

6-30-2006

A comparison of analog and applied setting research methods in evaluating motivational distortion

Steven Verstraete

Follow this and additional works at: <http://scholarworks.rit.edu/theses>

Recommended Citation

Verstraete, Steven, "A comparison of analog and applied setting research methods in evaluating motivational distortion" (2006). Thesis. Rochester Institute of Technology. Accessed from

This Thesis is brought to you for free and open access by the Thesis/Dissertation Collections at RIT Scholar Works. It has been accepted for inclusion in Theses by an authorized administrator of RIT Scholar Works. For more information, please contact ritscholarworks@rit.edu.

A Comparison of Analog and Applied Setting Research Methods in Evaluating
Motivational Distortion

Graduate Thesis/Project

Submitted to the Faculty

Of the School Psychology Program

College of Liberal Arts
ROCHESTER INSTITUTE OF TECHNOLOGY

By

Steven K. Verstraete

In Partial Fulfillment of the Requirements
for the Degree of
Master of Science and
Advanced Graduate Certificate

Rochester, New York

June 30, 2006

Approved: Scott P. Merydith, Ph.D.
Scott P. Merydith, Ph.D.(committee chair)

Dr. Jennifer Lukomski
committee member)

RIT
School Psychology Program
Permission to Reproduce Thesis

PERMISSION GRANTED

Title of thesis A Comparison of Analog and Applied
Setting Research Methods in Evaluating Motivational
Distortion

I Steven K. Verstraete hereby **grant permission** to the
Wallace Memorial Library of the Rochester Institute of Technology to reproduce my
thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Date: 7/1/06 Signature of Author: _____

PERMISSION FROM AUTHOR REQUIRED

Title of thesis _____

I _____ **prefer to be contacted** each time a
request for reproduction is made. I can be reached at the following address:

PHONE: _____

Date: _____ Signature of Author: _____

PERMISSION DENIED

TITLE OF THESIS _____

I _____ hereby **deny** permission to the Wallace
Memorial Library of the Rochester Institute of Technology to reproduce my
thesis in whole or in part.

Date: _____ Signature of Author: _____

Table of Contents

Chapter 1: Review of Literature.....	3
Chapter 2: Methodology.....	14
Chapter 3: Results.....	17
Chapter 4: Discussion.....	19
References.....	26
Tables.....	29
Figure.....	33

Running head: MOTIVATIONAL DISTORTION

A Comparison of Analog and Applied Setting Research Methods in Evaluating Motivational
Distortion

Steven K. Verstraete

Rochester Institute of Technology

Abstract

The purpose of this study was to evaluate the appropriateness of analog research designs in the assessment of motivational distortion. The responses of analog experimental groups instructed to fake-good (analog, $n = 22$) or respond honestly (classroom, $n = 22$) on the Sixteen Personality Factor (16PF) 5th Edition were compared to the responses obtained from a group of applicants to a graduate program in school psychology (applicant, $n = 38$). Results indicate a consistent pattern of increasing score desirability across groups, from classroom, to applicant, to analog, with significant differences ($p < .002$) between groups. Thus, the applicability of analog research findings to applied settings is called into question and Cattell's (1968, 1986) trait-view theory is supported.

A Comparison of Analog and Applied Setting Research Methods in Evaluating Motivational Distortion

While self-report questionnaires are one of the most common methods used in the assessment of personality, the accuracy of their results is inherently dependent on the honesty of the person being assessed (Schuerger, 1992). Thus, a test-taker can potentially influence the results of his or her assessment by falsely responding to the items of any given personality questionnaire. This phenomenon is broadly known as motivational distortion, or simply faking-good. Cattell (1968) described distortions as stemming from an interaction between the test-taker's personality and his or her reaction to the testing situation. This can occur either on the unconscious level, unbeknownst to the test-taker, or as a result of a conscious decision to fake one's responses. Paulhus (1984) confirmed this two-component view of social desirability using factor analysis. This study, consisting of more than 150 items and scale totals from various social desirability measures, indicated the presence of two primary factors: self-deception and impression management. Further confirmatory factor analysis lent additional support to this model of social desirability-related distortion by demonstrating its more adequate fit in comparison to competing two-factor explanations (i.e., attribution/denial). Paulhus and Reid (1991) elaborated on these two sources of response distortion. In self-deception, the distortion of test-takers' responses is assumed to stem from an innate desire to present themselves in a positive light to society in general. Impression management however, refers to intentional response distortions intended to present a specific image of the test-taker to a specific audience. While the latter source of distortion may seem more devious and threatening, it should be noted that both self-deception and impression management lead to inaccurate representations of the

test-taker's personality. Thus, they should both be taken into consideration when assessing motivational distortion in any situation.

Burger (2000) explains that there are generally two broad types of impression management distortions: socially-desirable responding, or faking-good, and malingering, or faking-bad. Faking-good involves attempting to present oneself in a more favorable light by responding in accordance with traits desirable in the situation at hand. This type of responding might be found in situations where a personality assessment is given as part of an application process for a job or for admission into college. In contrast, individuals who fake-bad respond in a manner that depicts them as possessing more negative attributes than they actually have. Faking-bad could possibly occur in situations such as a criminal trying to use an insanity defense, or a psychiatric patient who does not want to leave a secure hospital environment.

In an extensive review of the research on motivational distortion, Merydith and Wallbrown (1991) noted that as early as the turn of the 20th century, psychologists measuring personality have been aware of the potential for individuals to distort their responses in accordance with specific tasks. This knowledge was applied in a practical sense in a classic study by Meehl and Hathaway (1946). These researchers developed scales and procedures to identify and correct for faking-good and faking-bad on the Minnesota Multiphasic Personality Inventory (MMPI) which are still used today on the revised version, the MMPI-2 (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). The *L* (lie) scale assesses distortion by presenting the test-taker with the opportunity to endorse a number of items describing highly desirable, but extremely rare traits. If a high number of these items are endorsed, it is assumed that the test-taker is distorting his or her responses to look good (faking-good) (Greene, 1991). The *F* (infrequency) scale operates in a similar fashion by presenting items describing highly

atypical, deviant traits. If a test-taker endorses a large number of these items, it is assumed that he or she is intentionally trying to present as having some form of psychopathology (faking-bad). The *K* (correction) scale is designed to assist in the identification of test-takers who exhibit significant psychopathology, but whose MMPI profiles are normal. Based on *K* scale scores, clinicians can adjust five of the Clinical scale scores, Psychasthenia (7), Schizophrenia (8), Hypochondriasis (1), Psychopathic Deviate (4), and Hypomania (9), to correct for distorted responding (Greene, 1991).

Today, nearly all major objective self-report personality measures employ some method to identify this threat to the validity of their results. The 5th edition of the 16 Personality Factor (16PF) detects response distortions using the Impression Management (IM) scale (Russell & Karol, 1993). This scale is comprised of 12 items designed to assess the degree to which a test-taker endorses socially desirable attributes while denying attributes that are undesirable. High IM scores indicate higher levels of response distortion and suggest that the examiner should interpret the results of the test cautiously. Variables such as the reason for testing and the level of importance of accurate results should be considered.

Conn and Rieke (1994) found that IM scale scores correlated with a number of the 16PF Primary personality factors. In particular, they noted that individuals with high IM scores also tended to have Primary factor scores that contributed to a Global score profile of Extraversion, lower Anxiety, and higher Self-Control. In turn, low IM scores were related to Global profiles of Introversion, higher Anxiety, and lower Self-Control. Only the Reasoning (B), Dominance (E), and Sensitivity (I) Primary factors were found to have no significant correlation with IM scores. The 5th edition differs from past versions of the 16PF in that there are no procedures for correcting test results based on the output of the IM scale.

Extensive research evaluating the validity of measures designed to detect response distortions, such as the *L*, *F*, and *K*, scales from MMPI-2, and the MD scale from earlier versions of the 16PF, has indicated general support for their validity and usefulness. In a study investigating the validity scales of the original MMPI, Grow, McVaugh, and Eno (1980) found that as much as 81% of the variance associated with faking could be accounted for by utilizing the various detection and correction measures built into the test. However, Grow and colleagues did advocate a slightly modified use of the MMPI detection scales. Specifically, if an individual's *F* scale score exceeds his or her *K* scale score by 7 or more, faking-bad is indicated, or if the *K* scale score exceeds the *F* scale score by 11 or more, faking-good is indicated. While these researchers' findings suggested adjusted applications of the detection measures for clinical use, they continued to lend support to the use of the measures in general. Examining the newer MMPI-2, Wetter, Baer, Berry, Smith, and Larsen (1992) found that the revised test's validity scales were sensitive to both random responding and faking. Further, their results showed promise for techniques using a combination of the *F* and *VRIN* (Variable Response Inconsistency) scales to differentiate between distortions stemming from random responding and those resulting from faking.

Numerous other studies (e.g., Cassisi & Workman, 1992; Bagby, Nicholson, & Buis, 1995) have yielded similar results, establishing a strong body of support for the utility of validity scales on the MMPI/MMPI-II. However, as Merydith, Humphreys, and Ebener (1997) point out, most of the support for the validity of these scales is derived from studies using a common analog design. Specifically, participants are randomly assigned to groups and instructed to complete a given self-report measure either honestly, faking-bad, or faking-good. For example,

in Austin's (1992) study on the MMPI-2, college student volunteers were assigned to fake-good, fake-bad, and honest groups. The participants in the fake-bad group were instructed to complete the MMPI-2 as if they were trying to increase their disability benefits or beat a criminal charge using the insanity defense. The fake-good group was told to answer as if they were applying for a job and were trying to make a good impression. The honest group was simply told to answer truthfully. Both faking groups were instructed to fake in such a way so as to make it hard to tell they were faking. Typically, detection measures are then found to clearly differentiate faked and honest respondents. While the internal validity of this type of design is sound, the external validity is questionable. Samples of college students may not be representative of the test-takers in applied settings that measures of response distortion are used with. Therefore, it is unclear whether the results of such studies can be generalized to applied settings where personality measures are actually used.

An additional factor that should be considered when examining the utility of response validity scales for use in applied situations is the possibility of coaching. That is, in some real-world testing situations, test-takers may have gained prior knowledge of the existence of response validity measures and/or how to beat them (Storm & Graham, 2000). In fact, Wetter and Corrigan (1995) found that the majority of the law students and attorneys they surveyed felt that it was their duty to prepare and educate their clients about response validity measures prior to allowing them to take psychological assessments. Lees-Haley (1997) even reported an instance where an attorney attempted to limit the amount of time his client was allowed to undergo psychological assessment so as to limit the evaluator's ability to use methods other than MMPI-2 to assess the client's condition and truthfulness. Further, Baer and Miller (2002) noted

that even without the presence of attorneys, motivated individuals could conceivably do their own research prior to undergoing psychological assessment by using books on psychological testing available at many college or even public libraries. The possibility of coaching or prior knowledge of validity scales becomes more concerning when the results of studies comparing coached and un-coached responses and the utility of validity scales for detecting them are considered. Both Baer and Miller (2002) and Storm and Graham (2000) found that the ability of the standard MMPI-2 validity scales to detect faked responses was significantly diminished when respondents had received some kind of coaching on avoiding detection. Specifically, the latter study found that overall predictive power of the MMPI-2 validity scales, which ranged from .83 to .96 when identifying un-coached fakers, decreased to a range of .54 to .86 when attempting to classify coached fakers. Thus, the potential for coaching adds another concern for the validity of studies using analog-only research designs.

Some studies have directly examined the effectiveness of distortion measures in applied situations. Merydith, Humphreys and Ebener (1997) conducted a study using an older version of the 16PF (form C) to determine if its Motivational Distortion (MD) scale and associated correction procedures were appropriate to use with welfare recipients in mandatory employment-related evaluations. It was assumed that these participants had real motivation to present themselves in a socially desirable way, as their continued reception of welfare benefits was dependent on their participation in the welfare-to-work program. These researchers found that in general, the established detection and correction procedures were appropriate for use with the welfare population. However, the established procedures did not account for the observed relationship between high MD scores and lower scores on the Dominance (E) Primary factor.

This finding was consistent with the results of an earlier study by Wallbrown, Reuter, and Barnett (1989) examining the use of the MD scale with a sample of felons taking the 16PF during sentence-review. Taken together, these findings indicate that perhaps the applied settings used in the studies above tap into different motivations than the simulated faking situations traditionally used to validate distortion scales, thus resulting in different response styles between situations.

Such evidence begins to cast doubt on the appropriateness of studying motivational distortion using only experimentally simulated groups of faking and honest respondents. Zickar and Robie (1999) noted the problems inherent in trying to compare the type of faking found in simulated studies to that found in real-world situations. These researchers compared simulated faking and honest groups of military recruits using both validity scales and Item Response Theory (IRT). Although they found significant differences between faking and honest groups, especially with IRT methods, the researchers cautioned that the instructed faking they studied was likely much less subtle than faking in real-world situations.

All of these findings regarding differences in experimentally induced and real-world faking lend support to Cattell's (1968, 1986) trait-view theory. In this theory, Cattell proposes that any distortion behavior that results in anything other than a true representation of the individual stems from both the personality of the individual and the given situation. In addition, he posits that distortions are multi-dimensional in nature. The latter position is elaborated and delineated in Paulhus and Reid's (1991) description of distortions being comprised of both unconscious self-deception and intentional, deliberate impression management. However, it is

Cattell's (1968, 1986) description of the role of the situation in determining distortion that is most relevant.

In his 1986 writing, Cattell notes that there are traditionally thought to be two types of error in psychological testing: measurement errors and statistical inference errors. However, he describes a third error, which he deems to be more serious in nature than either of the first two. This error stems from "...the interpretation and use of the given numerical score" (Cattell, 1986, p.497); in essence, the incorrect interpretation of test data. This error refers to the failure to see all the complexities that lead to a test-taker responding in a given way. One must consider, therefore, every component, including the situation in which the test is taken, in order to accurately describe the response distortion present. In terms of socially desirable distortions, Cattell asserts that what is accepted as a "desirable" attribute can change across levels of specificity and situations. In the broadest sense, there are attributes desired by society as a whole. Then there are attributes desired by specific peer or social groups. Finally, Cattell asserts that there are attributes desirable for, and unique to every possible situation. With regard to personality assessment, there are different types and amounts of response distortion unique to different testing situations and purposes.

Cattell's (1968, 1986) trait-view theory, then, has serious implications for the study of deliberate distortion in personality assessment. If situational variables do indeed significantly alter the motives of the test-taker, and thereby the distortion tendencies of the test-taker as Cattell suggests, then serious doubt is cast on the external validity of the traditional simulated faking experimental design. The differences in situation between contrived simulations and applied settings are easy to see and undeniable. The question then becomes one regarding differences in

real-world situations. Does applying for a job elicit different motives for distortion than applying for parole? These situations are considerably more difficult to tease apart. Finally, when the implications of Cattell's theory are taken to their logical conclusion, it seems to suggest that different norms are needed for every unique situation in order to obtain a truly accurate measure of the motivational distortion present. In fact, Cattell (1968) has suggested that situational norms should be developed for at least the most commonly encountered test-taking situations (i.e., job application, correctional settings, etc.).

The results of studies examining faking in applied settings (e.g., Merydith et al., 1997; Wallbrown et al., 1989) have provided some support for Cattell's (1968, 1986) theory. In addition, a recent study by Stark, Chernyshenko, Chan, Lee, and Drasgow (2001) appears to lend further support. This study used Item Response Theory (IRT) methods to examine the effects of varied testing situations on the items and scales of the 16PF. Stark and his colleagues noted that attempts to study faking in applied settings have generally taken two forms. One method involves comparing groups of job-applicants with groups of non-applicants to ascertain situational differences, while the second method examines dispositional components of faking by comparing groups determined to be faking or honest based on their scores on some measure of distortion, such as the IM scale.

Stark et al. employed both methods by using data from Institute of Personality and Ability Testing from independent groups of job-applicants and non-applicants. Situational comparisons were made between the two groups, while dispositional components were examined by using IM scores to differentiate fakers and honest responders within each group. It was found that the IM scale functioned differently across situations, leading Stark and colleagues to

conclude that future research should not generalize results across applied situations. Instead, they suggest that separate groups of applicants and non-applicants be compared to study the effects of faking in each individual situation. These findings are congruent with Cattell's (1968, 1986) assertion that different situations elicit different styles of faking. However, Stark et al.'s (2001) study, while comparing applicants and non-applicants, used data taken from a wide variety of testing situations. The applicant sample was made up of data from a diverse range of employment situations, while the non-applicant sample used data from counseling, experimental, and developmental settings. However, no additional information is available about the composition of either group. It seems possible then, that there may be significant differences in response distortion between smaller, situationally-defined groups. Recalling Cattell's (1986) assertions, motivation to respond in a socially desirable manner exists at both broad levels, such as one's culture/society, and at very specific levels, as determined by the particular situation at hand. Thus, there is a need to evaluate motivational distortion in specific applied settings as compared to experimentally contrived situations.

The purpose of the present study is to contrast motivational distortion effects in simulated situations to that of specific applied settings. More specifically, this study intends to contrast 16PF profiles and IM scale scores from two applied settings to those from one experimental condition where participants are asked to fake good. This will add to the growing body of research on the effects of situational differences on elicited response distortions, and in turn, examine the appropriateness of using simulation studies to develop and validate distortion detection and correction techniques. In addition, this study will provide results which can be

applied in evaluation of Cattell's (1968, 1986) trait-theory approach to personality and its assessment.

Method

Participants

This study used archival data from two primary groups of participants: 28 school psychology graduate students currently enrolled in a specialist-level program at a large northeastern university, and 38 applicants to the same graduate program. Of the participants, 10 were male, and 56 were female.

Instrument

The 5th edition of the 16 Personality Factor questionnaire (16PF) (Russel & Karol, 1993) was used as the personality measure for this study. The 16PF is a standardized, norm-referenced measure of normal personality that provides a results profile based on 16 primary factors and 5 global factors (Conn & Reike, 1994). The 16 primary factors, warmth (A), reasoning (B), emotional stability (C), dominance (E), liveliness (F), rule-consciousness (G), social boldness (H), sensitivity (I), vigilance (L), abstractedness (M), privateness (N), apprehension (O), openness to change (Q1), self-reliance (Q2), perfectionism (Q3), and tension (Q4), were developed by Raymond B. Cattell (1945, as cited in Conn & Rieke, 1994) through an extensive factor analysis of common descriptors of human personality. The five global factors, extraversion, anxiety, tough-mindedness, independence and self-control resulted from intercorrelations among the primary factors.

The 16PF has been empirically shown to be a reliable and valid instrument for evaluating personality (Conn & Rieke, 1994). Normative data for the 5th edition was obtained from a sample of 2,500 people, with demographics matched to the 1990 U.S. Census. Temporal stability, as established by two-week test-retest methods, is reported to range from .69-.87 for the 16 primary

scales and from .84-.91 for the five global scales using a two-week interval. Social Boldness (H), Self-Reliance (Q2) are the most reliable primary factors, with reliabilities exceeding .85, while Reasoning (B) and Emotional Stability (C) are the least reliable, with reliability estimates of .69 and .75 respectively. Among the global factors, Extraversion is the most reliable, with a reliability of .91, while the Anxiety and Independence factors are the least reliable at .84. The Tough-Mindedness and Self Control factors have temporal stabilities of .87. In addition, the manual reports internal consistency coefficients (Cronbach's alpha) ranging from .68-.87 for the 16 primary factors.

Procedure

The archival data from participants was separated into three groups for comparison. The current graduate students completed the 16PF twice as an assignment for a course in personality assessment: once as a course exercise where they scored their own 16PF profiles and learned how to interpret them, and once responding as though they were applying for admission into the program and wanted to appear as an ideal candidate (essentially, faking good). These two conditions comprise two of the three experimental groups. In the Classroom condition, participants were read the conventional 16PF directions. In the latter fake-good Analog condition, participants were read the following script:

“For this session, disregard the instructions in your test booklet asking you to respond honestly and in the way which best describes you. Instead, imagine that you are applying for an internship or your first job as a school psychologist and that this test is part of the interview process. You want this position very much, and you know this prospective employer is looking for only those individuals who are best suited to the demands of the

position. Respond to the questions so as to make yourself look as good as possible to the interviewers. Remember, you want to appear as an ideal candidate for a school psychology position that you really want to get.”

The purpose of this class assignment is to provide students with an understanding of the use and interpretation of personality measures as well as the potential threats to validity associated with such measures.

The third condition used data from applicants to the program taking the 16PF as part of the standard admissions process; an applied situation assumed to produce a high motivation to fake one’s responses for the better. Thus, three groups will be formed: students in a neutral self-evaluative condition (Classroom), students positively faking responses in an analog condition (Analog), and applicants with high motivation to positively fake responses in an applied condition (Applicant).

Data Analysis

Data obtained from participants was first used to calculate means and standard deviations for each of the three groups (Classroom, Analog, and Applicant). Following this descriptive analysis, mean differences between groups were evaluated using three sets of t-tests. Comparisons between the Classroom and Applicant groups, and the Analog and Applicant groups, were conducted using independent measures t-tests, while comparisons between Classroom and Analog groups were made using dependent measures t-tests.

Results

Means and standard deviations from the descriptive analyses are displayed in Table 1 and Table 2 for Primary and Global factors, respectively. Primary factor mean differences from all three comparisons are shown in Table 3, and mean differences in Global factor scores can be found in Table 4.

The results of t-tests comparing the Classroom and Applicant groups yielded significant differences on four Primary factors: C: Emotional Stability ($t_{64} = 3.28, p < .002$), E: Dominance ($t_{64} = 3.95, p < .002$), F: Liveliness ($t_{64} = 4.36, p < .002$), and H: Social Boldness ($t_{64} = 4.03, p < .002$). Significant differences were also found on the Extraversion ($t_{64} = 4.07, p < .002$) and Independence ($t_{64} = 4.73, p < .002$) Global factors.

Results of t-tests conducted between the Analog and Applicant groups showed significant differences on nine of the Primary factors: C: Emotional Stability ($t_{64} = 7.60, p < .002$), F: Liveliness ($t_{64} = -3.69, p < .002$), G: Rule-Consciousness ($t_{64} = 6.83, p < .002$), H: Social Boldness ($t_{64} = 3.93, p < .002$), L: Vigilance ($t_{64} = -3.50, p < .002$), O: Apprehension ($t_{64} = -8.04, p < .002$), Q₂: Self-Reliance ($t_{64} = -3.99, p < .002$), and Q₄: Tension ($t_{57.19} = -14.13, p < .002$). The groups also differed significantly on the Impression Management (IM) scale ($t_{41.60} = 8.84, p < .002$) and two of the Global factors: Anxiety ($t_{64} = -12.17, p < .002$) and Self-Control ($t_{64} = 5.78, p < .002$).

The dependent measures comparison between the Classroom and Analog groups revealed significant differences on several factors. Ten Primary scales differed significantly: A: Warmth ($t_{27} = -4.28, p < .002$), C: Emotional Stability ($t_{27} = -8.82, p < .002$), G: Rule-Consciousness ($t_{27} = -13.62, p < .002$), H: Social Boldness ($t_{27} = -7.61, p < .002$), L: Vigilance ($t_{27} = 4.76, p < .002$), O: Apprehension ($t_{27} = 10.48, p < .002$), Q₁: Openness to Change ($t_{27} = -3.47, p < .002$), Q₂: Self-Reliance ($t_{27} = 6.50, p < .002$), Q₃: Perfectionism ($t_{27} = -6.51, p < .002$), and Q₄: Tension ($t_{27} =$

11.88, $p < .002$). Significant differences were also found on the Impression Management scale ($t_{25} = -9.86$, $p < .002$), and the Extraversion ($t_{27} = -5.79$, $p < .002$), Anxiety ($t_{27} = -15.29$, $p < .002$), Tough-Mindedness ($t_{27} = 3.75$, $p < .002$), Independence ($t_{27} = -5.02$, $p < .002$), and Self-Control ($t_{27} = -9.37$, $p < .002$) Global factors.

Discussion

The results of the present analyses give rise to some interesting findings. As Table 3 indicates, the two groups with the lowest number of significant differences ($p < .002$) between them across Primary and Global factors were the Honest Classroom and Applicant groups. Only four Primary (C, E, F, and H) and two Global (EX and IN) factors showed significant differences between the two experimental groups. In addition, the mean IM scales of the two groups were not found to be significantly different. By contrast, nine Primary (C, F, G, H, L, O, Q2, Q3, and Q4) and two Global factors (AX and SC) differed significantly between the Analog Classroom and Applicant groups. The IM scales of the two groups were also significantly different. The comparisons between the Honest Classroom and Analog Classroom groups yielded the highest number of significant differences: ten Primary factors (A, C, G, H, L, O, Q1, Q2, Q3, and Q4), all five Global factors (EX, AX, TM, IN, and SC), and the IM score. These findings seem to suggest that the profiles of the Honest Classroom and Applicant groups are relatively similar, while neither group's profiles closely approximate those obtained from the Analog Classroom group. This basic pattern of similarities and differences among the three experimental groups was not entirely unexpected. One could likely safely assume that the Honest and Analog Classroom groups would differ substantially, especially given the specific directions used with each respective group. Further, if one has any faith at all in the usefulness of personality measures in applied settings, one would also likely assume that the Applicant group's profiles would be more similar to honest test-takers' than they would be to test-takers' specifically instructed to fake their responses. However, inspection beyond solely statistically significant differences reveals another overall trend between the groups warranting further inspection.

Looking at the mean scale scores for the three experimental groups in Tables 1 and 2, a consistent pattern of rank-ordering among the groups emerges. Specifically, assuming that the scores of the Analog Classroom group represent the ideal, or most desirable, characteristics for school psychology graduate students, the mean scale scores of the groups tend to move from least desirable to most desirable in the following pattern: Honest Classroom: least desirable, Applicant: more desirable, Analog Classroom: most desirable. This pattern is further elaborated visually in Figure 1.

Before continuing the exploration of this trend, it should be noted that “ideal” scores in this case, are not necessarily high scores. The factors of the 16PF, and the constructs they represent, are bipolar, with the desirable end of the continuum defined by the given situation. For example, in the case at hand, higher scores on the Warmth (A) factor were found to be more desirable (as determined by the Analog Classroom group), while on the Tension (Q4) factor, lower scores were more desirable.

While the differences between the groups are not all statistically significant, particularly between the Honest Classroom and Applicant groups, the consistency of the pattern across the Primary and Global factors makes it worthy of consideration. Referring once again to Tables 1 and 2 and Figure 1, the results of the present study show that the obtained mean scores of 12 of the 16 Primary factors, and all five of the Global factors follow this general trend. The factors that do not fit the pattern are Reasoning (B), Dominance (E), Liveliness (F), and Abstractedness (M).

In the case of the Reasoning factor, in addition to not following the pattern, none of the three groups' mean scores differed significantly from one another. This is likely due to the problem-solving nature and content of the Reasoning scale. It is the only component of the 16PF

with pre-determined “correct” answers, making it difficult to “fake” in a positive way. Further, the Reasoning factor is one of the few 16PF Primary factors that is not significantly correlated with the Impression Management (IM) scale (Conn & Reike, 1994), suggesting that when individuals distort their responses on the 16PF to attempt to appear more desirable, the Reasoning factor remains largely unaffected. This latter point is also true with regard to the Dominance factor, but past research suggests that there may be more to consider when searching for the reasons behind the Dominance factor’s (and other 16PF factors lacking the “right/wrong” format of the Reasoning factor’s test items) non-compliance with the observed overall pattern.

Recalling the findings of past researchers discussed in the introductory section of the present study, particularly those of Wallbrown, Reuter, and Barnett (1989) and Merydith, Humphreys, and Ebener (1997), it has been found that in some applied settings, the Primary factors correlated with measures of motivational distortion do not necessarily exactly match the correlations found in the normative sample of the test authors. For example, Merydith and colleagues (1997) found that when working with a sample of Welfare recipients, the Dominance (E) factor of the 16PF form C was correlated with the Motivational Distortion (MD) scale, but that the instrument lacked any correction procedures to account for distortion on that particular factor. In sum, it seems that the lack of a significant correlation between elevated IM scores and any given Primary or Global factor in the normative sample of the 16PF is not necessarily a good predictor of such correlations in applied settings. Therefore, alternate reasoning is required to account for the failure of the Dominance, Liveliness, and Abstractedness factors to conform to the overall pattern at the heart of this discussion. However, that alternate line of reasoning may be closely intertwined with the failure of IM scale correlations alone to create a feasible explanation.

The findings of studies such as Merydith et al. (1997) lend support to Cattell's (1968, 1986) claims regarding the effects of varying testing situations on response styles. Cattell asserted that different testing situations carried with them different styles of faking necessary to appear desirable or undesirable in that particular situation. Thus, it seems that a more tenable explanation for the Dominance, Liveliness, and Abstractedness factors' deviation from the overall pattern may be that these factors are incorporated differently into the faking styles elicited by the situations presented by the present study. Specifically, it seems possible that the applicant situation elicited a stronger desire to appear assertive, proactive, and energetic (Dominance and Liveliness) than the analog situation. Therefore perhaps being in a real-life applicant situation draws out these desirable "candidate" traits, whereas being in an analog situation causes test-takers to concentrate more on traits desirable for actually *being* a school psychologist. Further, it also seems possible that test takers were unclear regarding the role of the Abstractedness factor as it applied to the situations they found themselves in, thus leading to largely neutral responding, and accounting for the lack of significant differences between groups on this factor.

With the exceptions to the observed overall pattern of increasing desirability examined, the present discussion can now focus directly on the pattern itself. Once again, the majority of the 16PF Primary and Global factor scores follow a general pattern of increasing desirability from the Honest Classroom group, to the Applicant group, and finally to the Analog Classroom group. In addition, based on the magnitudes of the differences between the groups, it appears that the Honest Classroom group's scores were the least desirable, the Applicant group's scores were slightly more desirable than those of the Honest Classroom group, and the Analog Classroom group's scores were considerably more desirable than those of the Applicant group.

To explain the significance of this pattern of differences, one must recall the existing theories on response distortion described in previous sections of this work.

Cattell (1968) and Paulhus and Reid (1991) described response distortions as stemming from two primary sources, thus giving rise to two basic forms of distortion. The first, self-deception, is theorized to be the result of an unconscious reaction to the test-taking situation. Impression management, the second form of distortion, consists of an intentional attempt to make oneself appear a particular way. In accordance with these existing theories, it is proposed here that the pattern of increasingly desirable scores found in the present study provides an illustration of the existence of, and the differences between, two separate types of response distortion that may closely approximate the preceding descriptions of self-deception and impression management. Specifically, using the Honest Classroom group's scores as a baseline, if the significantly elevated scores obtained from the Analog Classroom group are the result of impression management (since the instructions given to that group were consistent with the definition of impression management), the less significant, but consistent-across-groups elevation of the Applicant group's scores may very well stem from the unconscious response distortion that is characteristic of self-deception. At the very least, it seems probable that the differences between the Honest Classroom and Applicant groups and the differences between the Honest Classroom and Analog Classroom groups rise from sources, or types/styles of faking, that are qualitatively and/or quantitatively different. This suggestion is consistent with the findings of Zickar and Robie (1999), who noted that faking in applied, real-world settings, was likely much more subtle than the instructed faking examined by their, and other analog experimental designs. In addition, these findings lend support to Cattell's (1968,1986) theory

that different testing situations assert different effects on the types of response distortions present in each unique situation.

Based on the findings of the present study, it appears that the practice of using analog research designs to evaluate response distortion in self-report personality assessment may be prone to Cattell's "Third Error Source" (Cattell, 1986). That is, the analog paradigm seems to be failing to take into account all the complexities involved in response distortion and thereby misinterpreting the overwhelmingly positive findings regarding the validity of the distortion detection measures in question. In particular, the results of the present study indicate that not only are there likely situational factors that need to be accounted for, as Cattell and others have suggested, there also appear to be issues regarding the subtlety and source of distortions that warrant a significant degree of concern. By implication then, it seems that researchers evaluating the validity of response distortion detection measures need to take some degree of caution when interpreting the results of analog studies. Further, there may be cause to develop more sensitive techniques for evaluating both response distortion and the measures designed to detect it.

Another facet of issue surrounding response distortion and its detection is what precisely to do about it when it is found. When blatant impression management/faking is detected, it seems prudent to throw out the results, or at least interpret them with extreme caution. The numerous significant differences between the classroom and analog groups in this study show the disparities between the profiles of those who fake good and those who (presumably) do not. However, the consistent, but often not statistically significant differences between the applicant group and classroom groups pose a less obvious situation. While this study has provided evidence suggesting that some response distortion occurs undetected in applied situations, the question eventually becomes one of significance and meaning. That is, does the less intense

distortion found in the applicant group necessitate correction during profile interpretation? The scores obtained in the present study seem to indicate that by and large, the shape of the profile remains the same, as if a constant value was added/subtracted across factors. If the shape of the profile, or the magnitude of the factor scores relative to each other, remains the same, it seems that the obtained profile would still be highly usable, or even statistically the same, as the presumably honest profile. Further, if the slight elevations in factor scores for applicants are as consistent as the results of the present study suggest, perhaps it is more meaningful when an applicant profile lacks this mild elevation in scores. What does it say about an individual when his or her personality does not react to the situational variables present in a job/program application situation? These are questions that could be addressed in further extension and replication of this line of research.

In addition, the present study is limited by a relatively small sample size and the use of only one measure of personality in a single applied setting. Replication of this study/design with larger sample sizes is needed to further establish the credibility of the present findings. Future research should include multiple personality instruments from both normal and abnormal (i.e., MMPI-2) domains, and should investigate the stability of these findings across various applied situations.

References

- Austin, J.S. (1992). The detection of fake good and fake bad on the MMPI-2. *Educational & Psychological Measurement, 52*, 669-674.
- Baer, R.A., & Miller, J. (2002). Underreporting of psychopathology on the MMPI-2: A meta-analytic review. *Psychological Assessment, 14*(1), 16-26.
- Bagby, R.M., Buis, T., & Nicholson, R.A. (1995). Relative effectiveness of the standard validity scales in detecting fake-bad and fake-good responding: replication and extension. *Psychological Assessment, 7*, 84-92.
- Burger, J.M. (2000). *Personality*. Belmont, CA: Wadsworth/Thompson Learning.
- Butcher, J.N., Dahlstrom, W.G., Graham, J.R., Tellegen, A., & Kaemmer, B. (1989). *Minnesota Multiphasic Personality Inventory – 2 (MMPI-2): Manual for Administration and Scoring*. Minneapolis, MN: University of Minnesota Press.
- Cassisi, J.E., & Workman, D.E. (1992). The detection of malingering and deception with a short form of the MMPI-2 based on the L, F, and K scales. *Journal of Clinical Psychology, 48*(1), 54-58.
- Cattell, R.B. (1968). Trait-view theory of perturbations in ratings and self ratings: its application to obtaining pure trait score estimates in questionnaires. *Psychological Review, 75*(2), 96-113.
- Cattell, R.B. (1986). Dodging the third error source: psychological interpretation and use of given scores. In R.B. Cattell & R. Johnson (Eds.), *Functional psychological testing: principles and instruments* (pp. 496-543). New York: Brunner/Mazel.
- Conn, S.R., & Rieke, M.L. (Eds.). (1994). *16PF Fifth Edition Technical Manual*. Champaign, IL: Institute for Personality and Ability Testing, Inc.

- Greene, R.L. (1991). *The MMPI-2/MMPI an interpretive manual*. Boston: Allyn & Bacon.
- Grow, R., McVaugh, W., & Eno, T.D. (1980). Faking and the MMPI. *Journal of Clinical Psychology, 36*, 910-917.
- Lees-Haley, P.R. (1997). MMPI-2 base rates for 492 personal injury plaintiffs: Implications and challenges for forensic assessment. *Journal of Clinical Psychology, 53*(7), 745-755.
- Meehl, P.E. & Hathaway, S.R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology, 30*, 525-564.
- Merydith, S.P., Humphreys, J.K., & Ebener, D.J. (1997). Motivational distortion of the 16PF by welfare recipients. *Journal of Personality Assessment, 69*(2), 376-391.
- Merydith, S.P. & Wallbrown, F.H. (1991). Reconsidering response sets, test-taking attitudes, dissimulation, self-deception, and social desirability. *Psychological Reports, 69*, 891-905.
- Paulhus, D.L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*(3), 598-609.
- Paulhus, D.L. & Reid, D.B. (1991). Enhancement and denial in socially desirable responding. *Journal of Personality and Social Psychology, 60*(2), 307-317.
- Russel, M. & Karol, D. (1993). *16PF fifth edition administrators manual*. Campaign, IL: Institute for Personality and Ability Testing.
- Scheuriger, J.M. (1992). The Sixteen Personality Factor Questionnaire and its junior versions. *Journal of Counseling and Development, 71*, 231-244.
- Stark, S., Chernyshenko, O.S., Chan, K-Y., Lee, W.C., & Drasgow, F. (2001). Effects of the testing situation on item responding: cause for concern. *Journal of Applied Psychology, 86*(5), 943-953.

- Storm, J. & Graham, J.R. (2000). Detection of coached general malingering on the MMPI-2. *Psychological Assessment, 12*(2), 158-165.
- Wallbrown, F.H., Reuter, E.K., & Barnett, R.W. (1989). Motivational distortion on the 16PF primaries by male felons. *Measurement and Evaluation in Counseling and Development, 22*, 7-14.
- Wetter, M.W., Baer, R.A., Berry, D.T.R., Smith, G.T. & Larsen, L.H. (1992). Sensitivity of MMPI-2 validity scales to random responding and malingering. *Psychological Assessment, 4*(3), 369-374.
- Zickar, M.J. & Robie, C. (1999). Modeling faking good on personality items: and item-level analysis. *Journal of Applied Psychology, 84*(4), 551-563.

Table 1

Primary Factor Means and Standard Deviations Across Conditions

Factor/Scale	Classroom (n = 22)		Analog (n = 22)		Applicant (n = 38)	
	M	SD	M	SD	M	SD
A: Warmth	5.89	1.52	7.50	1.00	6.79	1.30
B: Reasoning	6.46	1.45	6.75	1.32	6.26	1.22
C: Emotional Stability	5.14	1.72	8.00	0.00	6.37	1.32
E: Dominance	3.93	1.25	5.00	1.25	5.11	1.16
F: Liveliness	5.43	1.29	5.82	1.12	6.89	1.39
G: Rule-Consciousness	4.39	1.13	7.93	1.46	5.42	1.48
H: Social Boldness	4.61	1.69	7.93	1.25	6.37	1.81
I: Sensitivity	5.36	2.31	6.32	1.49	5.84	1.70
L: Vigilance	4.71	1.38	3.32	1.12	4.66	1.78
M: Abstractedness	4.68	1.72	4.57	0.74	5.00	1.59
N: Privateness	4.46	1.86	4.00	1.31	4.13	1.70
O: Apprehension	5.79	1.23	2.79	1.23	5.71	1.61
Q1: Openness to Change	5.89	1.73	7.18	1.47	6.58	1.55
Q2: Self-Reliance	4.89	1.50	2.93	1.12	4.24	1.44
Q3: Perfectionism	4.79	1.55	7.46	1.17	6.08	1.85
Q4: Tension	5.36	1.47	2.18	0.48	4.71	0.96
IM: Impression Management	44.19	27.08	96.14	5.30	60.97	23.78

Table 2

Global Factor Means and Standard Deviations Across Conditions

Factor	Classroom (n = 22)		Analog (n = 22)		Applicant (n = 38)	
	M	SD	M	SD	M	SD
EX: Extraversion	5.91	1.45	7.90	1.05	7.27	1.26
AX: Anxiety	5.44	1.44	1.38	0.73	4.62	1.26
TM: Tough-Mindedness	5.56	1.57	4.18	1.31	4.73	1.42
IN: Independence	4.21	1.32	6.00	1.26	5.68	1.18
SC: Self-Control	5.02	1.20	7.42	1.02	5.52	1.50

Table 3

Mean Differences Between Comparison Groups Across Primary Factors

Factor/Scale	Classroom – applicant (n = 60)	Analog– applicant (n = 60)	Classroom – analog (n = 22)
A: Warmth	0.90	0.71	1.61*
B: Reasoning	-0.20	0.49	0.29
C: Emotional Stability	1.23*	1.63*	2.86*
E: Dominance	1.18*	-0.11	1.07
F: Liveliness	1.47*	-1.07*	0.39
G: Rule-Consciousness	1.03	2.51*	3.54*
H: Social Boldness	1.76*	1.56*	3.32*
I: Sensitivity	0.48	0.48	0.96
L: Vigilance	-0.06	-1.34*	-1.39*
M: Abstractedness	0.32	-0.43	-0.11
N: Privatness	-0.33	-0.13	-0.46
O: Apprehension	-0.08	-2.92*	-3.00*
Q1: Openness to Change	0.69	0.60	1.29*
Q2: Self-Reliance	-0.66	-1.31*	-1.96*
Q3: Perfectionism	1.29	1.39*	2.67*
Q4: Tension	-0.65	-2.53*	-3.18*
IM: Impression Management	16.75	35.17*	51.85*

* $p < .002$

Table 4

Mean Differences Between Comparison Groups Across Global Factors

Factor	Classroom – applicant (n = 60)	Analog – applicant (n = 60)	Classroom – analog (n = 22)
EX: Extraversion	1.36*	0.63	1.989*
AX: Anxiety	-0.82	-3.24*	-4.06*
TM: Tough-Mindedness	-0.83	-0.55	-1.37*
IN: Independence	1.47*	0.32	1.78*
SC: Self-Control	0.50	1.90*	2.40*

* p<.002

Figure Caption

Figure 1. Mean primary and global factor scores across groups.

Mean Sten Score

