Canadian Journal of Behavioural Science
2009, Vol. 41, No. 3, 121–132

0008-400X/09/$12.00 DOI: 10.1037/a0015589

Conventional and Actuarial Methods To Detect Response Distortion on the Basic Personality Inventory

Edward Helmes
James Cook University

Response distortion remains a significant issue in the assessment of psychopathology. Here four groups of psychiatric patients, each of 40 people, were asked to respond honestly or to distort their presentation as either worse, better, or a "normal" pattern of responses to the Basic Personality Inventory (BPI; Jackson, 1989); only those cases showing acceptable consistency in responding ("person reliability") were analysed. Performance of the conventional cutting points on the BPI validity scales were compared with results from linear discriminant analysis calculated for the patients and from those variables selected previously for university students by Helmes and Holden (1986). Preliminary analyses showed that the "good" and "normal" groups could not be separated; the normal group was therefore not included in subsequent analyses. Results showed better classification results for discriminant functions than for the use of the standard BPI validity measures. Contrary to findings using multiple regression, heuristic weights were the least accurate. Implications for applications of the BPI involving suspected response distortion are discussed, along with the unexpected finding of poor performance of heuristic classification functions.

Keywords: faking, random responding, Basic Personality Inventory, discriminant analysis, person reliability

Methods of detecting distorted response patterns on multiscale measures of psychopathology have a long history, dating to the early development of the *F*, *L*, and *K* scales of the Minnesota Multiphasic Personality Inventory (MMPI; Greene, 2000; Hathaway & McKinley, 1976). Methods based upon these scales in both the MMPI and MMPI-2 have generally worked well when applied appropriately in both simulation studies and in actual practise to detect attempts to over- or underreport levels of pathology (see Bagby, Marshall, Bury, Bacchiochi, and Miller (2006) and Wetter, Baer, Berry, and Smith (1992) for reviews of the MMPI/MMPI-2 and dissimulation).

In applied settings where such tests are used, the stakes of taking a measure of psychopathology are often high and there thusly may be significant external incentives to distort responses. Such contexts for presenting a distorted, inaccurate set of responses include the insanity defence for serious criminal charges (prompting overreporting) or seeking a highly desired job (prompting underreporting). More common situations in most clinical practises are likely those of distressed individuals who overemphasize their distress to gain therapist sympathy and those who seek to minimise real distress out of a fear of appearing "crazy" or to preserve a positive image of themselves.

There is thusly value in examining the performance of measures of psychopathology that have indices to detect response distortion to determine how well those indices perform with different samples. For measures of psychopathology, samples of people with genuine mental health problems provide a more realistic context for evaluating measures of response distortion than, for example, samples of university undergraduates, while forensic samples may present a quite different context again. Student samples are generally more highly educated and arguably subject to fewer psychosocial sources of stress than most samples from either mental health or forensic settings. The latter two populations may well perform differently from students if asked to distort their responses. The data presented here are from people with psychiatric problems and so represent a less common sample in the literature on response distortion. The study also addresses the question of how well a multivariate method for differentiating groups, discriminant analysis, performs in differentiating groups of patients who had been instructed to distort their responses to a measure of psychopathology. In most such simulation studies, experimenters have generally assumed that participants follow instructions to respond in a distorted manner from their true nature. If they do so approximately equally well, then discriminant analysis should recover the groups with a high degree of accuracy. Rogers,

This work was supported in part by funding from the Department of Psychiatry, University of Western Ontario. I thank Dr. Maryann Fraboni, Dr. Ronald R. Holden, and Dr. Douglas Langbehn and the anonymous reviewers for comments on earlier drafts. Portions of this work were presented at the 1990 Annual Meeting of the American Psychological Association. Additional analyses were conducted for this revision. Given the delay between data collection and this report, a reviewer of a draft pointed out that Arbisi and Ben-Porath (1995) noted changes in endorsement rates over time for responses to the Minnesota Multiphasic Personality Inventory (MMPI). The time span here is notably less than half the 50-year period for the MMPI and the developmental samples for the Basic Personality Inventory (BPI) and for the sample in this study should be quite similar.

Correspondence concerning this article should be addressed to Edward Helmes, Department of Psychology, James Cook University, James Cook Drive, Townsville, Qld 4811. E-mail: edward.helmes@jcu.edu.au

Sewell, Morey, and Ustad (1996) reported that discriminant analysis was superior to traditional validity measures in correctly classifying groups of students instructed to fake their responses. Whether this assumption holds with a mental health sample is evaluated here.

Response distortion may take different forms. Rogers (1997) summarises the various terms that have been applied to different patterns of response distortion. He discusses malingering, defensiveness, irrelevant responding, random responding, and hybrid responding as dissimulation. Taking a different approach, Nichols, Greene, and Schmolck (1989) classified random responding as one form of content nonresponsiveness (CNR), in which respondents do not respond in terms of the overt item content, which would also include Rogers' irrelevant responding. Nichols et al. labelled the primary form of response distortion as content-related faking (CRF), in which respondents do respond to the item content, but try to distort their image in either a positive or negative direction. This distinction between faking "good" and faking "bad" forms the basis for one of the more common manipulations in experimental studies of response distortion.

Common validity scales on multiscale inventories of psychopathology have generally been designed to detect both negative and positive dissimulation response styles. In addition to scales to detect dissimulation that has already occurred, efforts to limit attempts to distort responses through the test instructions have largely been warnings that the test has measures designed to detect attempts at presenting a deceptive set of responses (Eysenck, Eysenck, & Shaw, 1974; Goffin & Woods, 1995; Kluger & Colella, 1993). Such warnings may result in scores that can be regarded as "normal" profiles (e.g., Merbaum, 1972), as opposed to those that constitute the "faked good" test profiles that are more commonly studied. Presumably, normal profiles include admissions of some faults that are denied in attempts to produce overly positive ones, but there are few empirical studies that attempt to determine whether such distinctions can be supported in practise. This issue is one that is addressed in this study.

Generalization of Actuarial Methods

Given that a measure of psychopathology has a set of measures to detect dissimulation, how test users adopt those measures constitutes another issue. They may rely on the published recommended cutoff scores for the various measures to detect faking on the test or make use of more sophisticated multivariate procedures such as discriminant or profile analysis. A similar alternative is the use of decision algorithms, as used for detection of defensive responding on the PAI (Morey, 1991). Within the collection of actuarial methods, one question examined here is to what extent the use of discriminant analysis to differentiate faked profiles from honest ones can generalise to new samples. Discriminant analysis provides weights that can be consistently applied to sets of predictor variables where the goal is the allocation of new individuals to known groups. If the weights calculated in one study generalise to other samples, then such diagnostic classification (or the separation of those with valid (honest) test profiles from those with invalid (faked) ones) can be accomplished without new studies to derive new weights for every application of that test. When used in the explication of validity indicators, the explicit algorithms and fixed decision rules of discriminant analysis are likely to result in more consistently accurate predictions than simple clinical judgement using the same individual validity scales (Dana & Dawes, 2004; Dawes, Faust, & Meehl, 1993). For example, some of the interpretive guidelines for the PAI make this use of discriminant analysis, including the classification of profiles as faked (Morey & Lanier, 1998), as do some efforts to use intelligence test scores to detect deliberate attempts to appear impaired on these measures (Babikian & Boone, 2007).

Discriminant analysis is one multivariate approach to the classification of individuals, along with logistic and probit analyses (Wilkinson, Blank, & Gruber, 1996). One recognised weakness of multivariate approaches such as least squares regression and discriminant analysis is that weights may not generalise well across different samples, with the number and weighting of predictors and sample sise being relevant factors (Bernstein, 1988). Wainer (1976) suggested the use of unit weighting of predictors in multiple regression could lead to better generalization to new samples as an alternative to optimal calculated weights. This suggestion has often been supported in practise. For example, Raju, Bilgic, Edwards, and Fleer (1999) found that unit weights cross-validated better than ordinary least squares solutions in multiple regression in their large sample of air force recruits.

The application of unit weights to discriminant analysis is, however, much less common than its application in multiple regression and there is a much less extensive literature on the topic. Bernstein (1988) notes that discriminant analysis and multiple regression are equivalent only when there are two groups. In such cases, a dichotomous criterion is used to identify the two levels of group membership in both approaches, but the computational methods used to calculate the discriminant weights are different from those used in multiple regression, and normally the goals are different as well. In the case of regression, the computational goal is to maximise the multiple correlation coefficient, while in discriminant analysis the goal is to maximise the differences amongst the groups. Therefore, it becomes an empirical issue as well to determine whether the literature on unit weights derived from multiple regression is indeed applicable to applications involving discriminant analysis when more than two groups are involved. This issue is also addressed in this study, which is based upon related research that was completed using the same measure of psychopathology in a sample of university students (Helmes & Holden, 1986).

The Current Study

The measure of psychopathology used is the Basic Personality Inventory (BPI; Jackson, 1989), a comparatively new measure of psychopathology that has been used with psychiatric patients (McNeil & Reddon, 2001) and with both juvenile and older offenders (Jaffe, 1985; Kroner, Forth, & Mills, 2005; Kroner, Holden, & Reddon, 1997). Measures to detect response manipulation on the BPI include the Denial scale, which assesses the avoidance of personal problems and lack of insight into personal actions, and the Deviation scale, which is a collection of critical items reflecting severe psychological prob-

lems. Other validity measures on the BPI include a measure of social desirability, an index of consistency in the profile (person reliability), and counts of the number of true responses and of perseverative responses, in which the same response (T or F) is repeated in a string (Helmes & Holden, 1986; Holden, Helmes, Fekken, & Jackson, 1985).

There is a comparatively small literature on faking with the BPI compared with the equivalent literature with the older MMPI/MMPI-2. Holden and Jackson (1985) provide the first report of the use of the BPI in studies of faking. That article focused more upon the response style of social desirability and the nature of test-item content than upon profiles of distorted responses. That issue was addressed by Helmes and Holden (1986), which was the first report to provide information on distorted responses to the BPI. Bagby, Gillis, and Dickens (1990) reported that the BPI was better at detecting faking good than the Millon Clinical Multiaxial Inventory-II (MCMI-II; Millon, 1987), which in turn was better at detecting faking bad in this sample of university undergraduates. A later study by Steffan, Kroner, and Morgan (2007) used prison inmates instructed to fake good (FG) on the BPI or to fake bad in the form of one of three different forms of psychopathology. In contrast to the findings of Bagby et al. (1990), Steffan et al. reported that the BPI was better at detecting faking bad than faking good. None of these studies have investigated the question of whether the BPI has different profiles for responses that are faked "good" versus those that are faked "normal." That distinction is much less salient than the case of "good" versus "bad" that forms the basis for the majority of research on faking. Given the use of the BPI with adult and juvenile offenders who may be motivated to distort responses to the BPI in order to better their condition, there are grounds for investigating whether a distinct profile on the BPI exists for faking normal.

The purpose of this article is therefore threefold. The first goal is to determine whether patterns of scores on the BPI from psychiatric patients can be used to differentiate the distortion of responses for faking good from those for faking normal. This bears upon the issue of whether warnings to those taking a test with measures to detect distortion are likely to be effective in that individuals who are warned may moderate their attempts to present a positive image, leading to a pattern of scores closer to that of honest responding than of faking good. The second goal is to derive a new set of discriminant weights from a sample of psychiatric patients to contrast with those already developed for the BPI (Helmes & Holden, 1986), which were based upon university undergraduates who were instructed to feign their responses to the BPI. The second goal also includes a comparison of the success of the discriminant function weights with performance of the existing cutting points for the individual BPI validity measures reported by Helmes and Holden (1986). A third goal is to contrast the optimal discriminant weights with sets of simpler heuristic weights to determine their comparative accuracy in classification in a context which has a low level of external incentives for distortion, routine intake screening. Such a context is likely more typical of most clinical practises, in which the forensic and personal injury litigation cases that are most likely to involve deliberate and premeditated distortion are a small fraction of the cases. Incentives to under- or overreport psychopathology in such contexts are more likely therefore to arise from internal factors of the clients. Therefore the lack of strong incentives in this study to distort responses can provide information that may generalise more to clinical applications than to more forensic ones.

Method

Participants

The derivation sample comprised 160 Canadian psychiatric patients (55.6% male), with a mean age of 33.7 years (SD = 10.8) who were all inpatients of a tertiary psychiatric hospital. Participants were randomly assigned to a group instructed to appear to be worse than they actually were (fake bad, FB), to appear better than they actually were (FG), or to appear to be normal (fake normal, FN), while a fourth group took the BPI under standard instructions (straight take; ST). Three other persons began the task (two from the FB group and one from the FN group), but did not complete it, either failing to understand instructions to distort responses or being uncomfortable with the prospect of being dishonest and faking. All participants in the three distortion instruction groups were volunteers from the admitting wards of a psychiatric hospital that served a mixed urban/rural region, and were informed of the purposes of the study. All had consented to participate and were paid an amount of four dollars to compensate them for their time. No other compensation or other forms of external incentives were provided. The ST group comprised other individuals who had undertaken consecutive routine assessments using the BPI on the same admitting wards of the same hospital over the same period of time. None of this group refused consent for their results to be used and no court remand cases or others who might have other reasons to distort their responses in one direction or another were included.

The project was reviewed and approved by the University's Health Sciences Standing Committee on Human Research. A power calculation using the program by Pittenger (2001) gave a power of .75 to detect an effect size of .25 with a Type I error rate of .05 for four groups with 40 participants per group.

The overall sample was broadly representative of the adult patients of the hospital in terms of demographic characteristics and diagnostic groups. Most (52.5%) were single, with 17.5% presently married, 28.1% married in the past and now either separated or divorced, and 1.8% were widowed. The sample had a mean of 11.4 years of education (SD = 2.44 years), with a mean of 5.2 previous admissions (range 0 to 37, SD = 6.04). The largest diagnostic group was those with schizophrenia (50%), followed by those with personality disorders (18.1%), affective disorders (11.3%), alcohol and drug problems (7.5%), neuroses and adjustment and conduct disorders (4.4% each), and other psychoses (2.5%) using either the criteria of *Diagnostic and Statistical Manual of Mental Disorders, Third Edition* (*DSM–III*; American Psychiatric Association, 1980) or the *International Classification of Diseases, Ninth revision* (World Health Organization, 1977) for the hospital diagnosis at discharge from the index admission. The sample also included a small proportion of individuals with sexual disturbances and unclassified depressive disorders.

In terms of symptomatic history, 47.5% of the sample had present or past symptoms of depression, and 35.6% had attempted suicide. Over one third (34.4%) had some type of criminal charge against them in their past, and 21.3% had assaulted another person. Alcohol and drug abuse were fairly common: 34.4% had a history

of prescription or street drug abuse, while 36.3% had a history of alcohol abuse.

Procedure

The instructions to participants in the distortion groups were as follows, with the full set of instructions for *faking bad* being:

> Assume that you are in a situation where it would benefit you greatly to actually appear mentally disturbed on this questionnaire. As you read the items on the following pages of this test booklet, respond so that you present yourself as someone with serious psychological problems. In other words, try to *fake* this test so that the results will show that you are worse than you really are. Although you may feel that you would never represent yourself dishonestly, please try to do so for this study. However, *beware* that the inventory has certain features (which you want to avoid) designed to detect "faking." Do your best to fake out the inventory.

For those in the FG group, the phrases "very well-adjusted," "someone without any psychological problems or personality faults," and "better than you really are" were used in the first three sentences of the above paragraph. For the those in the FN group, the phrases were as follows: "perfectly normal," "someone with no serious psychological problems," and "quite normal, with some weaknesses, but no major faults or problems." These instructions for faking "good" and "normal" are typical of those normally provided as cautionary warnings with test applications intended to reduce response distortion (Eysenck et al., 1974; Goffin & Woods, 1995; Kluger & Colella, 1993). The intent of the wording used is to maintain the overall similarity to standard instructions for the test to the extent possible with comparatively modest variations to differentiate the three sets of instructions to distort responding.

Measures

The BPI is a 240-item self-report measure of psychopathology. It was developed using modern principles of scale development that emphasised construct and divergent validity and the suppression of response styles to measure 11 broad facets of psychological disturbance, with a 12th critical item scale (Holden, 2000; Holden & Jackson, 1992; Jackson, 1989). It has norms for adults and adolescents for dimensions that are related to antisocial behaviour, impulsivity, emotional disturbance, and psychotic tendencies. The BPI was scored for the standard 12 scales, and also for measures of social desirability (the "balanced" desirability scale), subject reliability (Holden et al., 1985), number of "True" responses, and number of perseverations (Helmes & Holden, 1986). The cutoff scores reported for the BPI validity scales in that study and used again here were .15 and 0 for the reliability index, 150 for a high and 65 for a low number of perseverative responses, 19 or over for high desirability and below 8 for low desirability, and under 20 "True" responses or over 180. Kroner and Reddon (1994) reported the validity indices of the BPI to be not correlated with the BPI content measures sufficiently to permit multivariate analysis without causing logical problems or violation of statistical assumptions.

The perseveration score is the number of repetitions of the previous response. For example, the pattern "TTTFF" would lead to a perseveration score of 3. The person reliability measure was calculated by first splitting the 20 items of each BPI scale into four subscales of 5 items each. The four subscales then formed two pairs of subscales for each BPI scale. The person reliability index was thusly a correlation calculated across the resulting 24 pairs of subscale scores.

Analysis

The first analysis was intended to refine the sample to exclude participants who showed signs of not following instructions. Consistency in responding to items throughout the test in the form of substantial correlation amongst parts of the test would appear to be a minimal criterion for successful dissimulation. The person reliability measure is intended to identify those with low correlations across the item content domains that may be caused by changes in response strategy over the length of the test. The criterion used was that the person reliability index had to be above the higher criterion of 0.15 used by Helmes and Holden (1986). This value has a probability of about .5 of being statistically significantly greater than 0 for a correlation with 24 pairs of scores.

To determine any differences between the FG and FN groups, a series of independent sample *t* tests was calculated to compare the scores on all 12 BPI scales and the 4 validity measures between the FG and FN groups. Cross-tabulations were used to calculate the number of cases in each group that exceeded the published cutoff scores for the BPI validity indices (Single Cutoff Scores in the tables) and for the calculation of sensitivity and specificity for predictions of faking group membership using these cutoff values. Sensitivity is the probability of criterion group membership being correctly identified using the given cutoff score. Specificity is the corresponding probability of correctly not belonging to that group being identified. Positive predictive power is the probability that a person in the defined group is correctly identified, while negative predictive power is the probability that a person not in the defined group is correctly identified by the test (Kessel & Zimmerman, 1993). The ST group was used as the comparison group in the latter calculations. Conventional least squares discriminant function analyses and other analyses were all calculated using SYSTAT 11 (SPSS, Inc., 2000).

A series of discriminant analyses were carried out to determine first, the extent to which discriminant analysis improved upon use of the conventional individual validity measures of the BPI and second, what level of classification was achieved using the optimal set of variables for differentiating those groups. A final set of classification functions was calculated to test how well simplified heuristic discriminant weights predicted group membership.

The variables used for the first discriminant analysis of all available measures included the basic 12 scales of the BPI, plus the four supplementary individual validity indices of social desirability, number of true responses, respondent reliability, and the number of perseverative responses (All Predictors in Tables 2 and 3). The second discriminant analysis (Optimal Patient Predictors) was repeated using only the variables from the set of 16 variables with the highest, or "best" standardised discriminant weights for discriminating the three groups as determined by values of 2.0 or better for the "*F* to remove" statistic (Optimal Patient Predictors in Tables 2 and 3). This value has a probability of being greater than 0 of about .10 in this analysis.

A third set of discriminant functions was derived based on the measures reported by Helmes and Holden (1986; Optimal Student

Predictors in Tables 2 and 3) for a group of university students to determine how well these variables generalised to the patient data. The variables used from the student sample of Helmes and Holden (1986) were desirability, number of perseverative responses, subject reliability, Denial, and Deviation.

The classification weights for the two sets of discriminant functions using the optimal sets of predictors were used to derive two additional sets of heuristic weights (Langbehn & Woolson, 1997) to determine how well simplified weights would classify cases. If the literature on multiple regression was fully applicable, then the heuristic weights should generalise better than the exact weights calculated for the sets of optimal variables. These heuristic functions were formed by rounding discriminant weights to values of 0, +/−1, with other weights formed by truncating decimal places for weights greater than +/−1 (Heuristic Patient Predictors). Preliminary analyses using the patient sample and different values of heuristic weights had tested a variety of simple unit weights (such as using only values of +1, −1, and 0) together with various types of simplified weights. Simplified heuristic weights were calculated for the student discriminant functions in the same manner and are reported as Heuristic Student Predictors in Tables 2 and 3.

Results

Table 1 reports the means and standard deviations for the four groups on the various measures of the BPI, as well as basic demographic information. The scores on the four individual validity measures are reported in the last four rows of Table 1. There was a striking similarity in scores on all measures for the groups instructed to FG and to FN.

Comparison of FN and FG Groups

The first step in the analyses was therefore to contrast the FN and FG groups using *t* tests. None of the 16 measures analysed showed any appreciable difference between the FG and FN groups. The distributions of scores in these two groups were also examined visually using density plots and box-and-whisker displays and were all found to be highly similar. As can be seen in the first and third columns of Table 1, the proximity of the FG and FN groups in terms of scores on virtually all the measures used is quite apparent. With no evidence that the patients discriminated between underreporting pathology and underreporting pathology while still admitting some problems, the FN group was dropped from further analyses. An alternative course would have been to merge the two groups to produce a larger sample, but given that the members of this group had been given a different set of instructions, it appeared that a more conservative approach not to analyse their data any further was more appropriate. The decision to drop the FN group was also based upon the larger number of studies using FG instructions than using some form of FN instruction. The next step was to evaluate the sample for consistency in responding.

Consistency in Responding

Individuals who adopt different strategies at different points during a test may change patterns of responding during the course of completing a set of items. Retention of only those with scores on the person reliability measure of 0.15 and higher left 32 people in the FG group, 34 in the FB group, and 35 in both the FN and ST groups. The five cases in the ST group that had person reliability values below 0.15 were also removed for consistency with the other groups, leaving a total of 101 cases in the 3 remaining

Table 1
Comparison of Basic Personality Inventory (BPI) Measures and Demographic Variables for the Four Groups of Instructions to Psychiatric Patients

	Fake good (n = 32)		Fake bad (n = 34)		Fake normal (n = 35)		Straight take (n = 35)	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	34.8	12.88	32.4	8.44	34.8	12.07	34.7	11.23
Education	11.7	2.86	11.1	2.19	11.5	2.22	11.2	2.38
Hypochondriasis	5.3	4.41	12.7	5.46	4.6	4.37	5.8*	3.38
Depression	4.8	4.84	14.0	5.82	5.7	5.62	7.9*	4.84
Denial	10.7	4.04	7.1	3.48	10.3	3.97	6.2*	2.90
Alienation	6.8	4.47	14.2	4.26	6.5	3.12	8.5*	3.50
Interpersonal Problems	5.7	4.53	13.6	4.22	5.0	4.07	4.7*	2.92
Anxiety	6.0	4.17	13.8	4.88	5.8	4.53	7.9*	3.85
Thought disorder	6.3	4.51	13.4	4.98	6.3	4.42	8.8*	4.41
Persecutory ideas	4.3	4.86	12.5	5.83	4.3	4.29	4.3*	3.21
Impulse expression	6.2	4.22	13.4	4.38	5.6	3.80	7.4*	2.67
Social introversion	4.6	3.89	12.6	4.96	6.4	4.49	7.3*	4.28
Self-depreciation	4.5	5.19	12.6	5.85	4.3	5.18	3.9*	3.50
Deviation	4.7	4.48	14.4	5.40	4.4	5.17	5.3*	2.86
Reliability	.55	.23	.56	.20	.49	.23	.41	.19
Desirability	15.6	4.75	6.9	5.36	15.3	4.59	15.5*	2.68
Number "True"	117.9	21.11	135.8	23.83	117.0	24.67	115.2*	16.95
Number of perseverations	107.3	33.50	111.9	30.85	102.5	30.03	112.5	18.27

* $p < .05$ at Bonferroni correction for 16 tests. Actual $p < .003125$.

groups. The deleted cases and the remaining ones were compared on demographic and clinical history variables, with a Bonferroni correction for the multiple tests. There were no significant differences on these measures between those cases that were deleted and those retained. Therefore the reduced sample was used in the subsequent evaluation of faking.

Detection of Faking the BPI

The next step was to examine the standard BPI validity measures to determine how well the conventional cutting points used with these measures differentiated the three groups. Nineteen of the 32 patients in the FB group scored below the lower cutoff score of below 8 for the desirability measure, as did three of the FG group, while no one in the ST group did so. In contrast, no one in the FB group scored above the high cutoff score for desirability of 19 or 20 out of 20, while 11 of those in the FG group and 5 in the ST group did so. No patient in any group scored below the low cutoff for the number of true responses measure of less than 20 "True" responses, while only 2 patients in the FB group did so for the high cutoff point of over 180 "True" responses out of 240 possible. For the measure of the number of perseverative responses, no patients in any group scored below the low cutoff score of under 65 repeated responses, while 2 in each of the FB and FG groups scored above the high cutoff of 150 such responses. These results are aggregated in the first row of Table 2 to summarise the collective performance of the 4 validity measures, labelled as Single Cutoff Scores.

Table 2 also reports the correct classification rates for the three groups for the various sets of classification functions arising from the discriminant analyses. Both functions were significant (Pillai's trace = 1.02, approximate $F(32, 168) = 5.51$, $p < .001$, canonical correlations = .79 and .64). The first discriminant analysis (Set II) had substantial weights for Alienation, number of perseverations, and Deviation for the first discriminant function. The second function had substantial weights for Depression, Denial, Alienation, Persecutory Ideas, Social Introversion, Self-Depreciation, and Social Desirability. The overall rate of correct classification was 79.2%, as summarised in the second row of Table 2. The lowest rate of correct classification was for the FG group, with 22 out of 32 people correctly classified. The FB group had 26 correct and the ST Group 32 correct. The majority of the FG errors were classed as ST.

A reduced set of measures was evaluated and also had both functions statistically significant (Pillai's trace = .89, approximate $F(10, 190) = 15.28$, $p < .001$, canonical correlations = .75 and .57). This function was based upon using the five variables from the above analysis of the three patient groups that had values of F to remove of 2.0 or more (Optimal Patient Predictors): Denial, Alienation, Social Introversion, Self-Depreciation, and number of perseverations. Note that only two of these variables (Denial and perseverations) are the same as those found in the student group of Helmes and Holden (1986). As with the analysis using the complete set of variables, this set of variables could correctly classify 75.2% of cases into the three groups. The classification errors were equivalent across all groups, with slightly more members of the FG group being classed as ST.

The third discriminant analysis used the five best variables from the Function 2 set reported for the undergraduate students reported by Helmes and Holden (1986) as applied with the current sample (Optimal Student Predictors). This analysis was to determine the extent to which the selection of variables from the previous study with students would generalise to the patient data and again had both functions significant; Pillai's trace = .83, approximate $F(10, 190) = 13.33$, $p < .001$, canonical correlations = .74 and .52. The overall classification rate was close in accuracy to the set of optimal predictors derived directly from the patient data (74.3% vs. 75.2%). The ST group had the most cases successfully classified (31 cases), but classification of the other two groups was poorer, with only about half the FB group correctly classified (17), and 27 of the FG group correctly classified. Once again, the most frequent classification error was of 11 FG cases classed as ST.

Heuristic Discriminant Functions

The initial efforts to form simplified heuristic functions attempted a simple unit weighting model of +1, −1 and 0 based upon the relevant discriminant classification functions, similar to that used successfully in multiple regression. It led to values of classification functions that were identical across cases, leading to no predictions for membership of some groups and all cases from some groups being classed into others. After several rounds of trial and error modifications, heuristic weights were devised that were similar in general configuration of weights to the original best weights for the five variables, but which did not involve decimal calculations. The new simplified heuristic function based on the present patient sample (Heuristic Patient Predictors) weighted only the Denial and Alienation scales on the function for underreporting, as did the function for overreporting, while the simplified heuristic function for the ST group weighted the Depression, Denial, and Self-Depreciation scales. The value of the constant

Table 2
Classification Results of Discriminant Analyses and Faking Measures

Variable	FG (% correct)	FB (% correct)	ST (% correct)	Most common classification	Total % correct classification
Single cutoff scores	34.4	61.8	88.7	16 FG as ST	61.4
All predictors	68.8	76.4	91.4	7 FG as ST	79.2
Optimal patient predictors	62.5	79.4	82.9	8 FG as ST	75.2
Optimal student predictors	53.1	79.4	88.6	11 FG as ST	74.3
Heuristic patient predictors	97.1	10.3	24.2	19 ST as FB	43.6
Heuristic student predictors	100	0	5.7	34 FB as FG	5.0

Note. FG = fake good; FB = fake bad; ST = straight take.

differed for these two response distortion functions (see the Appendix for the various classification functions that were used).

Forming heuristic weights for the student data from Helmes and Holden (1986; Heuristic Student Predictors) also required several preliminary analyses because of the wide discrepancy in the means of the various measures in the function. Results of using this simplified function with heuristic weights for the variables selected for students gave generally poor results for classification (see Table 2). Whereas the functions based on all variables or the sets of optimal predictors were correctly classifying around 70% to 80% of the patients, both the heuristic weight functions classified less than 50% of the cases correctly, with most of the correct classifications being of the FG group. The set of heuristic weights from the patient data classed most of the ST group as FG. The heuristic weights for the function based on students classed all but two of the ST group members as FG and all the FB cases as ST. The heuristic functions for the variables selected from the student sample were highly inaccurate overall.

Table 3 reports sensitivity, specificity, positive predictive power, negative predictive power and overall correct classification rates of faking good and faking bad for the various measures and discriminant functions reported. In order to calculate these figures, actual group membership in the ST group and either the faking bad or the faking good group was cross-tabulated with the corresponding predicted group membership for the set of individual measures as well as for each of the five discriminant analyses. Note that because these results are derived from the 2×2 tables in which the FG or FB group is contrasted with the ST group, the figures in Table 3 differ from those reported previously in Table 2. Table 3 also better illustrates the patterns of performance by the various sets of predictors. The most accurate model overall was the one based upon the current sample of patients and using all available BPI measures (All Predictors). These predictors gave an overall accuracy of just over 80%, with moderate values for all indices. The set of optimal predictors (Optimal Patient Predictors) from the present patient sample were also quite accurate, with better prediction of faking bad than of faking good. The best predictors from the study of students requested to fake responses (Optimal Student Predictors) had values for indices and overall accuracy very close to those for the best predictors for patients, again with predictions of FB better than for FG.

The two sets of heuristic functions were the least accurate in classification. The functions derived from the present patient sample (Heuristic Patient Predictors) was somewhat more accurate than chance, the opposite was true for the functions derived from the student sample, which had poor accuracy overall, especially for faking bad. For the heuristic functions for patients, the same pattern was seen in that the classification accuracy for faking bad was greater than for faking good; only faking bad had reasonable levels of predictive accuracy for the student heuristic functions.

Discussion

The results clearly show that disturbed individuals attempting to follow the instructions used here that distinguish "faking good" from "faking normal" do not do so successfully. The patients in this study did not differentiate these two sets of instructions. While this may be an overly fine distinction here and perhaps in other cases, there are applications, such as those in selection for critical positions (e.g., in law enforcement) where this distinction is important in that a denial of minor flaws may be seen as a negative attribute in a candidate (Rogers, 1997). It remains possible that different phrasing of instructions that stressed the differences more saliently would have resulted in better discrimination between the two instruction sets. For example, instructions for faking good may emphasise better than average levels of adjustment. A more powerful test would be a within-subject design in which the same people responded to both sets of instructions and differences might become evident with such a design using the current instructions. Pauls and Crost (2004) used such a design to evaluate different sets of instructions to fake. Additional studies of different instruction sets are also warranted, as initially expressed some time ago (Eysenck et al., 1974).

The results may also have implications for instructions for test administrators. Instructions for test takers that warn individuals of measures to detect faking may not actually influence those taking the tests. While this was not the major goal of this study, the results suggest that such warnings may be ineffective, at least with psy-

Table 3
Identification Rates for Basic Personality Inventory (BPI) Validity Measures and Discriminant Analysis Classifications

Predictors	Sensitivity	Specificity	Positive predictive power	Negative predictive power	Overall accuracy, %
Single FG measure	.34	.86	.69	.59	61.2
Single FB measures	.62	1.0	1.0	.73	81.2
All BPI measures FG	.69	.91	.88	.76	83.5
All BPI measures FB	.76	.91	.90	.76	84.1
Optimal patient FG	.63	.83	.77	.71	73.1
Optimal patient FB	.79	.83	.82	.81	83.6
Optimal student FG	.53	.89	.81	.67	71.6
Optimal student FB	.79	.89	.87	.82	84.1
Heuristic patient FG	.52	.32	.42	.42	41.7
Heuristic patient FB	.97	.28	.61	.89	64.1
Heuristic student FG	1.0	.06	.49	0	52.2
Heuristic student FB	0	.06	0	.06	2.9

Note. Sensitivity = probability of correct identification of a true case; specificity = probability of correct identification of a true noncase; positive predictive power = probability of a true case being correctly identified; negative predictive power = probability of a true noncase being correctly identified; FG = fake good; FB = fake bad. The four test properties can be calculated for any cutoff score on a measure. In this case, any single score over the cutoffs in the text were used, as were the classification outcomes of the five discriminant analyses.

chiatric patients. An alternative possibility is that such faking warnings need to include advice that detection of faking could have negative outcomes for the person. Rothstein and Goffin (2005) and Goffin and Woods found that a warning that included mention of possible negative consequences was effective in producing response patterns different from straight faking good. Both those studies used undergraduate student samples and whether patient groups perform similarly needs to be determined with additional studies.

Consistency in Responding

The use of the criterion of low person reliability as an index of following instruction set raises interesting issues. First, it is an uncommon validity measure. Second, screening samples for adherence to instructions to fake is rarely done, and little is known as to what proportion of respondents who are instructed to fake actually do so. Since its initial description by Holden et al. (1985), there have been few reports of the use of a person reliability measure. As a correlation amongst subscales of the full BPI scales, it assesses aspects of the consistency of responses to different parts of the test booklet (Tellegen, 1988). Third, direct self-report assessment of consistency is unlikely to be valid with psychiatric patients, given the prevalence of suspiciousness, defensiveness and denial in this group. Third, successful maintenance of a false self-image, even on a self-report instrument, does require consistency in maintaining that pose. Therefore finding that about 10% to 15% of respondents in the groups instructed to fake were unreliable in maintaining that pose suggests that the presence of psychological disturbance may make difficult the successful presentation of a false image over time. What is of interest is the finding of roughly the same rate in the group of ostensibly honest respondents. Person reliability is rarely assessed on self-report instruments, and so further study of the issue of low person reliability on self-report instruments is certainly warranted. To date, only the BPI appears to use such an index, and more widespread use and study of person reliability could expand significantly our knowledge in the area of distinguishing valid from invalid profiles.

The unreliability or inconsistency of responding suggests that at least a proportion of psychiatric patients have difficulty maintaining a consistent approach to responding, whether honest or distorted, to even a set of 240 items. Such individuals may have less meaningful responses overall (Clark, Gironda, & Young, 2003), regardless of the original set of instructions, whether to be honest or to present a distorted image of either common type. If this is the case, then measures of unreliable, even random responding, such as the Infrequency scales of the PRF (Jackson, 1984) and PAI (Morey, 1991), may be more useful as additional validity measures in self-report instruments than is commonly appreciated.

Faking Profiles for the BPI

In the present study, the patients' efforts at overreporting are, however, closer to a true psychiatric disturbance than is the average profile produced by students who were attempting to "fake bad" as in the Helmes and Holden (1986) study. Comparison with that previous study showed that the elevation of the patients' "fake bad" profile was lower than that simulated by the students by 2 to 3 points. In contrast, the FG profile from the patients was close in elevation to that reported for the simulating students. The "ST" profile here was much higher than that of the Helmes and Holden students, as would be expected. Also as expected was the lower use of validity indices by the classification functions derived for the patient groups in this study than their use by the equivalent functions for students in the Helmes and Holden (1986) study. Only the number of perseverative responses and the Denial scale were common to the patient and student functions using optimal sets of predictors. Steffan et al. (2007) did not use discriminant analysis in their analysis of faking the BPI amongst prison inmates, using the multivariate analysis of variance (MANOVA) instead. Their profiles of BPI scores under instructions to FG and to malinger also more closely resemble the profiles here in Table 1 than the profiles in Helmes and Holden (1986). Some scales appear to be higher in Steffan et al.'s Table 1 than those here, but the similarities are stronger than the differences between the two nonstudent samples. At the same time, the generalization of the selection of variables used in the Helmes and Holden study was good, with accuracy in classification close to the calculated optimal functions. This study did not, however, attempt to cross-validate the actual functions because the original report did not provide the actual discriminant functions to be used with raw scores of the selected variables. Such weights may not have generalised as well as in the present analysis in which new weights were calculated for the selected variables for the current data set.

These results also confirm that characteristic "faking" profiles for the BPI can be replicated, and that the measures to detect response distortion on the BPI have a degree of utility. At the same time, it is also clear that psychiatric patients can be correctly classified as distorting responses at a somewhat lower rate of about 70% to 80% when compared with the previous study of simulating university students by Helmes and Holden (1986) where from 75% to 90% of cases were correctly classified. Steffan et al. (2007) did not address this issue in their sample of inmates. It is important to remember that figures for test performance given in Table 3 are sensitive to base rates of faking good and bad and different settings may well have different base rates of external incentives for distortion and of clients apt to be internally motivated to distort their responses. The issue of the strength of natural, internal motivation to dissimulate raises an issue with regard to the credibility of simulation studies of faking. This has been the question of incentives to promote efforts at dissimulation (Rogers, 1997). It is not clear if making a larger sum of money contingent upon producing a successful (that is, undetectable) simulated response pattern would lead to different results in a study with patients such as this one.

Detection of Faking and Generalization of Discriminant Weights

The analysis of the performance of the individual BPI validity indices showed very mixed performance. There is only one indicator for underreporting or faking good, a high score on the social desirability scale. While the set of individual validity measures was quite good at classifying the honest ST group (over 85% accuracy in Table 2), the high scores on the social desirability scale correctly classified fewer than 35% of the FG cases, with most of the errors being to class these as in the ST group. There are more

measures for overresponding (faking bad) on the BPI than for underresponding, and they also performed better overall than the single measure of faking good by almost 30% (62% vs. 34% in Table 2).

It was clear that involvement of the four BPI validity measures was much less salient for all functions involving patients than was the case with the classification functions derived by Helmes and Holden (1986) from student data. This is consistent with the original predictions of shifts in the emphasis, but the observation of roughly equivalent classification accuracy in the Optimal Patient and Optimal Student Predictors results was less expected in that there appears to be very little shrinkage in predictive accuracy from the functions based on the selection of variables derived from students when applied to patients.

The discriminant functions based upon the use of all available predictors made fairly accurate distinctions between honest responding and both forms of response distortion. This result would be expected because all available information is being used and the functions are calculated on the present data, which means that some accuracy is gained by optimisation of random error that favours correct classification. Should the most accurate prediction equations from this study be used with new samples, the rates of correct classification would almost certainly be lower with such cross-validation. In reviewing the use of discriminant analysis of intelligence test scores for the detection of faking, Babikian and Boone (2007) refer to similar losses in predictive accuracy with replication, but the functions in the studies that were reviewed replicated quite well.

At the same time, a substantial number of individuals in all three groups was not accurately classified, which suggests in turn that a multivariate procedure designed to maximise group discrimination could not do so with very high accuracy, such as over 90% sensitivity and specificity. For example, no set of prediction results in Table 2 exceeded 80% in overall classification accuracy. This accuracy rate suggests that either the groups of respondents included individuals who were not following instructions and who were not removed by the screening out of people with low person reliability scores or that some respondents responded in manners that differed in significant ways from the majority of their peers who were given the same set of instructions. It is clear that these results need to be replicated in other samples and the prediction equations cross-validated for different populations.

The more parsimonious function based upon the set of the five "best" or optimal individual predictors from the present patient sample showed somewhat lower levels of accuracy. This set of functions had reasonable levels of accuracy in classification for both under- and overreporting, but were more accurate overall for overreporting (or faking bad). This result was consistent with the reports by Kroner et al. (1997) and Steffan et al. (2007) of better identification of faking bad than of faking good by the BPI in samples of offenders and is in contrast to the findings of Bagby et al. (1990) with student samples, where the opposite pattern was noted. Table 3 clearly shows that the BPI generally performs better here at detecting faking bad than faking good. While sensitivity was generally equal for both instruction sets, specificity was consistently higher for faking bad than for faking good in all predictor sets except the heuristic ones, suggesting that the BPI is better at determining who is not faking bad, while doing equally well at determining both who is faking good and faking bad.

Generalization of the variables derived from students (Helmes & Holden, 1986) was surprisingly good. The set of best predictors based on the student data (Optimal Student Predictors) was roughly as accurate as the "best" predictors derived for the patients (Optimal Patient Predictors), as seen in Table 3. The variation from expectation is that the student functions amount to a cross-validation of the selection of the original predictors and might consequently be expected to be less accurate than if the actual weights were used.

The two discriminant functions that were simplified heuristic weights would be appropriate for applications in which computer scoring or calculation of discriminant classification functions was not used. The simplified functions using such weights were, however, the least accurate in classification overall. While the simplified function for patients was well above chance levels in classifying the overreporting (FB) group, this was at the cost of being much less accurate in predicting underresponding in the FG group. The Heuristic Student Predictors were quite successful in classifying the ST and the underreporting FG group, but very poor with the FB group. Overall accuracy was still above the chance level of 33% classification for the FG group cross-tabulation table, but poor for the FB group with no cases correctly classified. While functions using exact weights for a small number of selected predictors generalised well across samples, the heuristic functions generalised poorly.

This result was unexpected in that the multiple regression literature suggests that unit weights generalise well to other samples under many conditions (Dana & Dawes, 2004). Such weights may, however, be less effective in the case of discriminant analysis with its different goals and methods of calculation for the case in which more than two groups are to be differentiated and classified. The classification accuracy for the discriminant functions using heuristic weights proved to be very sensitive to changes in the selection of variables and to the exact values of the weights. Clearly, this is an area in need of both empirical and statistical development. The use of similar heuristic functions in the neuropsychology literature (Babikian & Boone, 2007) appears to replicate fairly well in different samples, but that literature largely involves efforts to replicate the original study and not to derive new weights for new samples with only two groups to deal with. Monte Carlo studies may be particularly informative to address some of the above issues.

In addition, it was also clear that functions based on a limited number of predictor variables could be as effective as functions based upon a complete set of predictors. Dana and Dawes (2004) reported that use of the single best predictor was seldom as accurate as larger sets of predictors, but the present findings are less consistent with their simulation study. There is the important caveat to this in that it appears that the population to whom the functions are applied is also very important. Further studies of applications of the discriminant analysis with the BPI and similar measures across different populations will help to clarify this issue.

The overall rate of successful classification for all the discriminant functions evaluated was within the range of previous studies reporting rates of 81% reported by Helmes and Holden (1986) and the 73% to 78% reported by Bagby et al. (1990) for discriminant functions derived from students, as well as the figures for the PAI reported by Rogers et al. (1996). Despite the apparent similarity in classification rates, this study differs from Rogers et al. and Steffan

et al. (2007) in that instructions were to exaggerate general distress and not to feign specific disorders. The present results are also not congruent with the study of Bagby et al. (1990), who found that the BPI was slightly better at detecting faking good than at faking bad. Here only the heuristic functions tended to be more useful for underreporting, whereas all others were better at classifying overreporting. It may be that differences in the samples, with the present results being based in genuine patients, may account for some of the differences between the outcomes of these two studies.

Future Research

Obviously, there is more research into the value of discriminant analysis in the detection of faking to be done with other populations, particularly ones where faking is likely to be present and where external incentives exist. Such research is difficult, in part because of ethical concerns, but one value of the current study is its use of a reasonably large clinical sample. In addition, the base rate of faking in many populations is simply not known, and knowledge of that parameter is important in evaluating classification rates resulting from discriminant functions. There has been little research on the actual efficacy of warnings that faking might be detected by internal measures of a scale (Goffin & Woods, 1995; Rothstein & Goffin, 2005). This lack is likely related in part to the limited knowledge of actual base rates of response distortion. More experimental work is needed with different types of incentives to determine whether rewards for successful faking in fact improve performance. Does this mean that someone with the chance of a $500,000 settlement in a personal injury lawsuit is more likely to fake successfully than someone with the chance of a $100,000 settlement? Finally, more research into heuristic discriminant functions is warranted. Inspection of the output of the classification functions showed that values of the different functions for both sets of heuristic functions often differed only by one or two units or even only in the first decimal place. Minor differences in values led to drastic changes in classification results. These results are at variance with the literature on multiple regression, and Monte Carlo studies that systematically vary input parameters may help clarify the issues.

Limitations

Clearly, the instruction sets for faking good and faking normal did not work as intended. Only additional study with patient groups will determine whether this was because of the instructions, and what test administration instructions are most effective, notably whether the consequences of being detected are mentioned. A larger sample might have led to changes in the selection of optimal sets of predictors and different conclusions, while a greater monetary incentive might have led to higher rates of person reliability and consistency in responding.

Conclusions

The consistency apparent from the present results suggests that predictions of patients' distorted responses do not differ substantially in accuracy from those for students requested to distort. The results do, however, show a different combination of useful predictive variables, perhaps because of the patients' greater personal experience with psychiatric disturbance. The presence of ongoing psychiatric disturbance of some patients may also be reflected in the tendency to be inconsistent in the pattern of their responding when asked to distort their responses regardless of the exact instructions.

Taken together, these findings provide a basis for a practical actuarial approach to detection of dissimulation on the BPI. Results here are consistent with Morey and Lanier (1998), who also found that discriminant functions were more accurate than individual measures, even when those measures were applied using consistent decision rules. There are also implications for the conceptualization of the concepts of defensiveness, in that if one assumes disturbed individuals are motivated to distort their responses to appear undisturbed, some of them appear unable to do so in convincing manner.

Résumé

La distorsion de la réponse demeure une question importante dans la mesure de la psycholopathologie. Ici, quatre groupes de patients psychiatriques, chacun composé de 40 personnes, ont dû compléter le *Basic Personality Inventory* (BPI; Jackson, 1989); seuls les participants montrant une consistance acceptable dans leur façon de répondre (« fidélité de la personne ») ont été analysés. La performance aux échelles de validité du BPI mesurée avec les points de séparation conventionnels a été comparée aux résultats d'une analyse discriminante linéaire calculée auprès des patients à partir des variables sélectionnées précédemment par Helmes et Holden (1986) pour les étudiants universitaires. Des analyses préliminaires ont révélé que les groupes « bon » et « normal » ne pouvaient pas être séparés; le groupe « normal » a par conséquent été exclu des analyses subséquentes. Les résultats ont montré une meilleure classification avec les analyses discriminantes qu'avec les mesures traditionnelles de validité du BPI. Contrairement aux observations découlant de la régression multiple, la pondération heuristique était la moins précise. Les implications pour l'application du BPI lorsqu'une distorsion de la réponse est suspectée, ainsi que le résultat inattendu concernant la mauvaise performance des fonctions de classification heuristiques, sont discutés.

Mots-clés : feindre, réponses aléatoires, Basic Personality Inventory, analyse discriminante, fidélité de la personne

References

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

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.

Babikian, T., & Boone, K. B. (2007). Intelligence tests as measures of effort. In K. B. Boone (Ed.), *Assessment of feigned cognitive impairment: A neuropsychological perspective* (pp. 103–127). New York: Guilford Press.

Bagby, R. M., Gillis, J. R., & Dickens, S. (1990). Detection of dissimulation with the new generation of objective personality measures. *Behavioral Sciences and the Law, 8,* 93–102.

Bagby, R. M., Marshall, M. B., Bury, A. S., Bacchiochi, J. R., & Miller, L. S. (2006). Assessing underreporting and overreporting response styles on the MMPI-2. In J. N. Butcher (Ed.), *MMPI-2: A practitioner's guide* (pp. 36–69). Washington, DC: American Psychological Association.

Bernstein, I. H. (1988). *Applied multivariate analysis.* New York: Springer-Verlag.

Clark, M. E., Gironda, R. J., & Young, R. W. (2003). Detection of back random responding: Effectiveness of MMPI-2 and Personality Assessment Inventory validity indices. *Psychological Assessment, 15,* 223–234.

Dana, R., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics, 29,* 317–331.

Dawes, R. M., Faust, D., & Meehl, P. E. (1993). Statistical prediction versus clinical prediction: Improving what works. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 351–367). New York: Erlbaum.

Eysenck, S. B. G., Eysenck, H. J., & Shaw, L. (1974). The modification of personality and lie scale scores by special "honest" instructions. *British Journal of Social and Clinical Psychology, 13,* 41–50.

Goffin, R. D., & Woods, D. M. (1995). Using personality testing for personnel selection: Faking and test-taking instructions. *International Journal of Selection and Assessment, 3,* 227–236.

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

Hathaway, S. R., & McKinley, J. C. (1976). *Minnesota Multiphasic Personality Inventory manual* (2nd ed.) New York: Psychological Corporation.

Helmes, E., & Holden, R. R. (1986). Response styles and faking on the Basic Personality Inventory. *Journal of Consulting and Clinical Psychology, 54,* 853–859.

Holden, R. R. (2000). Are there promising MMPI substitutes for assessing psychopathology and personality? Review and prospect. In R. H. Dana (Ed.), *Handbook of cross-cultural and multicultural personality assessment. Personality and clinical psychology series* (pp. 267–302). Mahwah, NJ: Erlbaum Publishers.

Holden, R. R., Helmes, E., Fekken, G. C., & Jackson, D. N. (1985). The multidimensionality of person reliability: Implications for interpreting individual test item responses. *Educational and Psychological Measurement, 45,* 119–130.

Holden, R. R., & Jackson, D. N. (1992). Assessing psychopathology using the Basic Personality Inventory: Rationale and applications. In J. C. Rosen & P. McReynolds (Eds.), *Advances in psychological assessment* (Vol. 8, pp. 165–199). New York: Plenum Press.

Holden, R. R. & Jackson, D. N. (1985). Disguise and the structured self-report assessment of psychopathology. I. An analogue investigation. *Journal of Consulting and Clinical Psychology, 53,* 211–222.

Jackson, D. N. (1984). *Personality Research Form manual.* Port Huron, MI: Sigma Assessment Systems.

Jackson, D. N. (1989). *Basic Personality Inventory manual.* Port Huron, MI: Sigma Assessment Systems.

Jaffe, P. G. (1985). The utility of the Basic Personality Inventory in the assessment of young offenders. *Ontario Psychologist, 17,* 4–11.

Kessel, J. B., & Zimmerman, M. (1993). Reporting errors in studies of the diagnostic performance of self-administered questionnaires: Extent of the problem, recommendations for standardized presentation of results, and implications for the peer review process. *Psychological Assessment, 5,* 395–399.

Kluger, A. N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. *Personnel Psychology, 46,* 763–780.

Kroner, D. G., Forth, A. E., & Mills, J. F. (2005). Endorsement and processing of negative affect among violent psychopathic offenders. *Personality and Individual Differences, 38,* 413–423.

Kroner, D. G., Holden, R. R., & Reddon, J. R. (1997). Validity of the Basic Personality Inventory in a correctional setting. *Assessment, 4,* 141–153.

Kroner, D. G., & Reddon, J. R. (1994). Relationships among clinical and validity scales of the Basic Personality Inventory. *Journal of Clinical Psychology, 50,* 522–528.

Langbehn, D. R., & Woolson, R. F. (1997). Discriminant analysis using the unweighted sum of binary variables: A comparison of model selection methods. *Statistics in Medicine, 16,* 2679–2700.

McNeil, D. C., & Reddon, J. R. (2001). Utility and stability of the Basic Personality Inventory in psychiatric patients with longstanding psychotic disorders in a new psychiatric rehabilitation program over a two-year period. *Psychological Reports, 87,* 767–775.

Merbaum, M. (1972). Simulation of normal MMPI profiles by repressors and sensitizers. *Journal of Consulting and Clinical Psychology, 39,* 171.

Millon, T. (1987). *Manual for the MCMI-II.* Minneapolis, MN: National Computer Systems.

Morey, L. C. (1991). *Personality Assessment Inventory professional manual.* Odessa, FL: Psychological Assessment Resources.

Morey, L. C., & Lanier, V. W. (1998). Operating characteristics of six response distortion indicators for the Personality Assessment Inventory. *Assessment, 5,* 203–214.

Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trials. *Journal of Clinical Psychology, 45,* 239–250.

Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences, 37,* 1137–1151.

Pittenger, D. J. (2001). Power calculator: A collection of interactive programs. *Educational and Psychological Measurement, 61,* 889–894.

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1999). Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement, 23,* 99–115.

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

Rogers, R., Sewell, K. W., Morey, L. C., & Ustad, K. L. (1996). Detection of feigned mental disorders on the Personality Assessment Inventory: A discriminant analysis. *Journal of Personality Assessment, 67,* 629–640.

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review, 16,* 155–180.

SPSS, Inc. (2000). *SYSTAT.* Chicago: SPSS.

Steffan, J. S., Kroner, D. G., & Morgan, R. D. (2007). Effect of symptom information and intelligence in dissimulation: An examination of faking response styles on the Basic Personality Inventory. *Assessment, 14,* 22–34.

Tellegen, A. (1988). The analysis of consistency in personality assessment. *Journal of Personality, 56,* 621–623.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin, 83,* 213–217.

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

Wilkinson, L., Blank, G., & Gruber, C. (1996). *Desktop data analysis with SYSTAT.* Upper Saddle River, NJ: Prentice Hall.

World Health Organization. (1977). *International classification of diseases: Manual of the international statistical classification of diseases, injuries, and causes of death. Ninth revision. Congress, 1975.* Geneva: Author.

(*Appendix follows*)

Appendix

Classification Functions Used

Present Study All Variables

Set II. Fake Bad = −178.27 + 1.07*Hypochondriasis + 2.08*Depression + 2.33*Denial + 2.03*Interpersonal Problems + 1.94*Alienation – 1.18*Persecutory Ideas + 1.17*Anxiety + 3.08*Thinking Disorder + 1.92*Impulse Expression + 1.07*Social Introversion + 1.71*Self-Depreciation −1.07*Deviation + 24.19 * Reliability + .47*True Responses + 0.02* Perseverations + 11.07*Desirability

Set II. Fake Good = −170.11 = 1.06*Hypochondriasis + 1.96*Depression + 2.37*Denial + 1.88*Interpersonal Problems + 1.73*Alienation – 1.01*Persecutory Ideas + 1.26*Anxiety + 2.99*Thinking Disorder + 1.89*Impulse Expression + .81*Social Introversion + 1.76*Self-Depreciation − 1.44*Deviation + 25.66*Reliability + 10.89*Desirability + .46*True Responses + .05*Perseverations

Set II. Straight Take = −168.30 + 1.08*Hypochondriasis + 2.07* Depression + 2.04*Denial + 1.99*Interpersonal Problems + 1.28*Alienation − 0.92*Persecutory Ideas + 1.20*Anxiety + 3.11*Thinking Disorder + 2.02*Impulse Expression + 1.13*Social Introversion + 1.46*Self-Depreciation – 1.38*Deviation +22.41*Reliability + 11.01*Desirability + .44*True Responses + .06*Perseverations

Present Study Best Predictors

Set III. Fake Bad = −18.0 + 0.97*Denial + .77*Alienation + .44*Social Introversion −.13*Self-Depreciation +.11*Perseverations

Set III. Fake Good = −15.1 + 1.09*Denial + .42*Alienation + .22*Social Introversion – .19*Self-Depreciation + .13*Perseverations

Set III. Straight Take = −12.6 + .67*Denial + .15*Alienation + .52*Social Introversion - .33*Self-Depreciation + .14*Perseverations

Helmes & Holden (1986) Student

Set IV. Fake Bad = −44.7 + 2.74*Desirability + .13*Perseverations + 17.13*Reliability + .71*Denial + 2.76*Deviation

Set IV. Fake Good = −46.4 + 2.89*Desirability + .14*Perseverations + 18.36*Reliability + .83*Denial + 2.35*Deviation

Set IV. Straight Take = −43.2 + 2.95* Desirability + .15*Perseverations + 15.12*Reliability + .44*Denial + 2.35*Deviation

Patient Heuristic Weights

Set V. Fake Bad = −14 + Denial + Alienation

Set V. Fake Good = −9 +Denial + Alienation

Set V. Straight Take = −7 + Depression + Denial – Social Introversion

Helmes & Holden (1986) Student Heuristic Weights

Set VI. Fake Bad = −39 + Denial + 2.5*Deviation + 4*Reliability + 2.5*Desirability + perseveration/10

Set VI. Fake Good = −41 + Denial + 2*Deviation + 5*Reliability + 3*Desirability + .15*perseveration

Set VI. Straight Take = −41 + Denial/2 + 2*Deviation + 4*Reliability + 3*Desirability + .15*perseverations

Received June 9, 2008
Revision received February 2, 2009
Accepted February 3, 2009 ■