>
<th>Passed all validity scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honest Group</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td>Faking Group</td>
<td>26</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 3

*Validity Scale Classification of Successful Participants by Group*

<table>
<thead>
<tr>
<th></th>
<th>Failed at least one validity scale</th>
<th>Passed all validity scales</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honest Group</td>
<td>3</td>
<td>18</td>
</tr>
<tr>
<td>Faking Group</td>
<td>26</td>
<td>0</td>
</tr>
</tbody>
</table>
Discussion

The results of this study suggest that the use of specific validity scales (F, Fp, Ds-r2, and F-K) in conjunction with one another were effective in detecting feigned schizophrenia among participants who were informed of the basic symptoms of this disorder. The symptoms provided to test takers, when combined with the information shared relating to the validity scales designed to detect response distortion, offered the participants a better opportunity to succeed in feigning a disorder than normal, unaware test takers would have available.

Participants were given basic information about schizophrenia because it was believed that a person serious about feigning a psychiatric disorder successfully would very likely learn some basic information about the disorder intended to be faked. However, it is possible that giving information regarding the symptoms to participants would lead participants to overendorse symptoms thought to be features associated with schizophrenia. It is suggested further research compare groups of noninformed feigners, informed feigners, feigners that have been given validity scale information, and feigners that have been both disorder and validity scale informed would be suggested.

A troubling issue in this study is five subjects within the honest group who displayed a schizophrenic profile of responses. As mentioned, these five participants were either actually suffering from schizophrenia or were confused as to their goals when completing the test. Most puzzling is that participants were run in small, single condition groups. That is, a person in the honest condition was never exposed to the instructions for the faking group. Thus, it is odd that any person in the honest group could be unclear as to the desired test taking protocol.
A different problem may have occurred in the faking group. The subjects instructed to produce a schizophrenic profile had no intrinsic or extrinsic motivation to feign the requested disorder in a manner that would not be caught by the validity scales, and their commitment to the task may have been less than full. Obviously, the data demonstrate that all members of the faking group displayed a schizophrenic profile. Thus, these participants were motivated enough to produce a schizophrenic profile. What is unknown is whether their motivation to avoid identification by the validity scales was strong enough to avoid detection. Offering participants the opportunity for bonus remuneration as reward for a successful profile, or the most successful profile if there are zero successful profiles, would be an alternative tactic to increase motivation.

One possible concern not addressed in this study, due to the available subject population, is the accuracy of the selected validity scales used in this manner when a portion of the subjects are individuals previously diagnosed with schizophrenia. The main point of interest would be to determine whether people suffering from schizophrenia but answering honestly would fail one of the validity scales and be incorrectly identified as faking. It is entirely possible that the high rates of accuracy in this study will not be found when test takers are actually suffering from schizophrenia.

The use of a larger sample, more representative of the national population is another potential improvement to the study. It remains to be determined whether participants drawn from non-student populations would be more or less successful in feigning schizophrenia than the student sample used in this study. It is also not known whether age or education correlates with the ability to feign more successfully. It could be argued that broader life experiences might correlate with a higher feigning
proficiency. The sample size would need to be much larger than in the current study, both to meet the representation suggested and to allow for stronger predictive power. This would also allow the tactics and results to be generalized more broadly.

The data in the current study were analyzed in a noncompensatory model, in which failing any of the validity scales was interpreted as malingering. This format shows that the scales, used in conjunction with one another, appear to function very effectively in the detection of malingering. In looking to streamline this procedure and produce the most simple and accurate predictive possibilities, future research should be considered to see how few validity scales are necessary to obtain the same predictive accuracy. Past research has resulted in the determination of various different individual validity scales being considered the most powerful predictor of faking bad, always in a compensatory model if more than one were analyzed. The new study could be an exploratory analysis based on both the knowledge gained from the past research regarding the individual validity scales with the design of the present study, to find the combination of scales that most accurately predict malingering.

In sum, the MMPI-2 contains a number of validity scales for the detection of malingering, which when combined in a noncompensatory model are highly effective at accurately identifying subjects faking schizophrenia. The results of this study suggest that clinicians would find it fruitful to consider using groups of validity scales together in a noncompensatory model in contexts where detecting malingering profiles is a concern.
References


Psychological Assessment, 9, 106-112.


Appendix A:

Experimental Script
Honest Group Instructions

Welcome to our study on personality testing.

Pass out and read informed sign consent. Emphasize that participation is voluntary and anonymous.

Are there any questions about the study?

Collect informed consents and pass out tests and answer sheets. Be sure answer sheets are coded appropriately for group - first digit of ID should be 1 (for honest group).

For this study you will be taking a personality test called the MMPI. Please open the test booklet and read the instructions on the first page.

Read instructions out loud.
Then say:

Please don’t write on the test booklet. On the answer sheet, fill in sex and no other information. Remember this study is anonymous – we don’t want you to write your name or any other identifying information anywhere. When taking the test, please answer honestly.
Open the answer sheet. We’ll be answering the first 371 items plus nine more. When you answer number 371, find the next highlighted item and answer it. Repeat until you are finished. Be sure to answer in the right spot on your answer sheet. Does everyone see the highlighted items after number 371?
Are there any questions?

When people finish, check answer sheet to see if they answered the first 371 plus the highlighted items. Offer to answer any questions about the study (do so outside of the room to avoid disturbing the other test takers).
Faking Group Instructions

Welcome to our study on personality testing.

Pass out and read informed sign consent. Emphasize that participation is voluntary and anonymous.

Are there any questions about the study?

Collect informed consents and pass out tests and answer sheets. Be sure answer sheets are coded appropriately for group - first digit of ID should be 2 (for faking group).

For this study you will be taking a personality test called the MMPI. Please open the test booklet and read the instructions on the first page.

Read instructions out loud.

When answering the questions, we want you to answer the questions as if you are suffering from the mental disorder schizophrenia. Schizophrenia is characterized by disturbances of thinking, mood, and behavior.
Delusions and hallucinations may be present. Behavior may be withdrawn, aggressive, or bizarre.
Your goal is to be diagnosed as suffering from schizophrenia and NOT be caught by the lying, or faking, detection system built into the test. This faking detection system is designed to identify people who are lying when taking the test. To summarize, you should answer in a way that makes it appear you are schizophrenic without getting caught. Are there any questions?

Put transparency up and leave up during experiment (don’t dim lights).

Please don’t write on the test booklet. On the answer sheet, fill in sex and no other information. Remember this study is anonymous – we don’t want you to write your name or other identifying information anywhere. Please open the answer sheet. We’ll be answering the first 371 items plus nine more. After you answer number 371, find the next highlighted item and answer it. Repeat until you are finished. Be sure to answer in the right spot on your answer sheet. Does everyone see the highlighted items after number 371?

When people finish, check answer sheet to see if they answered the first 371 plus the highlighted items. Offer to answer any questions about the study (do so outside of the room to avoid disturbing the other test takers).