INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

*For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.

UMI Dissertation Information Service

University Microfilms International
A Bell & Howell Information Company
300 N. Zeeb Road, Ann Arbor, Michigan 48106
Holmes, Christopher Wells

ISSUES REGARDING FAKEABILITY AND THE MANAGERIAL POTENTIAL SCALE OF THE CALIFORNIA PSYCHOLOGICAL INVENTORY

Rice University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

M.A. 1986
RICE UNIVERSITY

ISSUES REGARDING FAKEABILITY AND
THE MANAGERIAL POTENTIAL SCALE OF THE
CALIFORNIA PSYCHOLOGICAL INVENTORY

by

CHRISTOPHER WELLS HOLMES

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

MASTER OF ARTS

APPROVED, THESIS COMMITTEE:

R. L. Dipboye
Professor of Psychology

D. M. Lane
Associate Professor of Psychology

Houston, Texas

May, 1986
ABSTRACT

ISSUES REGARDING FAKEABILITY AND
THE MANAGERIAL POTENTIAL SCALE OF THE
CALIFORNIA PSYCHOLOGICAL INVENTORY

by

CHRISTOPHER WELLS HOLMES

A revised version of the California Psychological
Inventory was administered to 120 college students under
three different instructional conditions in an examination
of the fakeability of the managerial potential scale of the
CPI. The instructional sets and the means for the Mp scale
in each condition were: (a) an honest instructional set, M
= 19.5; (b) a "fake good" managerial potential
instructional set, M = 24.9; and (c) a "fake bad"
managerial potential instructional set, M = 7.8. Analyses
indicated that the Mp scale is highly fakeable on both a
group and individual basis. However, different
interpretations of the overall magnitude of the fakeability
of the Mp scale emerge depending on how fakeability is
defined. The good impression (Gi) scale was found to be
fairly accurate at identifying faked inventory profiles.
Using a cutting score of T = 60, 87 % of the faked good
profiles were correctly identified while only 7 % of the
honest profiles were misclassified. Of several individual
differences variables examined, only SAT verbal scores were
predictive of faking ability.
ACKNOWLEDGMENTS

First and foremost I would like to thank my wife, Lisa, who has supported me, the professional college student, through all the trials and tribulations of my chosen career. This would not have been possible without her love and constant advice.

I would also like to thank my parents and family for the love and support they have given me through my long years of education. I will always remember my father's few choice words of advice, "What ever you end up doing for a living, by all means make sure you enjoy it."

I would also like to thank the members of my masters committee R. J. Harvey, Bob Dipboy, and David Lane for their expert advice and guidance. Special thanks goes to Dr. Harvey whose considerable programming and research skills helped make a conceivably immense task very manageable.

Last, but not least, I would like to thank my fellow members of the "First Year Class of '84" for their support and general insanity. I nicer bunch of people I've never met.
TABLE OF CONTENTS

Chapter 1. Issues Regarding Fakeability 1
  1.1 Introduction 1
  1.2 Previous Research 2
  1.3 Gordon's Proposed Method 9
  1.4 A Further Redefinition 11
  1.5 Individual Differences 14

Chapter 2. Issues Regarding the CPI 16
  2.1 Overview 16
  2.2 Managerial Potential Scale 17
  2.3 Fakeability and the CPI 19

Chapter 3. Study 23
  3.1 Overview 23
  3.2 Method 28
  3.3 Results 32
  3.4 Discussion 38

Appendix I. Items and Correct Answers for the Revised CPI 51

References 61
LIST OF TABLES

1. Variable Means and Standard Deviations 33
2. Intercorrelations Among Variables 35
4. Hierarchical Regression of Mp Change Scores on All Individual Difference Variables 37
5. Response Time Means and Standard Deviations for the Honest, Fake Good, and Fake Bad Conditions 38
CHAPTER 1
Issues Regarding the Concept of Fakeability

Section 1.1 Introduction

The concept of the fakeability of personality inventories or selection tests has long been of concern to psychologists. This concern is well placed considering the fact that the vast majority of these devices rely on self-report measures. Obviously, if these measures are susceptible to systematic distortions on the part of subjects, then any results obtained become of questionable utility (e.g. South, 1980; Thorton and Gierasch, 1980). Several studies have been conducted attempting to identify instruments which are easily faked. Unfortunately, many of these studies have provided contradictory findings regarding certain instruments (e.g. Rusmore, 1956; Bass, 1957). This problem appears to be due, in part, to the application of different operational definitions of fakeability and inadequate statistical testing (Gordon, 1978). If this is the case, then it becomes extremely important to subject any personality inventory to rigorous and uniform statistical testing designed to discover evidence of fakeability before it can be used as a selection or placement device. Also, of importance, is the need to explore individual differences in "faking ability" and to discover variables which may predict these differences. If such variables exist, then they may be used to temper decisions based on the results of
personality inventories in selection and placement situations.

The present paper has four purposes: (a) to examine the results of previous studies regarding the fakeability of several instruments; (b) to briefly discuss a new statistical method for measuring test fakeability outlined by Gordon (1978); (c) to propose a conceptual and operational redefinition of the concept of test fakeability, which follows logically from Gordon's proposal; and (d) to present the results of a study designed to address several key issues regarding test fakeability and the Managerial Potential (Mp) scale of the California Psychological Inventory (CPI).

Section 1.2 Previous Research

Some of the earliest evidence indicating that personality tests are fakeable comes from a study conducted by Mais (1951) regarding the Jurgenson Classification Inventory (JCI). Mais used a repeated measures design and found that subjects could increase their self-confidence scores an average of 12.8 from -5.9 (honest condition) to 6.0 (faked condition) when instructed to fake high self-confidence. This finding was especially interesting because the JCI was a forced-choice inventory designed specifically to deter faking. Longstaff and Jurgenson (1953) obtained similar results when subjects were instructed to fake high self-confidence. However, they also found that subjects
could not raise their scores a significant amount when instructed to fake good over-all scores. This led them to conclude that the inventory was useful in situations where the subjects did not know which specific traits were being assessed.

Maher (1956) found further evidence for the fakeability of forced-choice inventories when he tested his Study Activity Questionnaire (SAQ) for what he termed transparency. He found a total group mean increase of 14.2 points between an honest instructional set administration and faked instructional set administration of his questionnaire.

Rusmore (1956) conducted the first study investigating the fakeability of the Gordon Personal Profile (GPP). He did not find evidence indicating that the GPP was fakeable. However, he did not specifically instruct his subjects to try to fake the test. Instead he tried to simulate conditions in which the subjects, who were college students, would be motivated to fake the test. He did this by instructing them to imagine that they were applying for a job.

The fact that the faking of tests in selection or industrial situations does occur had already been demonstrated by Herzberg (1954) in a study using the Guilford-Zimmerman Temperament Survey. Further evidence for the occurrence of real world faking of selection devices was presented by Bass (1957). He compared the scores a group of
sales applicants obtained on the GPP with those obtained by a matched group of job incumbents. Bass found that the applicants scored significantly higher, on all sub-scales, than the incumbents. On some scales this difference was as large as five standard deviations from the overall mean. This led Bass to conclude that the GPP was of questionable utility in selection situations.

Evidence indicating that the Gordon Personal Inventory (GPI) is also fakeable was presented by Braun (1962) in the form of a repeated measures laboratory experiment. Braun (1962) instructed a group of college students in an industrial psychology class to take the GPI first under normal conditions and then immediately thereafter under instruction to "fake so as to appear the best possible candidate for an executive position in top management" (p. 611). Braun's results indicated that subjects were able to increase their scores significantly, from a mean of 14.95 in the honest condition to a mean of 37.26 in the faked condition. This finding was especially surprising because Gordon and Stapleton (1956) had previously reported finding a non-significant increase in mean total scores on the GPI between faked and honest test conditions for a group of high school students.

Schwab (1970) hypothesized that this difference was due to the order of administration of the faked and honest test conditions. Braun had presented the honest instructional condition first followed by the faked instructional
condition; whereas, Gordon and Stapleton had administered the instructional conditions in a reverse order. To test this hypothesis Schwab administered both the GPI and the GPP to a large group of college students in a counter-balanced repeated measures design. He found that when honest instructional conditions preceded faked conditions the mean differences between test administrations were significant on only 3 out of 8 occasions. However, when the tests were administered in the reverse order, faked followed by honest conditions, the mean differences between conditions were significant on 6 out of 8 occasions. This led Schwab (1970) to conclude that, "quite different interpretations of the magnitude of fakeability emerge depending on the order of instructions." (p. 675)

To further test the fakeability of the GPP and the GPI in a selection situation, Schwab and Packhard (1973) administered the inventories to two groups of female applicants at a medium sized electrical and manufacturing plant. Members of a control group were asked to take the inventories after they had been hired. They were told that the company was cooperating with researchers at a local university and that their answers would be mailed directly to the university. They were also told that their scores would have nothing to do with their employment. Presumably members of this group would not be motivated to fake their answers considering they had nothing to gain. Members of an experimental group were asked to take the inventories and
were told that the inventories were part of the final step in the selection process. Presumably members of this group would be highly motivated to present themselves in the best possible light; thus elevating their scores much as the applicants in Bass's (1957) study had. This was not the case. Schwab and Packhard reported finding no significant mean differences between the two groups scores on both of the inventories.

In another experiment, designed to be a reply to Schwab's (1970) study criticizing previous repeated measures designs, Braun and Farzel (1974) reported finding further evidence for the fakeability of the GPI and GPP. They used an independent measures design in which one group of college students was given an honest instructional set and another was told to fake like an ideal candidate for a top management executive position. They found significant mean differences between the two group scores for all of the scales on the GPI and for 3 out of 4 on the scales on the GPP.

Another forced-choice personality inventory which has been subjected to several fakeability studies is the Edwards Personal Preference Schedule (EPPS). Borislow (1958) presented evidence suggesting that the EPPS can be faked under highly structured personal and social desirability instructional sets but that it is not highly susceptible to faking under normal conditions. Borislow arrived at this conclusion after conducting a repeated measures experiment
involving a single honest instructional administration
followed by a series of faked instructional set
administrations of the EPPS with a group of 21 college
students. Likewise, Dicken (1959) found that when subjects
were instructed to simulate various traits related to the
EPPS they could change their scores substantially.

In a study attempting to detect the fakeability of the
EPPS in a selection situation, Kirchner (1962) found that
applicants for a retail sales position scored significantly
higher than incumbents on four of the EPPS scales. Kirchner
interpreted this as evidence for the fakeability of the EPPS
in selection situations. Orpen (1971a, 1971b) found further
evidence for the fakeability of the EPPS using both college
students, who were instructed to pretend they were
applicants in a selection situation, and real life
applicants for positions in a large insurance company.

The Ghiselli Self-Description Inventory (GSDI) is
another forced-choice inventory which has been studied for
evidence of fakeability, albeit somewhat less extensively
than the GPP, GPI and EPPS. Krug (1958) found that college
students could raise their scores significantly on the
initiative, intelligence, and self-assurance scales of the
inventory if they were given instructional sets specific to
each one of these traits prior to a re-test administration.
Krug also found that bias introduced by a specific set
generalized to other scales in the inventory causing
slightly inflated scores overall. He concluded that this
finding was, "particularly disturbing if the inventory is to be used as a selection instrument, since it increases the number of potential sources of a high overall score." (Krug, 1958, p. 92)

Kaufman, Hakmiller and Porter (1959) found that a group of college students could not increase their scores on the supervisory ability or intelligence scales of the GSDI when they were instructed to pretend they were applying for top or middle management positions. Braun (1962) found similar results regarding the supervisory ability scale. He investigated the effects of three different faked instructional sets on re-test scores for all the scales of the inventory. The three sets were: "general faking (answer so as to put yourself in the best light), occupational title (answer so as to appear an ideal candidate for a designated position), and a specific set to appear high on supervisory qualities..." (Braun, 1962, p. 639). The results indicated that the subjects could raise their scores on all the scales, except for the supervisory ability scale, regardless of the instructional set to which they were assigned.

Several other self-report scales have been shown to be susceptible to faking. These include the Strong Vocational Interest Blank (Bordin, 1943; Longstaff, 1948; Kirchner, 1960; Becker, 1963; Abrahams, Neuman, and Githens, 1971) and the Kruder Form D Occupational Interest Survey (Bridgman and Hollenbeck, 1960). Evidence for the fakeability of the Personal Orientation Inventory and the Personal Orientation
Dimension Scale has also been found (Braun and La Faro, 1969; Ecker and Watkins, 1975; Snyter and Allik, 1981).

Finally, evidence has been gathered indicating that many empirically keyed selection tests are also fakeable. These include generic biodata and life history blanks (Klein and Owens, 1965; Doll, 1971; Schrader and Osburn, 1977) and even specific, tailor made devices (Goldstein, 1971; Thorton and Gierasch, 1980).

Section 1.3 Gordon's Proposed Method

With so much evidence amassed it is difficult not to conclude that a vast majority of personality inventories and selection devices are fakeable, to some degree or another, and that individuals will attempt to fake them under certain conditions. What is interesting is that many different operational definitions and statistical tests have been used to study the fakeability of these devices. It is possible that these differences help to account for the variety of findings regarding the fakeability of certain devices (e.g. Braun, 1962; Schwab, 1970). Gordon (1978) provided an excellent critique of the problem when he stated that:

Because the major concern of test users is that fakeable instruments may provide inaccurate measurements of the interests, aptitudes, and personality of the respondent, studies designed to investigate the magnitude and direction of impression management should employ fakeability criteria operationalized in terms directly bearing on the matter of the accuracy of measurements. However, accuracy is a measurement characteristic seldom discussed by psychologists. Consequently, it is not surprising that the type
of evidence contained in most studies of fakeability is incomplete at best because of failure to deal with all aspects of measurement accuracy. (p. 772)

Gordon then suggested that repeated measures fakeability designs should be analyzed under the framework of Naylor's (1967) model of measurement accuracy. He proposed that a faked test condition is analogous to a fallible measure whereas an honest test condition is an analogous to a less fallible measure. If this is the case then fakeability can be measured in terms of the inaccuracy of the instrument. According to Naylor's (1967) model, inaccuracy can be seen as a function of three sources of "error".

The first is average or constant error attributable to the measuring instrument or to a situation and it should effect each score in the same manner. Gordon (1978) stated that, "The instruction to fake good or fake bad is a situational factor which is likely to produce such a general shift in the mean of the responses on some tests" (p. 774). However, this error should not change the relative positions of the scores in the faked and honest conditions.

The second source is systematic error attributable to individual differences. This source could potentially change the position of the individuals' scores in the two conditions while also raising or lowering the overall mean difference between tests.

The third source is random error which would not have a significant effect on the mean difference because random
error is assumed to be uncorrelated over repeated measures. However, it could change the relative positions of the individuals' scores in the two conditions.

Gordon further proposed that in order to fully assess the fakeability of any instrument, statistical tests must be used which are sensitive to both constant error (due to instructional sets) and systematic error (due to individual differences). He then suggested using a test for significant difference in the mean scores of the two test conditions (honest vs. faked) to estimate constant error, followed by a calculation of the variance in the difference or change scores in order to estimate the amount of error due to individual difference variables. The larger the variance of the change scores the more likely that individual differences in faking ability exist. Unfortunately, only a subjective judgement on the part of the experimenter can be used to assess whether the majority of this variance is due to individual differences or random error. If the change scores variance is extremely large, and it appears to be due mostly to individual differences, then the usefulness of the test in a selection situation must be questioned.

Section 1.4 A Further Redefinition of Fakeability

Gordon (1978) brought up an extremely important point regarding research in test fakeability. In essence, what previous researchers had been doing was simply accepting the
null hypothesis when they failed to find significant mean
differences between honest and faked test conditions. This
fallacy often lead them to conclude that the particular
instrument under investigation was highly resistant to
deliberate dissimulation attempts on the part of individuals
taking the test. As Gordon pointed out these inappropriate
conclusions were based on a poor operational definition of
fakeability which lead to inadequate statistical testing.
However Gordon's proposed remedy to this situation is still
subject to a problem which may also have affected the
findings in previous studies.

The problem is possible ceiling effects in the
magnitude of the change scores, for certain individuals, due
to the use of an "honest" test score as a baseline or
standard for which to compare "faked" test scores against.
Some individuals may score high enough on a certain
instrument so that even if they are very capable of faking
the attribute it measures, this ability will not adequately
manifest itself when the honest test score is subtracted
from the faked test score. Thus using the difference
between faked and honest scores is not an adequate measure
of the fakeability of an instrument. A concrete
illustration may help to explicate this problem.

Take, for example, two individuals, Subject A and
Subject B, who are both in a repeated measures experiment
designed to determine if a particular instrument which
measures managerial potential is fakeable. Subject A
obtains a score of 30 out of a possible 32 on the test, in an honest instructional condition, and Subject B obtains a score of 20 in the same condition. Following this condition the subjects obtain scores of 32 and 28, respectively, under an instructional set to "fake so as to appear as the best possible managerial candidate for a certain position."

Subtracting the subjects' honest scores from their faked scores we get change scores of 2 and 8, respectively. Thus it appears that Subject B is better at faking the instrument than Subject A, given the implicit definition of faking ability involved in this operationalization of the concept (i.e. faking ability is the ability to raise ones score on a scale above what ones "true" score is on the scale). However, Subject A obtained a perfect score in the faked condition whereas Subject B missed 6 possible points in the faked condition. Obviously, in an absolute sense subject A was better at discerning what the correct answers were for the instrument, and therefore better at faking the instrument, than Subject B. The decision to regard the instrument as fakeable, in such an experiment, could depend largely on the relative differences between the subjects honest test scores not on their ability to perceive the correct answers for the instrument. In a worst case scenario an experimenter could have a small sample consisting completely of subjects similar to subject A and conclude that the instrument under question was not fakeable due to a non-significant mean difference between test
conditions and a small variance in the change scores.

In order to truly test the fakeability of any instrument it is apparent that a new operational definition of fakeability, which is not dependent on honest test scores, must be employed. The most logical solution to this problem is to use a repeated measures design in which subjects are instructed to "fake good" and "fake bad" on the instrument in question. If subjects are able to identify the correct answers for the instrument (i.e. the instrument is highly fakeable) then they should be able to fake their scores on the instrument equally as well in either direction. Using this method the fakeability of the instrument would then be defined in terms of the difference between the test scores in the fake good condition and those in the fake bad condition. Gordon's method of examining both the mean difference between test conditions and the variance of the change scores would then be employed to form a truly rigorous test of the fakeability of the instrument.

Section 4.5 Individual Differences

If there are individual differences in faking ability, as Gordon's proposal suggests, then it follows that there may be reliable measures which predict these differences. Unfortunately, a review of the of the literature showed relatively few studies addressing this issue.

Several studies found that college students faked more than job incumbents when instructed to do so (e.g. Kirchner,
1960; Klein and Owens, 1965; Abrahams, Neuman, and Githens, 1971). However, this does not necessarily mean that college students as a group are "good fakers". Given the circumstances college students are probably more willing to fake a test than job incumbents and thus they appear as better fakers.

Schrader and Osburn (1977) found no differences between males and females in faking ability on a biodata instrument. Holmes (1985) found similar results in a study regarding the fakeability of the supervisory ability scale of the Ghiselli Self Description Inventory.

Finally, in a study having a direct bearing on the current study, Dicken (1960) found that expert psychologists were somewhat better at faking personality traits on the California Psychological Inventory than college students. Not only were they better at faking their scores on the inventory, but they were also able to accomplish this without raising their scores on the good impression scale which was designed to detect faking. This was a very interesting finding considering that it is logical to assume that an integral part of the ability to fake a test, on an individual level, is the ability to do so without being detected. This study, and its implications will be discussed further in subsequent chapters. Suffice it to say the relative paucity of research in this area makes it a fruitful one for further research.
CHAPTER 2

Issues Regarding the California Psychological Inventory

Section 2.1 Overview

The California Psychological Inventory (CPI) was developed in the late forties and early fifties in research tracing its origins to early studies involving the Minnesota Multiphasic Personality Inventory (Gough, 1957). The original test consisted of fifteen scales designed to measure personality dispositions related to the every day, interpersonal behaviors of "normal" individuals (Gough, 1951).

The CPI has undergone several revisions since its introduction. Currently, the test consists of four hundred and eighty true-false items with eighteen embedded scales.

The scales are specifically designed to measure what Gough (1968) calls "folk concepts" which are, "aspects and attributes of interpersonal behavior that are to be found in all cultures and societies, and that possess a direct and integral relationship to all forms of social interaction" (p. 57). Gough's decision to use folk concepts is based on his belief that the best method for constructing a test is to examine the setting in which the test is to be used, and then to develop measurements based on the constructs already in operational usage there. Thus, for Gough, theory is secondary to empirical analysis.

This emphasis becomes apparent following an examination
of the methods used in developing the various scales. Of
the eighteen scales, thirteen were developed using an
external criterion method in which a large pool of items
were administered to groups of individuals representing
opposite extremes of a personality trait. Item analyses
were then carried out in an attempt to identify questions
which differentiated the groups in a reliable manner. The
remaining five scales were developed using a more rational-
intuitive method involving internal consistency analysis
(Megargee, 1972).

The scales have been used successfully in a variety of
ways (i.e. as part of the whole test, in combination with
other scales, or by themselves) in situations ranging from
clinical assessment to research attempting to predict
academic success (e.g. Holland, 1959; Howell, 1966; and
Gough and Kirk, 1970). Several experimental or special
purpose scales have also been derived from the test for use
in specific situations or contexts. These include the
Amenability scale (Rudoff, 1959), the Anxiety scale
(Leventhal, 1966), and the Managerial Potential scale
(Goodstein and Schrader, 1963). It is to this last scale,
that a large part of of this study is addressed.

Section 2.1 Managerial Potential Scale

Goodstein and Schrader (1963) developed what they
termed a "good manager" scale using an external criterion
method similar to the method used to develop thirteen of
the original scales. They administered the CPI to a sample of 603 managers from eight U.S. Army Ordinance Corps Field Service Depots, and to a comparison sample of 1,748 nonmanagerial employees at the same depots. Using item analysis, they were able to identify 206 items which differentiated the two groups at the .01 level of confidence. They also found that individuals in top management scored significantly higher than those in line supervision. By dividing the sample of managers into three groups; top management (n=106), middle management (n=245), and line supervision (n=252), they were able to find positive correlations between scale scores and an overall index of job performance within each group. The obtained coefficients were .25, .27, and .12, respectively. For all 603 managers, the correlation between the scale scores and the index of job performance was .23 (p < .01).

Orpen (1972) found similar results by applying the scale to a sample of 186 men. He contrasted scores of 62 persons in managerial positions with the scores of 124 persons in nonmanagerial positions. He found a significant difference between the mean scores of the two groups and a significant correlation of .27 between the scores of the managerial group and peer ratings of "competence".

Gough (1984) published a shortened version of this scale which came out of efforts to revise the CPI. Using data from a split sample of 200 Air Force officers and data gathered from a sample of 49 bank managers compared to
normative data gathered from archival records, Gough was able to identify 34 items which differentiated the groups with a high degree of confidence. Gough also reported finding correlations of .88 and .89 between the old scale and the new scale in normative samples of 1,000 men and 1,000 women. Finally, he reported finding correlations of .23, between the new scale scores and performance ratings in the sample of 100 officers used to evaluate the items; and .20, between the new scale scores and performance ratings in a holdout sample of 100 officers.

Considering these results, it is apparent that this new scale could be used to identify individuals who exhibit managerial potential. If this is the case, then the scale is potentially very useful in selection or placement situations. However, before it is used extensively as such a device it should be subjected to a rigorous examination to determine if it is fakeable.

Section 2.3 Fakeability and the CPI

Unfortunately, there has not been a great deal of research regarding the fakeability of the CPI. The research that has been conducted, regarding this issue, has focused on using the well being (Wb) scale to detect "negative" dissimulation and the good impression (Gi) scale to detect "positive" dissimulation.

The Wb scale is one of the eighteen original scales. It consists of 44 items designed to "identify individuals
relatively free from self-doubt and disillusionment" (Gough, 1975, p.10). The Wb scale is also used in clinical settings to identify individuals who are deliberately trying to lower their scores on the entire test in order to highlight their personal problems. When Wb scores fall way below all the other scores (approaching zero in most cases) it is taken as a very good indicator of deliberate negative dissimulation (see Megargee, 1972). Validity coefficients for this scale range from .26. for a comparison between Wb scores and expert ratings of "health and vitality" in a sample of 100 military officers; to .27, for a comparison between Wb scores and self-ratings on general physical fitness in the same sample (Gough, 1975). Test re-test reliabilities for this scale range from .71 to .75 (Gough, 1975). Several studies have shown that when groups are asked to "fake bad" on the CPI the group's mean Wb score decreases drastically. In one study the mean Wb score for a group of 354 college students who were instructed to "feign anxiety" was 11.81 as compared to a normative group mean of 37.46 for similar individuals taking the inventory under normal instructions (Gough, 1969).

The Gi scale is also one of the eighteen original scales. It consists of 40 items designed to "identify persons capable of creating a favorable impression, and who are concerned about how others react to them" (Gough, 1975, p.10). The Gi scale is also used to identify individuals who deliberately attempt to raise their scores on the test
in various settings. Validity coefficients for the Gi scale range from .32, for a comparison between Gi scores and Im scores from the Strong Vocational Interest Blank in a sample of 152 adult males (Magargee, 1972); to .60, for a comparison between Gi scores and K scale scores from the Minnesota Multiphasic Personality Inventory in a sample of 150 adult males (Gough, 1975). Test re-test reliability coefficients for this scale range from .68 to .81 (Gough, 1975). The research regarding the usefulness Gi scale in detecting faking is also fairly positive.

Canter (1963) found that alcoholic patients were able to significantly improve their overall CPI scores when instructed to do so. However, in all cases the patients also raised their Gi scores considerably more than their scores on the other scales.

Dicken (1960) attempted to determine whether the CPI could be selectively altered. Four groups of 20 students were administered the CPI under standard conditions, and then again under instructions to role-play dominance, responsibility, intellectual efficiency, and flexibility. The first three groups were able to significantly raise their scores on the Do, Re, and Ie scales while the individuals in the forth group were not able to raise their scores on the Fx scale. A fifth group which was instructed to make the best impression possible produced generally elevated scores on all the scales. In every case Gi scores increased substantially, and in every group except one
(flexibility), it was the Gi scores that had the greatest increase of all the eighteen scales. By applying a cutting score of T=60, after all the scores had been standardized with a mean of 50 and standard deviation of 10, Dicken was able to detect 79 per cent of the dissimulated records while misclassifying only 3 per cent of those obtained with the standard administration.

As mentioned earlier in Chapter 2, Dicken also found that expert psychologists were better at faking the inventory than the college students. He had two groups of ten psychologists take the CPI under standard and role-playing conditions. He found that, in all cases, these subjects were able to alter their relevant scores successfully without significantly increasing their Gi scores. This finding, although it casts somewhat of a shadow on the validity of the Gi scale, does lend support for the notion of individual differences in faking ability.
CHAPTER 3

Study

Section 3.1 Overview

The present study was designed with four specific purposes in mind. These were: (a) to compare the merits of the different methods for operationalizing fakeability outlined in Chapter 1; (b) to examine the possible fakeability of the managerial potential (Mp) scale of the California Psychological Inventory (CPI); (c) to further examine the usefulness of the good impression (Gl) scale as a tool for detecting positively faked test profiles; and (d) to attempt to identify individual difference variables which may predict "faking ability".

In order to accomplish these goals in a feasible manner a decision was made to administer a shortened version of the CPI to a group of college students using an repeated measures design with three instructional conditions. This design was chosen in order to allow all subjects to take the test under: (a) an honest or normal instructional condition; (b) a "fake good managerial potential" instructional condition; and (c) a "fake bad managerial potential" instructional condition. This allowed a comparison of the results obtained using the different methods for operationalizing fakeability.

The shortened version of the CPI was used mainly because of constraints on the available subject pool. Had
the full length CPI been administered, the experiment would have required at least 3 hours of participation for each subject. There was also the question of possible random answering and boredom effects if the subjects had been required to take the full inventory three times. Finally, previous research has shown that administering shortened versions or specific scales of the CPI does not greatly alter the validity of the scales (e.g. Milbraith & Klien, 1962; Leventhal, 1970). In light of this evidence, it appears that use of a shortened version of the CPI has considerable merit for research purposes.

The Mp scale was, of course, included in the revised version of the CPI so that it could be tested for resistance to deliberate dissimulation attempts on the part of the subjects. If the Mp scale is resistant to deliberate faking then it could possibly be used as an aide to selection or placement decisions either by itself, or as a component of the whole CPI. This is very important if the scale is to be used in industry because it is highly likely that individuals in charge of hiring would try to use the scale by itself, disregarding the other scales.

The Gi and Wb scales were included in the revised version of the CPI for two reasons. First, if the Mp scale is highly fakeable then it becomes important to know if the Gi scale can be used to detect individuals who fake in a positive manner. If the Gi scale is highly accurate in identifying such individuals, then the Mp scale could still
be used in conjunction with the G1 scale in a selection or placement situation. If, however the G1 does not identify positively faked profiles accurately then the utility of both scales, as aides to personnel decision making, is highly suspect.

The second reason for including the G1 and Wb scales in the new version was so that they could be used as control variables in regressing the subjects change scores (difference between their Mp scores in the fake good and fake bad conditions) on various individual differences measures. This was done because, as mentioned earlier in Chapter 2, an integral component of the ability to fake a test on an individual level is the ability to do so without being detected. Therefore faking ability, in this experiment, was defined as the ability to systematically distort scores on the Mp scale without causing severe distortions of scores on the G1 and Wb scales in the different conditions. From this it follows that the only truly interesting individual differences measures which may predict faking ability, are ones which do so over and above any predictive ability of the Wb and G1 scales.

Two other scales from the CPI were included in the revised inventory. These were the sociability (Sy) scale and the flexibility (Fx) scale. The Sy scale consists of 36 items designed to, "identify persons of outgoing, sociable, participative temperament" (Gough, 1975, p.10). Validity coefficients for this scale range from .24, for a comparison
between Sy scores and "sorority-joining" in a sample of 174 freshman women (Hase and Goldberg, 1967); to .44, for a comparison between Sy scores and peer ratings in a sample of 190 freshman women (Hase and Goldberg, 1967). Test re-test reliability coefficients for this scale range from .68 to .84 (Gough, 1975). The Sy scale was included because I hypothesized that there would be a positive relationship between sociability and faking ability. This hypothesis is supported, somewhat, by the results of previous study (Holmes, 1985) regarding individual differences and faking ability on the supervisory ability scale of the GSDI. In this study a small but positive correlation \( r = .25, p < .10 \) was found between subjects change scores and their scores on the interaction orientation scale of the Orientation Inventory (Bass, 1977).

The Fx scale consists of 22 items designed to "indicate the degree of flexibility and adaptability of a person's thinking and social behavior" (Gough, 1975, p.11). This scale was validated by attempting to show negative relationships with measures of "rigidity". Gough (1975) reports coefficients ranging from -.36, for a comparison between Fx scores and expert ratings of rigidity in a sample of 40 medical school seniors; to -.58, for a comparison between Fx scores and scores on the California F scale in a sample of 180 college students. Test re-test reliability coefficients for this scale range form .49 to .67 (Gough, 1975).

This scale was included because I hypothesized that
there would be a positive relationship between flexibility and faking ability. Although I found no empirical evidence supporting this hypothesis it seems to be a logical assumption that such a relationship would exist.

The questions from the Self-Monitoring Scale (Snyder, 1974) were also included in the revised inventory. This was done because I hypothesized that there would be a positive relationship between level of self-monitoring and faking ability. Research regarding the Self Monitoring Scale offers fairly strong support for this hypothesis. Lippa (1976) found that people categorized as high self-monitors were better at adopting the behavior of another type of person, when asked to role-play, than those categorized as low self-monitors. Snyder (1976) also found that high self-monitors describe themselves as flexible, adaptable, and shrewd; whereas, low self-monitors describe themselves as consistent and principled. If this is the case, then it seems highly likely that the more self-monitoring an individual is the better he would be at faking a personality inventory.

Subjects were also asked to report their SAT scores. These scores were later checked for accuracy through access to the students' records. The SAT scores were obtained so that they could be used as an index of general intelligence. This was done because I hypothesized that there would be a positive relationship between general intelligence and faking ability. Support for this hypothesis can be found
from the considerable evidence indicating that general intelligence is predictive of a variety of skills and abilities in a variety of settings (e.g. Hunter and Pearlman, 1981).

Finally, latency measures of response time to each question in each condition were obtained. This was done to for two reasons. First, so that the subjects change scores could be regressed on their composite response times for the faked conditions. This was done to determine if there was a relationship between overall response time, in the faked conditions and faking ability. The second reason for including these measures was so that comparisons between overall response times in the three conditions could be made. I hypothesized that subjects would take longer to answer the questions in the faked conditions than in the honest conditions.

Section 3.2 Method

Subjects. One hundred and twenty male and female undergraduates from a small southern university participated in this experiment in groups ranging in size from 8 to 19 individuals for a total of 10 sessions. Each subject received 1 hour of participation credit for a class they were enrolled in.

Procedure. The revised inventory, consisting of 188 true/false questions, was administered to the subjects using
an interactive computer program, IBM PASCAL, and 19 IBM XT personal computers. The subjects were randomly assigned to one of six treatment conditions, with the restriction of equal cell size, based on all possible ordered combinations of the three instructional sets (honest, fake good, fake bad). This counter-balanced design was used in an attempt to eliminate possible order effects (see Schwab, 1970). The questions were also administered to each subject in a random order for each condition in an attempt to eliminate any possible order effects due to question presentation.

The instructional set for the honest condition was adapted from the original instructions for the CPI. The exact instructions used were as follows:

The following inventory contains a series of statements. Read each one, decide how you feel about it, and then indicate your answer by pressing the proper key. If you think you agree with the statement, or feel that it is true about you, press T for TRUE. If you disagree with the statement or feel that it is not true about you press F for FALSE.

The instructional set for the fake good managerial potential condition was adapted from a set of instructions used by Schrader and Osburn (1977) in an experiment regarding the fakeability of a biodata instrument. The exact instructions were as follows:

Imagine that you have just been hired by a large, Fortune 500 firm and that you are currently undergoing an assessment process which will play a large part in determining the position in which you will be placed. You are aware of the fact that the company has openings in several positions, and that you are a part of a small group of individuals hired to fill these positions. The positions are very similar in
terms of starting salary, benefits, and room for advancement. However, you are also aware of the fact that there are only a few positions open in management. You feel that these are very desirable positions and you have decided to present yourself as the best possible candidate for these positions. As part of the assessment process you have been asked to fill out the following inventory. Because you really want to be a manager you have decided to fake the inventory so that you will appear as the best possible candidate for a managerial position. Please answer the following questions as if you were trying to "beat the inventory" and present yourself as the best possible candidate for a managerial position. The questions will appear on the screen followed by a prompt for your answer. Press T for TRUE and F for FALSE. Remember that you are trying to fake this inventory so as to appear as the best possible managerial candidate!!

The instructional set for the fake bad managerial condition was a simple variation of the one used for the fake good condition. The exact instructions were as follows:

Imagine that you have just been hired by a large, Fortune 500 firm and that you are currently undergoing an assessment process which will play a large part in determining the position in which you will be placed. You are aware of the fact that the company has openings in several positions, and that you are a part of a small group of individuals hired to fill these positions. The positions are very similar in terms of starting salary, benefits, and room for advancement. However, you are also aware of the fact that there are only a few positions open in management. You feel that these are very undesirable positions and you have decided to present yourself as the worst possible candidate for these positions. As part of the assessment process you have been asked to fill out the following inventory. Because you really do not want to be a manager you have decided to fake the inventory so that you will appear as the worst possible candidate for a managerial position. Please answer the following questions as if you were trying to "beat the inventory" and present
yourself as the worst possible candidate for a managerial position. The questions will appear on the screen followed by a prompt for your answer. Press T for TRUE and F for FALSE. Remember that you are trying to fake this inventory so as to appear as the worst possible managerial candidate!!!

The computers were arranged in rows of carrels in the experimental room with one computer per carrel. As subjects entered the room they were instructed to take a seat at an appropriate computer. The subjects were then given a brief verbal description of how to operate the computers and asked if they had any questions regarding the computers. After any pertinent questions were answered the subjects were instructed to type in their subject number and press the return key to start the program. The subject numbers had previously been written on small slips of paper and place beside the computers. Immediately after the subjects pressed the return key the instructions for the first condition appeared on the screen. The subjects were allowed to read the instructions and answer all the following questions at their own pace. At the bottom of each set of instructions was a line directing the subjects to press any key to continue. Once they pressed any key the first question appeared on the screen inside a double edged boarder. Instructions directing the subjects to press "T" for true and "F" for false also appeared with each question. A new question appeared on the screen following each response. This process continued until the subjects answered all 188 questions in the first condition. After
the questions had been answered in the first condition, instructions for the next condition appeared on the screen and subjects were instructed by the computer to continue. This process continued until the subjects had completed all three conditions.

After the subjects had completed this section of the experiment a set of prompts appeared instructing the subjects to type in their name, the class number for which they wished to receive experimental credit, their SAT verbal score, and their SAT quantitative score. Following these prompts were a set of instructions directing the subjects to contact the experimenter, who was seated in the front corner of the room. As each subject completed the experiment he/she was given a written debriefing form and allowed to leave the room.

Section 3.3 Results

General. The analyses performed in this experiment were based on the subjects scores for: (a) the Mp scale in all three conditions (honest, fake good, and fake bad); (b) the Gi scale in the fake good condition (c) the Wb scale in the fake bad condition; (d) the Sy scale in the honest condition; (e) the Fx scale in the honest condition; (f) the Self-Monitoring scale in the honest condition; (g) the SAT verbal test; (h) SAT quantitative test; and (i) their overall combined response times to the questions in the faked conditions, recorded in seconds from the presentation
of a question to the onset of a response. The mean and standard deviation for each of these variables is presented in Table 1.

Table 1  
Variable Means and Standard Deviations

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>MP (honest condition)</td>
<td>19.50</td>
<td>4.56</td>
</tr>
<tr>
<td>MP (fake good condition)</td>
<td>24.95</td>
<td>3.51</td>
</tr>
<tr>
<td>MP (fake bad condition)</td>
<td>7.78</td>
<td>3.42</td>
</tr>
<tr>
<td>GI (fake good condition)</td>
<td>32.04</td>
<td>5.79</td>
</tr>
<tr>
<td>WB (fake bad condition)</td>
<td>5.06</td>
<td>1.45</td>
</tr>
<tr>
<td>SY (honest condition)</td>
<td>26.08</td>
<td>5.10</td>
</tr>
<tr>
<td>FX (honest condition)</td>
<td>10.93</td>
<td>3.70</td>
</tr>
<tr>
<td>SM (honest condition)</td>
<td>16.71</td>
<td>3.21</td>
</tr>
<tr>
<td>SATV</td>
<td>614.5</td>
<td>96.9</td>
</tr>
<tr>
<td>SATQ</td>
<td>656.5</td>
<td>76.5</td>
</tr>
<tr>
<td>RT</td>
<td>1141.7</td>
<td>313.4</td>
</tr>
</tbody>
</table>

Note: MP = managerial potential scale, GI = good impression scale, WB = well being scale, SY = sociability scale, FX = flexibility scale, SM = self-monitoring scale, RT = total response time for faked conditions.

Fakeability. In order to allow a comparison of the results obtained using the different methods for measuring fakeability, and also to examine the fakeability of the Mp scale, a one-way analysis of variance (ANOVA) was conducted comparing the Mp scores from the different instructional conditions. This showed a significant overall difference between conditions, F(2, 238) = 642.36, p < .01. Planned comparisons between the condition means were made using students t tests. These tests indicated that the means for all three conditions (honest, fake good, and fake bad) were significantly different from each other, p < .01.
Following these comparisons the subjects scores for the Mr scale in the honest condition were subtracted from their scores for the scale in the fake good condition. The resulting change scores had a mean of 5.44 and a variance of 27.48. Next, the subjects scores for the Mr scale in the fake bad condition were subtracted from their scores in the fake good condition. The resulting change scores had a mean of 17.18 and a variance of 26.90.

In order to examine the accuracy of the G1 scale in identifying positively dissimulated test profiles the subjects scores on the G1 scale in the honest and faked good conditions were converted to T scores. Following this a cutting score of \( T > 60 \) was used to classify profiles as the belonging to the faked good group (c.f. Dicken, 1960). This resulted in the correct identification of 89% of the faked good profiles while only 7% of the honest profiles were misclassified.

**Individual Differences.** Simple correlations between all of the variables to be used in predicting faking ability were computed as a preliminary step in examining the individual differences measures. The resulting correlation matrix is presented in Table 2 on the following page.

Next, a series of hierarchical regressions were conducted in order to examine the relationships between faking ability and the individual difference variables. First the Mr change scores (difference between fake good and fake bad conditions) were regressed on each variable.
Table 2
Intercorrelations Among Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>MPH0</th>
<th>MPFG</th>
<th>MPFB</th>
<th>CS</th>
<th>GI</th>
<th>WB</th>
<th>SY</th>
<th>FX</th>
<th>SM</th>
<th>SATV</th>
<th>SATQ</th>
<th>RT</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPH0</td>
<td></td>
<td>.18</td>
<td>.02</td>
<td>.11</td>
<td>-.03</td>
<td>.08</td>
<td>(.50)</td>
<td>(.22)</td>
<td>(-.36)</td>
<td>.16</td>
<td>.04</td>
<td>-.09</td>
</tr>
<tr>
<td>MPFG</td>
<td></td>
<td></td>
<td>-.12</td>
<td>(.75)</td>
<td>(.69)</td>
<td>(-.07)</td>
<td>.07</td>
<td>(.21)</td>
<td>-.04</td>
<td>.06</td>
<td>.10</td>
<td>-.08</td>
</tr>
<tr>
<td>MPFB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FX</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SATV</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SATQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: MPH0 = managerial potential scale (honest), MPFG = managerial potential scale (fake good), MPFB = managerial potential scale (fake bad), CS = change scores for the Mp scale between the fake good and fake bad conditions, GI = good impression scale (fake good), WB = well being scale (fake bad), SY = sociability scale (honest), FX = flexibility scale (honest), SM = self-monitoring scale (honest), RT = total response time for faked conditions. () p < .01. [] p < .05.
separately using the GI scores from the fake good condition and Wb scores form the fake bad condition as control variables. The results from these analyses are presented in Table 3.

Table 3
Separate Hierarchical Regressions Of Mp Change Scores on Individual Difference Variables Controlling for GI and Wb Scores

<table>
<thead>
<tr>
<th>Variable</th>
<th>RSQ(Unique)</th>
<th>RSQ(Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SATV</td>
<td>.0196*</td>
<td>.5383**</td>
</tr>
<tr>
<td>SATQ</td>
<td>.0009</td>
<td>.5196**</td>
</tr>
<tr>
<td>RT</td>
<td>.0002</td>
<td>.5189**</td>
</tr>
<tr>
<td>SM</td>
<td>.0020</td>
<td>.5207**</td>
</tr>
<tr>
<td>FX</td>
<td>.0104</td>
<td>.5291**</td>
</tr>
<tr>
<td>SY</td>
<td>.0012</td>
<td>.5199**</td>
</tr>
</tbody>
</table>

Note: Variable labels are the same as in Table 2. RSQ(Unique) = unique variance accounted for by each variable above that of the Mp and Wb scales. RSQ(Total) = total variance accounted for by each model. * p < .05. ** p < .05.

Following these analyses all of the variables were included in a hierarchical regression analysis in order to further examine the inter-relationships of the variables and faking ability. The main purpose of this analysis was to determine each variables unique relationship with faking ability when all the variables were included in the model.
The results from this analysis are presented in Table 4.

Table 4
Hierarchical Regression of Mq Change Scores on All Individual Difference Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>RSQ(Incr)</th>
<th>RSQ(Unique)</th>
<th>RSQ(Total)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GI</td>
<td>.3111**</td>
<td>.2265**</td>
<td></td>
</tr>
<tr>
<td>WB</td>
<td>.2076**</td>
<td>.1904**</td>
<td></td>
</tr>
<tr>
<td>SATV</td>
<td>.0195*</td>
<td>.0170*</td>
<td></td>
</tr>
<tr>
<td>SATQ</td>
<td>.0030</td>
<td>.0015</td>
<td></td>
</tr>
<tr>
<td>RT</td>
<td>.0013</td>
<td>.0018</td>
<td></td>
</tr>
<tr>
<td>SM</td>
<td>.0028</td>
<td>.0042</td>
<td></td>
</tr>
<tr>
<td>FX</td>
<td>.0025</td>
<td>.0024</td>
<td></td>
</tr>
<tr>
<td>SY</td>
<td>.0035</td>
<td>.0035</td>
<td>.5513**</td>
</tr>
</tbody>
</table>

Note: Variable labels are the same as those in Table 2. The order of entry of the variables following the Gi and Wb scores was determined using the variables simple correlations with the change scores. RSQ(Incr) = incremental variance accounted for by each variable. RSQ(Unique) = unique variance accounted for by each variable. RSQ(Total) = total variance accounted for by the model. ** p < .01. * p < .05.

**Response Time.** In order to obtain measures of the overall condition response times, the response times to the individual questions were summed for each condition. The means and standard deviations of the condition response times are presented in Table 5 on the following page.

A one way ANOVA indicated that that there was a significant overall difference between the response times in each condition, $F(2, 238) = 5.72, p < .01$. Planned follow up comparisons were made between the means using students t tests. The results indicated that the mean for the honest condition was significantly different form those for the faked conditions, $p < .05$. However, the means for the
faked conditions were not significantly different from each other, \( p > .05 \).

**Table 5**

<table>
<thead>
<tr>
<th>Condition</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honest</td>
<td>629.99</td>
<td>209.86</td>
</tr>
<tr>
<td>Fake Good</td>
<td>581.99</td>
<td>184.31</td>
</tr>
<tr>
<td>Fake Bad</td>
<td>559.67</td>
<td>191.57</td>
</tr>
</tbody>
</table>

Section 3.4 Discussion

Judging from these results it appears that the Mp scale is highly fakeable, when it is scored and interpreted independently of the other scales, regardless of how fakeability is operationalized. In both cases the overall mean difference between the relevant conditions was significant indicating that the Mp scale is susceptible to "constant error" faking due to instructional sets. In both cases the variances of the relevant change scores were also quite large indicating that the Mp scale is susceptible to "systematic error" due to individual differences in faking ability. Thus it appears that Gordon (1978) was correct in identifying individual differences as a problem area in research on test fakeability. Following Gordon's (1978) logic it is apparent from these results that the Mp cannot be used independently of the rest of the inventory in a selection situation.
However, what is also interesting is the fact that quite different interpretations of the overall magnitude of the fakeability of this scale emerge when using the different definitions of fakeability. When this aspect of fakeability was defined as the difference between the fake good and fake bad conditions the mean difference was 17.18. This was quite a bit larger than the resulting mean difference of 5.44 when this type of fakeability was defined, using the traditional method, as the difference between the fake good and honest scores. This finding coupled with a subjective examination of the actual questions and answers to the Mp scale (presented in Appendix I) seems to indicate that the "correct" answers to a large proportion of the questions on the scale are easily identifiable. However, the extent of the transparency of this scale was underestimated when fakeability was defined using the traditional method. It may be argued that the faking bad of an inventory has little relevance to the "real world". However, the purpose of laboratory research in this area is to determine if the correct answers to such inventories are easily identifiable before they are used in the real world. Therefore it is apparent that the most rigorous test of the fakeability of an instrument is one which uses the new method of measuring fakeability i.e. an examination of the difference between a set of scores obtained under instructions to fake good and a set of scores obtained under instructions to fake bad followed by an
examination of the variance of the difference scores).

If the results of this study can be generalized to previous studies, then it becomes apparent that there is a possibility that the fakeability of other instruments has been underestimated. Considering the robust findings regarding the fakeability of several of these instruments using the traditional definition of fakeability (e.g. Mais, 1951; Maher, 1956; and Braun, 1962) one can only imagine what the findings may have been if the new definition had been used.

The inadequacies of the traditional method for defining fakeability may also help to account for contradictory findings regarding the fakeability of other instruments. It is possible that some researchers may have found non-significant differences between conditions, while others did, because of ceiling effects in the change scores due to high honest test scores or because of individual differences in faking ability. If this is the case then the differences between the results of some studies may simply be a function of differences, between samples, of the subjects honest test scores or their ability to fake the inventory in question. This problem may have been compounded by small sample size considering the vast majority of the previous studies used samples of less than 30 subjects.

Although the results of this study indicate that the Mp scale is highly fakeable by itself they also seem to indicate that the Gl scale is highly accurate in identifying
positively faked profiles. Considering this finding it is highly likely that the CPI as a whole is not fakeable. However, before this is taken as conclusive evidence in support of the idea of using these scales together in a selection or placement situation several caveats must be made.

First, it must be remembered that the subjects in this experiment were college students who had nothing to lose if they were "caught" faking. In fact, they were specifically instructed to try and "beat the inventory". Because of this it appears that they misrepresented themselves on this inventory with a vigor. This caused their GI scores to increase drastically in the fake good condition. The mean GI score in this condition was 32.04 which translates to a T score of 70. Thus it was exceedingly easy to detect faked profiles using a cutting score of T = 60. Whether such blatant faking occurs in real world situations is another matter entirely. The evidence regarding this phenomenon is mixed. Goldstein (1971) compared information from application blanks provided by 111 prospective applications for nursing home positions with information obtained from their previous employers. He found considerable discrepancies between the two accounts in the areas of length of previous employment and ending salary. He also found that 17 of the applicants had never been employed by the companies they indicated as previous employers. Also, some previous research regarding the fakeability of other
personality inventories indicates that applicants tend to score considerably higher on personality inventories than matched groups of incumbents (e.g. Herzberg, 1954; Bass, 1957). However, other researchers have found no differences between the scores of similarly matched groups of individuals on other inventories (e.g. Schwab and Packard, 1973). Whether this is an indication of differences in the fakeability of the inventories or differences in faking attempts is an open question. Research in the area of impression management has also shown that individuals are less likely to misrepresent themselves, or will alter their presentations, when they believe there is a chance of being "caught" (e.g. Schlenker, 1975). Considering the mixed evidence it is highly likely that individuals in the real world might be more subtle in their faking attempts than the subjects in the study were. If this is the case then it is possible that the Gi scale would not detect such faking.

Second, it must be remembered that 7% of the honest test profiles were misclassified as being faked profiles in this experiment. It is possible that the individuals who were so identified may have faked their answers even in the honest condition. However, considering the circumstances surrounding the study, this seems highly unlikely. If this scale has a 7% false hit rate in a controlled study such as this then it will probably be even less accurate in a real world setting.

These first two problems would probably be minimized if
the CPI was administered and interpreted as a whole by an experienced psychologist. Unfortunately, this would be extremely costly thus defeating one of the main purposes for using such a standardized instrument. Therefore this scale could only be used, by novices, as an indicator of possible faking. This would severely limit its utility in making personnel decisions. When would a personnel specialist decide to eliminate a possible candidate from the running because of a high Gi score and when would he choose to ignore the score?

Third, it is possible that the Gi scale may measure attributes which are desirable in certain business environments. Evidence supporting this notion, in a tangential manner, comes from a study conducted by Ruch and Ruch (1967). They found that K scores on the Minnesota Multiphasic Personality Inventory (MMPI) were the second best predictors of successful performance ($r = .39, p < .01$) for a sample of 182 sales representatives. Considering that the Gi and K scales are very similar in what they measure, it is possible that Gi scores might also predict successful sales or even managerial performance. If high Gi scores are related to managerial success then by using them to screen out individuals a personnel specialist might, in the process, screen out the best candidates.

Finally, it must be remembered that the demonstrated criterion related validity coefficients for the Mp scale are relatively low ranging from .20 to .23. Whether
inferences made from this scale would add to the predictive efficiency of other measures such as cognitive abilities tests is an open question (Gough, 1984). If the Mp scale does not have such properties then the fakeability of the scale and the validity of the G1 scale as a lie detection scale become moot points. Clearly more research is needed regarding both the Mp and G1 scales before they can be used in selection or placement situations.

Turning now to examine individual differences in faking ability we find mixed results. That there are individual differences in faking ability of the Mp scale is hard to deny. In fact one individual raised his/her score on the scale 26 points from the fake bad to the fake good condition while another individual actually lowered his/her score 4 points between the two conditions. Following Gordon's (1978) logic, as mentioned earlier, this finding acts to compound the problems in trying to use the scale as a personnel device. If the scale had been susceptible to constant error faking alone then adjustments in cut off points could be made to allow for possible elevated scores. However such adjustments become useless in the presence of individual differences. If the G1 scale was extremely accurate at identifying faked profiles then it could be used to eliminate such problems. Unfortunately, as the discussion above has shown there are still several potential problems with using the G1 scale in the real world.

Examining the attempt to predict faking ability using
other instruments, we find somewhat weak support for the relationship between faking ability and general intelligence using SAT verbal scores as an indicator of such intelligence. The findings did show a significant, albeit small (RSQUARE = .017), unique relationship between SATV scores and faking ability even when all of the variables were included in the regression model. That such a relationship was found could be construed as not entirely surprising. There was a significant verbal component in the task at hand. However, the subjects were required to do much more than simply read the questions. They were required to delve beyond the surface meanings of the questions and interpret the implications of their answers to each question in terms of managerial potential. Considering this it appears that, once again, we find evidence in support of the predictive validity of general intelligence; even in such a unique task as faking a personality inventory.

The results regarding the relationships between the personality variables and faking ability were somewhat disappointing. None of the personality measures included in this study showed a significant relationship with faking ability in any of the analyses. There could be several reasons for this.

First, it is possible that personality variables in general are not related to the ability to fake a paper and pencil test. When individuals try to fake such a test all
they are equipped with is some general idea as to what the
test is supposed to measure. They are not generally allowed
access to social or interpersonal cues that are present in
more normal situations. Thus highly sociable or flexible
individuals who are very good at picking up on the
appropriate actions and responses in social situations, and
therefore good at presenting themselves in a favorable light,
may be at a loss in the relatively sterile environment of a
test taking situation. Previous research implies that high
self-monitors are at their best in role-playing when they
have an "audience" (Lippa, 1976). If this is the case then
it seems possible that high self-monitors would not be good
at faking paper and pencil tests which are usually
administered in an anonymous or individualized setting.

Second, it is possible that the sample of college
students used was too homogeneous in it's personality make
up to truly test the relationships between these variables
and faking ability. Previous research regarding other
personality attributes, such as locus of control (cited in
Phares, 1976) lends credence to the idea of a restriction of
range for personality variables in samples of college
students.

Finally, it is possible that the unusual nature of the
circumstances under which these measures were obtained (i.e.
the measures were obtained from an "honest" instructional
condition which was was imbedded in a series of faked
instructional conditions) caused the subjects to alter their
responses, either consciously are subconsciously. This would have altered the reliability and validity of the measures, thus decreasing the chance of finding relationships between the variables and faking ability.

There also appears to have been no relationship between the amount of time the subjects took to answer the questions in the faked conditions and faking ability. This is not entirely surprising either. The subjects answered the questions in these conditions fairly rapidly. It is highly likely that they did not take the time to mull over questions they were not to sure about, opting instead to go with their first impression for most of the questions. Also, it is important to note that 26 of the items on the MP scale are keyed for "false" while only 8 of the items are keyed for "true". This may have facilitated a flat response rate for the questions. Once the subjects "locked on to" the correct answers to the questions they may have simply answered the majority of them in the same manner depending on what condition they were in. Considering these two points it is highly likely that the variance of the overall response times is simply a function of differences in reading ability or general motivation to finish the experiment.

What is interesting are the differences in overall response times between the conditions. The subjects actually took longer to answer the questions in the honest condition than in the faked conditions. This is the exact
opposite of what was predicted. The most parsimonious explanation for this finding is that the scale was so easy to fake, for the majority of the subjects, that they answered very rapidly in the faked conditions. This seems highly likely considering the overall magnitude of the fakeability of this scale. The subjects probably took longer in the honest condition because they were less sure of how they really felt about the questions.

The present study has answered several questions regarding fakeability research in general and specifically the fakeability of the Mp scale. However, it has also raised several questions that need to be addressed in future research.

The fakeability of the Mp scale and the validity of the Gi scale as a detection device need to be further examined, both in the laboratory and in the field. In the laboratory, further studies need to be conducted examining the usefulness of the Gi scale as a detection device. It is highly likely that subjects would be capable of altering their scores on the Mp scale without being detected by the Gi scale if they knew that there was a possibility of their being "caught". If this is the case then there is little hope for using these scales as aides in making personnel decisions.

In the field, further studies need to be conducted examining the validity of these scales as well as the prevalence of real world faking. Most of the studies
regarding the latter issue were conducted over a decade ago. Considering current economic conditions, it seems highly likely that this problem has intensified in recent years. One possible study would use a variation of the method of examining application blank information for accuracy introduced by Goldstein (1971). Prospective applicants for a position could be given a modified version of the CPI containing the Mp and Gi scales. They would also be asked to fill out application blanks. The applicants Gi scores could then be examined for evidence of severe elevation and the information on the application blanks could be checked for accuracy by contacting their previous employers. This would give an indication of the degree of faking occuring. The applicants Gi scores could also be correlated with the incidence of misinformation on the application blanks in an attempt to further validate the Gi scale. Ideally applicants would then be hired on the basis of other information. Later the performance ratings of the individuals could be correlated with their Mp scores in an attempt to further validate the Mp scale. Also of interest would be an examination of any possible relationship between the measures of faking and performance. It is possible that faking ability or the willingness to fake is related to managerial ability.

Finally, further research needs to be conducted examining individual differences in faking ability. This is a relatively neglected area of research which needs to be
explored further before any general conclusions may be drawn regarding the phenomenon.
APPENDIX I

Items and Correct Answers for the Revised CPI

Managerial Potential Scale

1 F
I have had very peculiar and strange experiences.
2 F
In most ways the poor man is better off than the rich man.
3 T
I take a rather serious attitude toward ethical and moral issues.
4 F
Most people would tell a lie if they could gain by it.
5 T
I have no dread of going into a room by myself where other people have already gathered and are talking.
6 F
I get pretty discouraged sometimes.
7 F
I don't blame anyone for trying to grab all he can get in this world.
8 F
Most people will use somewhat unfair means to gain profit or an advantage rather than to lose it.
9 T
I wake up fresh and rested most mornings.
10 F
I frequently notice my hand shakes when I try to do something.
11 F
Teachers often expect too much work from the students.
12 F
I often act on the spur of the moment without stopping to think.
13 F
My way of doing things is apt to be misunderstood by others.
14 F
I have had more than my share of things to worry about.
15 F
I am quite often not in on the gossip and talk of the group I belong to.
16 F
I like to keep people guessing what I'm going to do next.
17 F
Most people are honest chiefly through fear of being caught.
18 F
Most people inwardly dislike putting themselves out to help other people.
19 T
If given the chance I would make a good leader of people.
20 F
The future is too uncertain for a person to make serious plans.
21 F
I am often bothered by useless thoughts which keep running through my head.
22 F
A person does not need to worry about other people if only he looks after himself.
23 F
When prices are high you can't blame a person for getting all he can while the getting is good.
24 F
I think most people would lie to get ahead.
25 T
I usually feel that life is worthwhile.
26 F
It is hard for me to act natural when I am with new people.
27 F
I feel that I have often been punished without cause.
28 F
I have never done any heavy drinking.
29 T
I think I am usually a leader in my group.
30 T
I enjoy planning things, and deciding what each person should do.
31 F
Success is a matter of will power.
32 T
People seem naturally to turn to me when decisions have to be made.
33 F
It seems that people used to have more fun than they do now.
34 F
My skin seems to be unusually sensitive to touch.

Good Impression Scale

1 F
Some people exaggerate their troubles in order to get sympathy.
2 T
I always follow the rule: business before pleasure.
3 F
I gossip a little at times.
4 F
There are a few people who just cannot be trusted.
5 F
It is hard for me to start a conversation with strangers.
6 F
I sometimes pretend to know more than I really do.
7 F
Sometimes I feel like smashing things.
8 F
I hate to be interrupted when I am working on something.
9 F
Sometimes I feel like swearing.
10 F
Sometimes I cross the street just to avoid meeting someone.
11 F
I like to boast about my achievements every now and then.
12 F
I must admit I often try to get my own way regardless of what others may want.
13 F
Sometimes I think of things too bad to talk about.
14 F
I must admit that I often do as little work as I can get by with.
15 T
I like to listen to symphony orchestra concerts on the radio.
16 F
I do not always tell the truth.
17 T
I always try to consider the other fellow's feelings before I do something.
18 T
I feel as good now as I ever have.
19 T
I enjoy hearing lectures on world affairs.
20 F
Criticism or scolding makes me very uncomfortable.
21 F
If I am not feeling well I am somewhat cross and grouchy.
22 F
I feel nervous if I have to meet a lot of people.
23 T
I do not mind taking orders and being told what to do.
24 F
Most people are secretly pleased when someone else gets into trouble.
25 T
The most important things to me are my duties to my job and to my fellow man.
26 F
When things go wrong I sometimes blame the other fellow.
27 F
Sometimes at elections I vote for men about whom I know very little.
28 T
I would like to belong to a discussion and study club.
29 F
I am apt to show off in some way if I get the chance.
30 F
Sometimes I just can't seem to get going.
31 F
I must admit that I have a bad temper, once I get angry.
32 T
I have never deliberately told a lie.
33 F
There have been a few times when I have been very mean to
another person.
34 F
At times I have been very anxious to get away from my
family.
35 F
Sometimes I rather enjoy going against the rules and doing
things I'm not supposed to.
36 F
There have been times when I have worried a lot about
something that was not really important.
37 F
Every now and then I get into a bad mood, and no one can do
anything to please me.
38 F
Most people would tell a lie if they could gain by it.
39 F
I get pretty discouraged sometimes.
40 F
I often act on the spur of the moment without stopping to
think.

Well Being Scale

1 F
I am made nervous by certain animals.
2 F
Several times a week I feel as if something dreadful is
about to happen.
3 F
I find it hard to keep my mind on a task or job.
4 F
Once a week or oftener I feel suddenly hot all over, without
apparent cause.
5 F
I can remember "playing sick" to get out of something.
6 T
I usually expect to succeed in things I do.
7 F
I am so touchy on some subjects that I can't talk about
them.
8 T
I have very few quarrels with members of my family.
9 F
At times I have a strong urge to do something harmful or
shocking.
10 F
I don't seem to care what happens to me.
11 F
I am afraid to be alone in the dark.
12 F
I have nightmares every few nights.
13 F
I have a great deal of stomach trouble.
14 F
I have been afraid of things or people that I knew could not hurt me.
15 T
Any man who is able and willing to work hard has a good chance of succeeding.
16 T
I hardly ever feel pain in the back of the neck.
17 F
When I was a child I didn't care to be a member of a crowd or gang.
18 F
When I am feeling very happy and active, someone who is blue or low will spoil it all.
19 F
Everything tastes the same to me.
20 F
Much of the time my head seems to hurt all over.
21 F
My people treat me more like a child than a grown-up.
22 F
Some of my family have habits that bother and annoy me very much.
23 F
No one seems to understand me.
24 F
I dream frequently about things that are best kept to myself.
25 F
I have reason for feeling jealous of one or more members of my family.
26 F
There are certain people I dislike so much I am inwardly pleased when they are catching it for something they have done.
27 F
My mouth feels dry almost all the time.
28 F
When I am cornered I tell that portion of the truth which is not likely to hurt me.
29 F
Life usually hands me a pretty raw deal.
30 F
I have one or more bad habits which are so strong that it is no use fighting against them.
31 F
I am bothered by acid stomach several times a week.
32 T
I get all the sympathy I should.
33 F
I have felt embarrassed over the type of work that one or more members of my family have done.
34 F
I have often felt guilty because I have pretended to feel more sorry about something than I really was.
35 F
The things some of my family have done have frightened me.
36 F
My skin seems to be unusually sensitive to touch.
37 F
I am troubled by attacks of nausea and vomiting.
38 F
I would have been more successful if people had given me a fair chance.
39 F
Almost every day something happens to frighten me.
40 F
My family has objected to the kind of work I do, or plan to do.
41 F
There seems to be a lump in my throat much of the time.
42 T
I usually feel that life is worthwhile.
43 F
I think most people would lie to get ahead.
44 F
Sometimes I cross the street just to avoid meeting someone.

Sociability Scale

1 T
I enjoy social gatherings just to be with people.
2 T
A person needs to "show off" a little now and then.
3 F
When in a group of people I usually do what the others want rather than make suggestions.
4 T
As a child I used to be able to go to my parents with my problems.
5 T
I seem to be about as capable and smart as most others around me.
6 T
I liked school.
7 F
A windstorm terrifies me.
It is very hard for me to tell anyone about myself.

I usually feel nervous and ill at ease at a formal dance or party.

I have at one time or another in my life tried my hand at writing poetry.

I can be friendly with people who do things which I consider wrong.

I like to be the center of attention.

When in a group of people I have trouble thinking of the right things to talk about.

I was a slow learner in school.

I am likely not to speak to people until they speak to me.

I do not dread seeing a doctor about a sickness or injury.

It makes me uncomfortable to put on a stunt at a party even when others are doing the same sort of thing.

I have a tendency to give up easily when I meet difficult problems.

I would like to wear expensive clothes.

I like parties and socials.

I should like to belong to several clubs or lodges.

Once in a while I laugh at a dirty joke.

At times I have worn myself out by undertaking too much.

I love to go to dances.

People pretend to care more about one another than they really do.

I like to read about history.

I am a good mixer.

I have no fear of water.

I like to read about science.

In school I found it very hard to talk before the class.
31 T
I like science.
32 F
I am bothered by people outside, on streetcars, in stores, etc., watching me.
33 T
I have no dread of going into a room by myself where other people have already gathered and are talking.
34 F
I am quite often not in on the gossip and talk of the group I belong to.
35 T
If given the chance I would make a good leader of people.
36 F
It is hard for me to act natural when I am with new people.

Flexibility Scale

1 F
I often wish people would be more definite about things.
2 F
It is annoying to listen to a lecturer who cannot seem to make up his mind as to what he really believes.
3 F
I find that a well-ordered mode of life with regular hours is congenial to my temperament.
4 F
It is hard for me to sympathize with someone who is always doubting and unsure about things.
5 T
I often start things I never finish.
6 F
Our thinking would be a lot better off if we would forget about words like "probably," approximately," and "perhaps."
7 F
I never make judgments about people until I am sure of the facts.
8 F
A strong person will be able to make up his mind even on the most difficult questions.
9 F
For most questions there is just one right answer, once a person is able to get all the facts.
10 F
I like to have a place for everything and everything in its place.
11 F
I don't like to work on a problem unless there is the possibility of coming out with a clear-cut and unambiguous answer.
12 F
It bothers me when something unexpected interrupts my daily routine.
13 F
Most of the arguments or quarrels I get into are over matters of principle.
14 F
I am known as a hard and steady worker.
15 F
I don't like things to be uncertain and unpredictable.
16 F
Once I have my mind made up I seldom change it.
17 F
I think I am stricter about right and wrong than most people.
18 F
I am in favor of a very strict enforcement of all laws, no matter what the consequences.
19 F
I always see to it that my work is carefully planned and organized.
20 F
I set a high standard for myself and I feel others should do the same.
21 F
The trouble with many people is that they don't take things seriously enough.
22 F
People who seem unsure and uncertain about things make me feel uncomfortable.

Self-Monitoring Scale

1 T
I'm always trying to figure myself out.
2 T
I'm concerned about my style of doing things.
3 F
Generally, I'm not very aware of myself.
4 T
It takes me time to overcome my shyness in new situations.
5 T
I reflect about myself alot.
6 T
I'm concerned about the way I present myself.
7 T
I'm often the subject of my own fantasies.
8 T
I have trouble working when someone is watching me.
9 F
I never scrutinize myself.
10 T
I get embarrassed very easily.
11 T
I'm self-conscious about the way I look.
12 F
I don't find it hard to talk to strangers.
13 T
I'm generally attentive to my inner feelings.
14 T
I usually worry about making a good impression.
15 T
I'm constantly examining my motives.
16 T
I feel anxious when I speak in front of a group.
17 T
One of the last things I do before I leave the house is look in the mirror.
18 T
I sometimes have the feeling that I'm off somewhere watching myself.
19 T
I'm concerned about what other people think of me.
20 T
I'm alert to changes in my mood.
21 T
I'm usually aware of my appearance.
22 T
I'm aware of the way my mind works when I work through a problem.
23 T
Large groups make me nervous.
References


