

LIBRARY

Michigan State

University

PLACE IN RETURN BOX to remove this checkout from your record.
TO AVOID FINES return on or before date due.
MAY BE RECALLED with earlier due date if requested.

DATE DUE	DATE DUE	DATE DUE

SCORE INFLATION ON BIODATA AND SITUATIONAL JUDGMENT
INVENTORY ITEMS: A COLLEGE ADMISSIONS QUANDARY

By

Lauren Jill Ramsay

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF ARTS

Department of Psychology

2003

ABSTRACT

SCORE INFLATION ON BIODATA AND SITUATIONAL JUDGMENT INVENTORY ITEMS: A COLLEGE ADMISSIONS QUANDARY

By

Lauren Jill Ramsay

Tests used in a selection context must be carefully examined to ensure that if they are susceptible to inflation, that vulnerability does not affect their utility. Biodata and situational judgment measures could be used to support college admissions decisions. This study explores the relationship between score inflation on these two tests, and situational factors influencing performance: motivation to perform well, coaching on how to perform well, and a warning statement not to respond dishonestly. Both motivation and coaching were found to predict performance. Conscientiousness and emotional stability were associated with social desirability measured by self deception and impression management, and extreme inflation appeared to be driven more by situation than by personality. Item-level analyses showed that biodata items that did not require elaboration were more susceptible to inflation, as were items that were less objective, less controllable, and more college-relevant. Criterion-related validity was not negatively affected by inflation in this study, and corrections to selection decisions based on three indices of inflation (bogus items, an inflation index, and high scores on impression management) did not result in any change in the quality of candidates selected, in terms of performance outcomes. Implications of the findings are discussed.

ACKNOWLEDGEMENTS

This research was conducted with support from the College Board, New York.

TABLE OF CONTENTS

LIST OF TABLES.....	vi
LIST OF FIGURES	viii
INTRODUCTION	1
LITERATURE REVIEW	5
Definitional Issues	5
Biodata	8
Situational Judgment Inventory	12
Coaching	13
Individual Differences in Faking	19
Effects on Validity	21
Inflation Controls	25
Warning	26
Composition of Items	27
Statistical Control using Social Desirability or Bogus Items	30
Study Development.....	34
METHOD	45
Sample.....	45
Study Design.....	46
Procedure	47
Manipulations	49
Measures	50
RESULTS	57
Situational Differences.....	57
Individual Differences	67
Item Differences.....	73
Validity and Selection Decisions	82
CONCLUSION.....	91
Discussion.....	91
Limitations	97
Practical Implications.....	98

APPENDICES	104
Appendix A – Mael’s Taxonomy	104
Appendix B – Sample Questionnaire.....	106
Appendix C – Sample Biodata Items.....	120
Appendix D – Sample Situational Judgment Items	122
Appendix E – Wording of Instruction Sets.....	124
Appendix F – Informed Consent Form – Motivated Group	129
Appendix G – Informed Consent Form – Not Motivated Group.....	132
Appendix H – Faking Study Protocol.....	135
Appendix I – Coaching Directions	141
Appendix J – Coaching Handout	143
Appendix K – College GPA Release Form	156
Appendix L – High School GPA and SAT/ACT Score Release Form	158
Appendix M – Twelve Dimensions of College Performance.....	160
Appendix N – Bogus Biodata Items	164
Appendix O – Inflation Index Items	166
Appendix P – Calculation for Inflation Index Selection.....	169
Appendix Q – Biodata Item Types Coding Instructions.....	172
Appendix R – SJI Item Types Coding Instructions	174
REFERENCES	176

LIST OF TABLES

Table 1 – Sample Elaborated Biodata Item	29
Table 2 – Expected Score Levels	39
Table 3 – Study Manipulations and Participants per Cell.....	46
Table 4 – Coefficient Alpha and Descriptives for Ratings of Biodata Item Characteristics.....	55
Table 5 – Coefficient Alpha and Descriptives for Ratings of SJI Item Characteristics.....	56
Table 6 – Analysis of Variance Results for Biodata with and without the Inclusion of Covariates	59
Table 7 – Means and Standard Deviations of Biodata across Conditions	61
Table 8 – Analysis of Variance for SJI with and without Covariates.....	63
Table 9 – Means and Standard Deviations of SJI Responses for Various Study Conditions.....	64
Table 10 – Correlation Matrix	69
Table 11 – Overall Means and Standard Deviations of Four Item Characteristics Fisher z for Biodata.....	73
Table 12 – Analysis of Variance Results for Fisher z for Four Item Characteristics for Biodata	74
Table 13 – Means and Standard Deviations of Four Item Characteristic Fisher z for Various Study Conditions for Biodata.....	75
Table 14 – Overall Means and Standard Deviations of Four Item Characteristics Fisher z for SJI	77
Table 15 – Analysis of Variance Results for Fisher z for Four Item Characteristics for SJI.....	78
Table 16 – Means and Standard Deviations of Four Item Characteristic Fisher z for Various Study Conditions for SJI	79

Table 17 – Correlations and Descriptive Statistics for Predictors and Criteria82

Table 18 – Zero Order and Partial Correlations between Situational Judgment and Biodata and Two Criteria Controlling for Measures of Faking.....83

Table 19 – Coefficient Alpha of Biodata and SJI Scales for Various Study Conditions84

Table 20 – Adjusted *r* for Biodata and SJI Scales in Predicting GPA and Absenteeism Across Various Study Conditions.....85

Table 21 – Descriptives for High Performance and Reference Groups.....87

Table 22 – Descriptive Statistics for Selection Ratio of .10.....89

Table 23 – Descriptive Statistics for Selection Ratio of .25.....89

Table 24 – Descriptive Statistics for Selection Ratio of .50.....90

LIST OF FIGURES

Figure 1 – Conceptual Model	36
Figure 2 – Expected Interaction of Warning, Coaching, and Motivation on Test Performance	40
Figure 3 – Interactions between Coaching, Motivation, and Warning for Biodata Performance	66

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z

INTRODUCTION

Colleges have historically considered traditional academic measures, such as high school GPA and SAT/ACT scores, in predicting academic performance in college, which informs admission decisions. However, there is more to the college experience than academic performance alone, and as colleges are pressed to measure student performance along broader criteria, they need to consider a similarly broad range of predictors. Biodata questions and situational judgment inventories have shown promise as two such examples of predictors of college student performance (Oswald, Schmitt, Ramsay, Kim, & Gillespie, in press), and in general, the prospect of using a more inclusive range of instruments is attractive for gathering a wealth of information on applicants, given the aim of increasing the diversity of applicants accepted to college. Nevertheless, such measures require thorough development and validation efforts because applicants to college may be motivated to distort their responses (e.g., “fake good”) and inflate their scores on these tests. *Faking* can be defined as a conscious effort to manipulate responses to make a positive impression (see Zickar & Robie, 1999), and people can and do fake on certain tests when motivated to do so (McFarland & Ryan, 2000; Viswesvaran & Ones, 1999). It is reasonable to assume that applicants may manipulate their responses to a test in a way that enhances the likelihood of their success in a selection process (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990), and such manipulation may affect the utility of the tests (Rosse, Stecher, Miller, & Levin, 1998). This manipulation may be motivated by a desire to appear socially desirable or job desirable (Ones, Viswesvaran, & Reiss, 1996). As Corr and Gray (1994, p. 433) succinctly put it, “to some extent it is *de rigueur* for successful job applicants/incumbents to engage in some form of impression

management.” Nevertheless, some inflation goes beyond low-key impression management, and may include outright lying and deliberately providing false information. As it is difficult to establish the boundaries of what is honest and what is dishonest responding, I will use the term *inflation*, rather than *faking*, unless I am referring to literature that specifically uses the term faking. I seek to understand the extent to which individual difference characteristics affect test responses, how responses might be distorted under different situational constraints, and how both these factors affect the utility of biodata and situational judgment inventories designed for college admissions.

While we know that individuals are capable of inflation on some items on biodata tests and situational judgment inventories, we do not know enough about situational factors that contribute to inflation, or methods of identifying it and limiting its effects. If eventually biodata and situational judgment measures might be used to contribute to the information collected by colleges in making admissions decisions and in student development contexts (Oswald, et al., in press), then applicants will be motivated to achieve high test scores, and many will likely demand coaching programs to help prepare for the test. In other words, if there is a move to supplement ability measures with the noncognitive measures considered in this paper in making college admissions decisions, the evaluation process may become more vulnerable to coaching and other forms of manipulation.

Some researchers regard faking studies to be of little value, however, when conducted outside a real selection experience. I argue that there are important reasons to examine faking in tests to be used in a selection context, before they are actually used for

decision-making. Inflation and coaching may be problematic in some contexts, and considering the motivation to perform well on selection tests in a college-admissions context, these issues must be addressed when considering tests such as biodata and situational judgment inventories for college admissions. Past research makes a similar recommendation: With the development of the Assessment of Background and Life Experiences (ABLE), a test containing personality and biodata content, that test was not regarded as ready for implementation until research had been conducted on the effects of coaching and faking (White, Young, & Rumsey, 2001). In the present context, without studying the causes, contexts, and outcomes of test-takers' manipulating their responses to these new tests, and instead simply implementing a new test in a college admissions context, universities would be opening themselves up to intense criticism and probable litigation for using a test the appropriateness of which has not yet been supported. It is essential that test users understand the extent to which a test's validity may be affected by coaching or inflation, and how warning statements, biodata elaboration requirements and correction scales might be useful if such tests were used in a college admissions context. To that end, this study tests a sample of freshman students, most of whom are in their first semester of college. With this sample I will be capturing information on individuals who have not yet had college experiences and are very similar in age and experience to those actually applying to college.

Much of the research on faking has focused on personality tests, allowing enough studies for Viswesvaran and Ones (1999) to meta-analyze the extent to which personality trait scales can have their scores inflated. They conclude that in between-subjects design studies of the Big Five factors and social desirability scales, faking good instructions in

laboratory settings result in response inflation of about half a standard deviation across **the Big Five**, and more than one standard deviation in social desirability scales, compared **to those** with instructions to respond honestly. As is the case in employment selection, by **inflating** scores, some applicants to college may be able to distort and thereby improve **their ranking** in comparison to others, consequently affecting who is selected, under a **top-down** selection system (e.g., Rosse, et al., 1998).

While identifying inflation is one avenue toward ensuring a reliable selection **process**, another is discouraging responses that are dishonest or untrue. Following is a **review** of literature related to inflation that is relevant to biodata measures and situational **judgment** tests. I discuss the effectiveness of various methods to reduce inflation, such as **warning** statements and biodata elaboration requirements, and the use of items with **characteristics** that make them more resistant to inflation.

I address identifying and correcting for score inflation using various scales. I also **examine** the impact of test coaching, as coaching will almost certainly occur in the event **that** any noncognitive measure is used in high-stakes decision-making. In many cases I **have** tapped the literature on personality and integrity tests when it is likely to apply to **responses** to biodata and situational judgment items.

LITERATURE REVIEW

Definitional Issues

The literature covers a range of studies that describe personality test score **inflation** somewhat generally as faking (see Viswesvaran & Ones, 1999). Score inflation **could be** partially explained by a number of effects that are not always differentiated. Effects sometimes classified under the umbrella of response bias are socially desirable **responding**, job-desirable responding, self-deception, impression management, and lying. **Considering** all these effects as one thing does little to further our understanding of how **and why** individuals inflate their scores on certain tests. These different precipitators of **score** inflation are each described below.

Socially desirable responding may occur when the characteristics being captured **by a** test or test item are transparent to the respondent, and those characteristics are **regarded** as attractive by the respondent or more generally attractive in the respondent's **culture**. Personality tests are criticized for being susceptible to socially desirable **responding**, where conscientiousness and emotional stability factors of the Big Five, for **example**, are regarded as socially desirable and adaptive. This transparency potentially **makes** personality measures, such as those capturing the Big Five, susceptible to **inflation**. On average, the Big Five measures have been shown to be inflated by about **half a** standard deviation under instructions to fake good (Viswesvaran & Ones, 1999). **To be** able to control for this tendency, social desirability scales have been created to **gather** information about the tendency of individuals to respond in a socially desirable **manner**. However, these scales themselves are fakeable, with Viswesvaran and Ones

(1999) showing in their meta-analysis a faking effect size of more than one standard deviation on social desirability scales.

Paulhus (1984) presents evidence of the two-component model of socially desirable responding; self-deception and impression management. *Self-deception* is the unconscious inclination that an individual has towards claiming that desirable characteristics apply to them. *Impression management* is the conscious dissembling that an individual engages in to present a favorable impression. Paulhus' Balanced Inventory of Desirable Responding (BIDR) is an example of a social desirability measure that captures both dimensions. Paulhus demonstrated that impression management tended to be more susceptible to situational change than self-deception. It is this impression management component that is regarded as being most closely linked to faking. While these two factors of social desirability are conceptually useful, they have not consistently been empirically supported by other researchers, and there are more nuanced ways to look at them, including the positive and negative attributes of specific items (see Paulhus & Reid, 1991).

Job-desirable responding reflects the recognition that individuals understand that, for particular jobs, different job behaviors or characteristics will be appropriate and desirable. If the test being taken is capturing traits that are transparent to the respondent, then respondents who understand which characteristics are desirable in the particular work context can respond in a job-desirable way. Dwight and Alliger (1997) show that job-relatedness of integrity items was positively related to their fakability, and this may be because job-relevant items are more transparent, which makes job-desirable responding easier. Miller (2001) describes how coaching for a test can focus on traits

desirable for a particular job, where job applicants can modify their responses to a test (or **interview**) to suit the target audience (e.g., an HR recruiter, a direct supervisor).

Lying is a term that should be used with caution, because it implies an intent to **deceive**. Nonetheless, the term has been used to refer to people with scores on “**lie scales**” that in some cases simply comprise unusual responses rather than responses that **are necessarily** lies. Scales that are made up of bogus items reflecting fictitious **experiences** that the individual could not possibly have had (e.g., Anderson, Warner, & **Spencer**, 1984), are the closest that research in this area gets to capturing outright lying.

By identifying these definitional issues I seek to point out that being too quick to **simplify** the issue of faking may result in missing separate and important factors in score inflation. The present study examines score inflation in two tests whose data show initial **promise** for use in a college admissions context: biodata and situational judgment **inventories**. Inflation on these two tests may be a result of any number of the different **factors described** previously. Social desirability is evident in certain biodata and **situational judgment** items, where some responses will be regarded as more socially **acceptable** than others. For example, it may be apparent that an item addressing **multicultural** tolerance has certain response options that reflect socially desirable **behavior**. Similarly, there are biodata and situational judgment items that reflect the **typical expectations** of a college student, and those with information about those **expectations** should not find it difficult to identify questions and response options that **reflect desirable** behavior, such as attending class. Some individuals may have a natural **inclination** to present themselves in a positive light, and this tendency toward impression **management** is expected to vary across individuals, even when they are faced with the

same questions and situational factors as others. These various situational and individual **facets** of social desirability all have a place in research on non-cognitive predictors such as **biodata** and situational judgment inventories.

Having provided some background with the definitional issues that relate to this **research**, I will now address the literature on two tests that are of interest in a college **admission** context: biodata and situational judgment inventories.

Biodata

Biodata measures comprise items related to the examinee's background and **experiences**, and they have long been used as a tool in personnel selection (see discussion in Stokes, 1994, in *The Biodata Handbook*). While biodata items have demonstrated utility **in** adding to the information that we have about a candidate, they are not **impervious** to inflation. Lautenschlager (1994) discusses a selection of twelve biodata studies **that** support the notion that individuals may distort or dissimulate on biodata items, **even** though they are designed as measures of actual background and factual **experience**. Studies have used a range of participant groups, including job incumbents, **job applicants**, and students playing various roles. Accuracy in biodata has been **operationalized** in different ways: correlational accuracy, level of mean differences, and absolute **accuracy**. *Correlational accuracy*, according to Lautenschlager, refers to **consistency** in correlation with an external criterion. *Level of mean difference* refers to **the variation** in the mean score between participant groups in studies where there are different **conditions** for different groups. *Absolute accuracy* refers to consistency in responses at the individual subject level.

Lautenschlager discusses twelve studies using biodata items that relate to faking, going back as far as 1950, providing a map of the development of research in this area. Early studies, such as that of Keating, Patterson and Stone (1950), capture accuracy of biodata by calculating the correlation between the report of the individual and report of the supervisor. Correlations were very high (as high as .98 for duration of employment). While it is possible that correlations would be high if everyone was inflating their responses equally, I suspect that in this case such strong relationships probably occur because of the clear verifiability of the information requested. Today's biodata items tend to cover a broader content area, they may be clearly job-relevant, and at the same time be more difficult to verify as many experiences take place independently of the knowledge of others, making them more susceptible to inflation. Regarding the type of biodata items that may be susceptible to inflation, several studies are illuminating. Klein and Owens (1965) conducted a faking manipulation of a biodata questionnaire under instructions to respond as would a typical, creative research scientist. Under this manipulation, this study found item-type differences, where objective items were less susceptible to inflation than subjective items. Doll (1971) extended the idea of item type by including continuous (Likert scale) and noncontinuous (multiple choice) items, and found that continuous item responses were more susceptible to faking. Also included in this study was a warning of the presence of an existing lie scale, which also reduced faking. A study conducted by Cohen and Lefkowitz (1974) concluded that the MMPI's K-scale could identify people who may be responding in a socially desirable fashion. Thornton and Gierasch (1980) experimented with an empirically keyed biodata inventory, and they concluded that empirical keying did not result in the inventory being

any less vulnerable to faking. Empirical keying, however, has since been demonstrated to be one effective method of limiting faking (Kluger, Reilly, & Russell, 1991). Pannone (1984) attempted a different approach to identify faking – he included a bogus item that asked about experience on a nonexistent piece of equipment. This method continues to generate some interest as a means of identifying faking.

Graham, McDaniel, Douglas and Snell (2002) provide another useful review of biodata studies that relate to faking and validity issues. They cite the study by McManus and Masztal (1993) who used Mael's (1991) taxonomy to investigate the relationship of item characteristics and validity, concluding that historical, external, objective, and verifiable items are less susceptible to faking. Mael's taxonomy has also prompted other research to be discussed in more detail later (see *Composition of items*).

Stanush (1997) conducted a meta-analysis to establish the susceptibility of self-report measures to faking. She concluded that self-report measures are susceptible to distortion under two experimental conditions: under instruction on how to self-present, or under a condition of honest responding ($d = .64$). However, inflation as a result of instructional set was larger than that explained by motivation, which Stanush uses to support the argument that the validity of self-report measures is not likely to be negatively affected by real-life motivation. I question this conclusion, because it is difficult to capture all aspects of the effects of real-life motivation in a study, and evidence about mean differences is different than evidence for similar criterion-related validity. Stanush (1997) does find that biodata inventories tended to be more susceptible to inflation than personality inventories ($d = .94$ vs. $d = .45$, respectively; see also McFarland, 2000). Studies of biodata accuracy have provided valuable information on

the usefulness of these tests; however, the present study addresses the unresolved issue on the effects of inaccuracy in biodata on the predictive validity of the test.

Miller (2001) demonstrates that the Conscientiousness Biodata Questionnaire (CBDQ), a biodata measure designed to capture the personality trait of conscientiousness, was readily faked in a study where participants were provided with a coaching session that provided trait-related information. While understanding the usefulness of biodata as accurate or inaccurate is important, it is also valuable to attempt to reduce inaccuracy. Kluger, et al. (1991) note that findings of response bias in biodata items have been mixed. They argue that empirically keying the item response options creates items that are less susceptible to faking, and also have statistical advantages. With empirical keying, it can be difficult for fakers to guess or know where they will lose or gain points with faking, whereas with item-keying, fakers can operate effectively once the direction of the desirability of responses on the continuum is determined. They note that faking good may also include presenting behaviors that are not socially desirable, but that are perceived as desirable in candidates for a particular position (i.e., job desirable). The authors found that, generally, participants asked to respond as if they were real applicants for a job yielded higher scores than the honest group, had lower empirically keyed scores, and more extreme empirically keyed scores. Gore (2001) also explores the fakability of biodata questions and notes empirical keying as a potential way to reduce response inflation in biodata questions. Considering that the biodata questions generated by Oswald, et al., (in press) are not empirically keyed and have response options that are largely continuous Likert-type scales, similar to most personality self-report measures, it

is worth investigating whether these items are more susceptible to inflation, as the present study does.

Situational Judgment Inventory

Situational judgment inventories (SJIs) typically comprise sample problem scenarios for which the respondent must choose a response representing an appropriate course of action. A review addressing situational judgment items can be found in McDaniel and Nguyen (2001). Motowidlo, Dunnette, and Carter (1990) are cited for their primary study in developing situational judgment items. Items tap a broad range of content areas and dimensions, and they can vary tremendously in complexity. While they are most often presented written in paper-and-pencil format, administration can vary, and may even include the presentation of video vignettes (e.g., Lievens, Coetsier, & Decaesteker, 2000). Key characteristics identified by McDaniel and Nguyen are that situational judgment inventories are typically correlated with ability and experience, as well as conscientiousness, emotional stability, and agreeableness.

While situational judgment measures are useful in that they can address a variety of skills, and they add to the prediction of performance, the method could still permit some level of inflation, depending on how the items are written. If SJIs are found to be less vulnerable to inflation than personality and biodata measures yet also show adequate predictive validity, they may be more likely to be used in a selection context. Nguyen (2002) found that respondents to a situational judgment test for customer service selection were able to raise their scores when instructed to score favorably, depending on the phrasing of the test questions. When the items were phrased in a fashion that required a Most Likely/Least Likely response instruction, participants were able to raise their

scores. However, when the items required a Best/Worst response choice, respondents were not able to raise their scores. Vasilopoulos, Reilly and Leaman (2000) examined a sample of job applicants, and administered a battery of tests that includes situational judgment items. All participants also responded to the BIDR impression management scale, and based on their total score, they were then dichotomously categorized as being high or low on impression management. Although not the focus of their study, Vasilopoulos, et al. found that in comparing groups with low, moderate and high job familiarity, the difference in mean situational judgment scores between those low on impression management and high on impression management grew as job familiarity grew. Mean differences in standard deviation terms (d) were .18, .22, and .73, respectively. These results support the idea that situational judgment items are susceptible to inflation due to socially desirable responding, and that knowledge about the job for which the situational judgment test is written is valuable in gaining a higher score. However, this social or job desirable responding could actually be adaptive behavior that reflects useful knowledge.

Now that I have described the characteristics of biodata and situational judgment inventories, I will discuss coaching as a means of improving performance.

Coaching

Where there is a strong motivation to do well on a test, such as gaining admission to college, people are likely to use whatever resources are available to them to do the best that they can, including test coaching programs. Coaching can be defined as getting outside guidance (White, et al., 2001), and such external interventions can improve scores on selection tests (see discussion in Sackett, Burris, & Ryan, 1989). Kulik, Bangert-

Downs and Kulik (1984), in their meta-analysis of coaching, conclude that studies of improvements in SAT scores reflect a $d = .15$ improvement as a result of coaching. For a range of other ability tests, they found a $d = .51$ coaching effect size for test-retest study designs, and a $d = .27$ coaching effect in studies without a pre-test. Ability tests are less susceptible to coaching than are some non-cognitive tests. Alliger and Dwight (2000), through their meta-analysis, show that coaching instructions and instructions to fake good both tended to improve scores on integrity tests ($d = .90$ with faking instructions, and $d = 1.32$ with coaching for overt tests, and $d = .38$ with faking instructions, and $d = .36$ with coaching for personality-based measures). As integrity tests are non-cognitive in nature, these results may generalize to those for biodata measures. Miller (2001) provided a brief coaching session to participants in a study of their responses to personality and biodata items. Miller's coaching consisted of explaining and defining the Big Five traits, describing how the traits are measured, and providing sample items. The trainer also specified that the most important trait in predicting job performance was conscientiousness, although other traits were also covered so that the participants were able to discriminate between personality characteristics. That study demonstrated that brief coaching resulted in significantly improved performance on the personality dimension of interest in the study – conscientiousness – as well as biodata items tapping conscientiousness, along with post-training knowledge tests that were created for the study.

While tests that are pure g measures are stable, because of the stable nature of g , personality-related tests are not as reliable. As biodata and situational judgment tests are not pure ability measures, but are instead based on personal experience and a broad range

of competencies, it is highly likely that coaching would produce greatly improved performance on these measures, as coaching could make salient the most desirable response pattern. This would allow the participants to modify their self-presentation strategy to maximize their score.

Messick and Jungeblut (1981) found that score improvement was positively related to the amount of time spent in coaching, though there is also a point of diminishing returns, and there can be effective brief coaching. Messick and Jungeblut acknowledge the variety of training that is included under the “coaching” umbrella; some authors restrict their usage of the coaching term to mean practice on sample items and last-minute cramming, while others use it to mean special full-time instruction that could extend for months. The authors were examining coaching for the SAT, a cognitive ability test that requires knowledge and skills typically acquired over a long period of time. The effectiveness of brief coaching has been variable, dependent on the type of coaching provided. The most brief of the SAT coaching programs evaluated in the Messick and Jungeblut (1981) study provided 30-60 minutes each of coaching for Verbal and Math sections of the SAT, resulting in a significant improvement in scores, even though the underlying skills required for high scores on the SAT are supposed to be acquired over long periods of time. Similarly non-cognitive measures are susceptible to brief coaching. Klubeck and Bass (1954) found that 30 minutes of training to improve performance in leaderless group discussion was effective for those who were already high in leadership skills, but less effective for other individuals, while Petty (1974) demonstrated that 15 minutes of training could be effective in improving performance in leaderless group discussions. Sackett, et al. (1989) note that exercise-specific training,

which is effectively training to the test, rather than more general skill development training, increases the effectiveness of coaching. The same authors also comment that there has been an absence of research on coaching on personality tests. Some eleven years later, Miller (2001) states that coaching on personality selection tests has received little attention, and found that an exercise-specific training program that included sample items was effective in improving scores on biodata and personality tests. A 15-minute coaching session that included time to complete a learning outcome measure was sufficient to generate an effect size of $d = 1.66$ on the conscientiousness scale of the NEO-FFI when comparing the scores of those given coaching along with instructions to fake to those receiving control group training and instructions to be honest (coaching and faking instruction effect). The effect size dropped to $d = .48$ when looking at just the coaching effect, by comparing those who received coaching along with faking instructions to those with control group coaching along with faking instructions, to hold the faking instruction condition equal across the groups. Effect sizes for differences on the CBDQ varied by subscale, with the subscales measuring organization ($d = 1.23$) and attention to detail ($d = 1.14$) showing the greatest effect sizes for the coaching and faking instruction effect. The subscales of planfulness and deliberate/rational showed a coaching effect size of $d = .60$.

Cunningham, Wong, and Barbee (1994) conducted a study that included a specific explanation of the content of the test, effectively providing very brief coaching on the underlying rationale of an overt integrity test. This coaching was provided simply by providing written statements about the nature of the test prior to the test administration. For example, the following statement was used to provide information

disparative q

positive attitude

may be recomme

may want to keep

projective quest

project the image

They also see o

He may want to

This w

illustrated the

related areas of

improvement in

greater improve

transitions betw

accomplished b

to be more easi

standing on the

to be effective in

Another

waiting for his

1944, for exam

high ability (a

1.17) on identit

about punitive questions: “For example, honest individuals tend to have relatively **punitive** attitudes towards themselves and to those who commit crimes. They are more **likely** to recommend punishment than forgiveness for those who commit crimes. You **may** want to keep these ideas in mind as you take the test” (p. 650). Similarly, for **projective** questions, participants were coached, “For example, honest individuals tend to **project** the image that they are honest, and deny any temptations towards dishonesty. **They** also see other people, both friends and strangers, as being as honest as they are. **You** may want to keep these ideas in mind as you take the test” (p. 650).

This written form of coaching was effective in raising scores by about 10%, and **demonstrated** that specific information about the test led to improved performance in **related** areas of the test (in this case, information about projective tests leading to greater **improvement** in projective questions, and information about punitiveness leading to **greater** improvement in punitive questions). Considering that coaching could make the **dimensions** being examined in a biodata or situational judgment test more overt (as **accomplished** by Miller (2001) for biodata), and overt tests of integrity have been shown **to be** more easily faked than covert tests (Alliger & Dwight, 2000), relatively brief **coaching** on the dimensions captured by biodata and situational judgment items is likely **to be** effective in raising scores on these tests.

Another issue that needs to be examined is whether there are differential effects of **coaching** for high and low ability applicants. Kulik, Kulik, and Bangert’s meta-analysis (1984), for example, found that effect sizes of practice on aptitude tests were greater for **higher** ability ($d = .82$) and middle ability ($d = .40$) samples than for low ability groups ($d = .17$) on identical tests in a test-retest study. Also, ethnicity may have an effect on

existing effectively

and middle

existing Ryan

programs in a self

construed the

access to coaching

influences might

That coach

existing on blood

some research has

eg. Bunting &

college admissions

W. Jones & V

depending to be

existing would

relation context

where appear

existing on the

limited education

for would expect

to be elevated

coaching effectiveness, namely non-Whites may be more likely than Whites to be in the lower and middle-ability groups, and may be less likely to improve as a result of coaching. Ryan, Ployhart, Gregarus, and Schmit (1998), in discussing test preparation programs in a selection context cite a study presented by Holden (1996) that demonstrated that coached Caucasians improved more than minorities. Differential access to coaching along with differential results based on ability or other individual differences might exacerbate the problem of adverse impact in college admissions.

That coaching can be effective is cause for closer examination of the effects of coaching on biodata and situational judgment tests for use in college admissions. While some research has been conducted on the coaching of children for educational selection (e.g., Bunting & Mooney, 2001), and there has been an examination of coaching for college admissions tests such as the SAT, as I have discussed, and MCAT (e.g., Jones, 1987; Jones, & Vanyur, 1987), there is a dearth of work on the coaching of students responding to biodata and situational judgment tests in a college admissions context. As coaching would probably become widely available if such a test was to be used in a selection context, it is important to consider changes in its predictive validity, especially as there appear to be differences in students who undertake coaching. Those who seek coaching on the SAT tend to be from more affluent families, have parents with more formal education, and are more likely to be Asian American (Powers & Rock, 1998). One would expect that as coaching became available, scores of those coached would tend to be elevated, and the predictive validity of the test may be reduced over time.

Conscient

and striving (E

are conscient

as that will cau

in college. No

is be individual

scientific uses

depending (Onc

ling in biolut

ling condition

McFarland and R

ation on pers

individual variab

and a sample o

later. The orde

dependents com

struction cond

Participants in th

18% in test secon

personality char

measures includ

tioned.

Individual Differences in Faking

Conscientiousness is linked to achievement-striving, and extraversion is linked to **status-striving** (Barrick, Stewart, & Piotrowski, 2002), which suggests that individuals **who** are conscientious and extraverted may tend to be more motivated to perform well on **tests** that will earn them a symbol of achievement and status, such as getting accepted **into** college. Not only are there possible individual differences in motivation, there may **also be** individual differences in ability to inflate responses, as emotional stability and **conscientiousness** have been found to be positively related to socially desirable **responding** (Ones, Viswesvaran, & Reiss, 1996). Studies that have been conducted on **faking** in biodata have been criticized for making the assumption that everyone in a **faking** condition inflates their scores to the same extent (e.g., Becker & Colquitt, 1992). **McFarland** and Ryan (2000) argue that there are individual differences in response **inflation** on personality-related measures, and they conducted a study examining **individual** variability in faking across non-cognitive measures. They conducted a study **using** a sample of students with honest vs. fake instructional sets as a within-subjects **factor**. The order of instructions was manipulated as a between-subjects factor. The **respondents** completed an integrity test and biodata inventory under the manipulated **instruction** conditions, and then all respondents completed a self-monitoring scale. **Participants** in the faking condition were offered a financial incentive whereby the top **15%** in test scores would receive \$15. **McFarland** and Ryan concluded that the **personality** characteristics found to be positively associated with faking on noncognitive **measures** include integrity and conscientiousness, while neuroticism was negatively **related**.

Mers

contract. The

efficiency

establishing

and like cond

and self-mon

advised dis

presentation

were told to f

their respon

dition. The

relationships

and non-fake

It is ap

often require

and emotional

visesvaran

test or pattern

helpful

Having

differences. I

Mersman and Schultz (1998) suggested that the ability to fake is an independent **construct**. They conducted a study with a sample of students who completed personality, **self-monitoring** and social desirability scales, and the tendency to fake good was captured **by establishing consistency** with Saucier's (1994) Big Five Mini-Markers under honest **and fake** conditions. Mersman and Schultz found that faking good was not correlated **with** self-monitoring, impression management or self-deception, and conclude that **individual** differences in faking capacity are unrelated to those constructs of self-**presentation**. I view this conclusion with caution as in this study, the participants who **were told** to fake did not have any real motivation to do so. As with many studies, the **honest** respondents may not have been providing responses that were entirely without **inflation**. The combination of these two factors in this study may have minimized the **relationships** between faking and individual differences because the gap between fakers **and non-fakers** was restricted.

It is apparent that people do not all inflate their responses to the same extent, **when** required to fake, and there is at least preliminary evidence that conscientiousness **and emotional** stability are related to faking (McFarland & Ryan, 2000; Ones, **Viswesvaran**, & Reiss, 1996). However, further study contributions on what particular **trait** or pattern of traits make some people better able to fake than others would be **helpful**.

Having discussed the relevant literature on biodata, SJIs, coaching, and individual **differences**, I will now consider inflation as it relates to validity.

It is

missions

regarding

across a

difficult

are more

Dragon

recognized

methods

ABLE data

responses

likely poss

less than

likely fake

validity. The

and coached

suggesting

demonstrated

interviews

was not

Some

Smith and

Effects on Validity

It is apparent that the high value that applicants to college place on college admissions is likely to precipitate motivated responding on college admissions tests. Regarding ability, self-reports of ability have generally been found to be quite accurate across a number of research studies (Mabe & West, 1982), and with *g*-loaded tests it is difficult to inflate one's scores beyond one's ability level. Noncognitive tests, however, are more malleable as the best response can in many cases be guessed. Zickar and Drasgow (1996) acknowledge that as tests are fakeable, it may be useful to focus on how to recognize the fakers after they have taken the test. One way to do this is with detection methods that identify patterns of inconsistent responding. They used two samples in the ABLE dataset from Project A. Polytomous item response models were used to score responses, and IRT model fit was examined. They then computed person-fit indices to identify possible fakers. They found that fakers who had been coached were easier to detect than those who were simply *ad lib* faking. They examine a theta-shift model to identify fakers, conducting statistical manipulations to test for the effects of faking on validity. Their results show theta differences between honest responders, *ad lib* fakers, and coached fakers, and they found larger effect sizes than shown by Ones, et al., suggesting that faking may have more negative implications for validity than has been demonstrated. However, they did not report correlations between personality scales and outcomes under the different faking conditions, so the issue of criterion-related validity was not addressed directly.

Some authors claim that construct validity is not changed by score inflation. Smith and Ellingson (2002), in comparing a sample of applicants and a sample of

patients (non-ap

cross all scales

measures did not

feasibility is tr

examined by in

and non-applied

are different th

preparation.

Anderson

the effect of in

healthy adjustm

scores due to so

roles). So

related validity

self-deception a

Using regression

samples of true

voluntary con

validity of the p

controlling for

validities were

meta-analysis

effectiveness

students (non-applicants) found that applicants did not simply inflate their responses **across** all scales, and the relationship between social desirability scales and personality **measures** did not change significantly across samples. They concluded that social **desirability** is trait-based rather than situation-driven, and that construct validity is not **attenuated** by inflation. However, as there were slightly different patterns for applicants **and non-applicants**, where applicants scored higher on some positive dimensions that **were** different from the dimension for the non-applicant group, that conclusion may be **premature**.

Anderson, Warner, and Spencer (1984) note that there are mixed opinions about **the effect** of inflation due to socially desirable responding. Inflation can be regarded as a **healthy** adjustment to a specific situation, and some researchers have found that inflated **scores** due to socially desirable responding do predict performance in certain jobs (e.g., **sales roles**). Some researchers provide evidence that faking has little effect on criterion-**related** validity. Barrick and Mount (1996) conducted a study examining the effects of **self-deception** and impression management on the predictive validity of the Big Five. **Using** regression analyses and latent-variable modeling, they concluded that, in two **samples** of truck drivers, conscientiousness and emotional stability were valid predictors of **voluntary** turnover and supervisory ratings of performance, and that the predictive **validity** of the personality measures was not significantly negatively affected by **controlling** for self-deception and impression management. They do note that adjusted **validities** were generally slightly smaller. Ones, Viswesvaran and Reiss (1996), in their **meta-analysis**, conclude that correcting for social desirability does little to increase the **effectiveness** of Big Five measures as predictors, however, it would be inappropriate to

truly claim as

looks at a narrow

model rather than

inspects that so

pecially because

Vijayvaran &

emotional stability

capture all that

Some re

Attitudes, view in

predictors of pe

Smith & Ellings

Graham McDan

likely enhanced

study comparing

conditions, with

items according

validity of item

For example, ch

for the honest c

through super

responders, an

as well as the

broadly claim as a result of their findings that faking does not affect validity. Their study **looks** at a narrow perspective – the use of specific social desirability measures in their **model**, rather than measures that are more situation- rather than personality-driven. Also, **I suspect** that social desirability scales do not fully capture the extent of inflation, **partially** because these measures, too, are susceptible to socially desirable responding (**Viswesvaran & Ones**, 1999). They are also correlated with conscientiousness and **emotional** stability (Ones, Viswesvaran, & Reiss, 1996), and they do not appear to **capture** all that explains score inflation.

Some researchers are more suspicious of inflation in self-reports of skills and **abilities**, viewing the inaccuracies as a real problem that may make the reports invalid **predictors** of performance, and less useful as differentiators between applicants (e.g., **Smith & Ellingson**, 2002; Topping & Gorman, 1997). According to a recent paper by **Graham, McDaniel, Douglas and Snell** (2002), “For biodata, the degree of prediction is **likely** enhanced by the accuracy of the self-report information” (p. 574). They conduct a **study** comparing responses to biodata items under honest-responding and faking-good **conditions**, with job performance as the criterion of interest. They categorize biodata **items** according to various item characteristics, and demonstrate that criterion-related **validity** of items with different characteristics varies across honest and faking conditions. **For example**, they show that for items rated as verifiable through hard records, validity **for the** honest condition was .16, but .02 for the faking group. Items rated as verifiable **through** supervisors or coworkers showed validity coefficients of .12 for honest **responders**, and .03 for fakers. It is apparent from this study that the item characteristics **as well** as the situational instruction set are related to validity.

Me

consent

be of di

practical s

Dragon W

personality

the Sixtee

applicants

applicants

applicant

to PF is ne

On

job destr

best possi

well as in

behavior i

performan

academica

leadership

soon. Suc

perceived

ters Coa

Miller (20

McFarland (2000), who found that the criterion-related validity of conscientiousness was not significantly affected from a statistical perspective, stated that the .05 difference in validity between the honest and applicant conditions might be of practical significance in a real selection context. Stark, Chernyshenko, Chan, Lee, and Drasgow (2001) cite disagreement about the effects of faking on the validity and utility of personality tests. They look at both trait and situational perspectives in faking, and using the Sixteen Personality Factor Questionnaire (16PF), in a sample of applicants and non-applicants, they showed significant differential item functioning (DIF) across groups of applicants and non-applicants, where the items appeared to operate differently for applicant and non-applicant groups. They conclude that the construct validity of the 16PF is negatively affected by faking among applicants.

Ones, Viswesvaran, and Reiss (1996) note that job applicants may be faking in a job desirable manner rather than a socially desirable manner. Presenting oneself in the best possible light in a college admissions context could be in a job desirable fashion, as well as in a socially desirable direction. In the college context, the issue of job-desirable behavior is limited by the individual conception of what is valued as effective college performance. While some may consider job-desirable college performance as academically focused behavior, others may recognize the broader goals that include leadership behavior, interpersonal skills, multicultural tolerance, artistic appreciation, and so on. Such differences in perceptions of the criterion space may affect individuals' perceived job-desirable behavior, and possibly also the criterion-related validity of the tests. Coaching is one way to provide information about what is job-desirable behavior. Miller (2001) demonstrates that those who receive coaching are far better equipped to

system well o

the top of

toward the b

would have

hands of p

while thro

washing

selected an

Fair

of teaching

identify it, a

result of an c

familiarity (s

construct val

admission to

admissions co

traditions.

Having

problems. I w

relation.

As I e

mention when

perform well on personality and biodata items, with those who received coaching rising **to the** top of the distribution of performance scores, and those without coaching falling **toward** the bottom. In a top-down selection system based on test performance, coaching **would** have had a real effect on who was selected in this context, by filling the upper **bands** of performers with individuals whose performance has been enhanced by coaching, **while** those who have not had the opportunity to improve their performance through **coaching** might fall into lower performance bands, and find themselves not getting **selected** under a top-down system.

Faking is apparent in testing in many selection contexts, and if faking as a result **of coaching** affects test validity, it is essential that steps are taken to discourage faking, **identify** it, and correct for it. However, score improvements after coaching could be a **result** of an artificial increase in the score, or a real improvement as a result of task **familiarity** (see Sackett, et al., 1989). Our goal is to have admissions tests that have both **construct** validity and predictive validity. Students are motivated to perform well to gain **admission** to college, and if biodata and situational judgment tests are to be used in an **admissions** context, we need to understand better how they perform under coached **conditions**.

Having noted important areas of the literature on inflation and possible related **problems**, I will now discuss ways in which researchers have attempted to control **inflation**.

Inflation Controls

As I expect that faking is possible with non-cognitive measures, and likely in a **situation** where there is a motivation to perform well on those measures, it is pragmatic to

to prevent
warning states
characteristics

Warning

status and
filing on non-

lowering pr

cluded the m

for dishonest

would result

= 61 for dem

received an op

importance of

warrants gre

distinction" (p

warning states

They found st

filing group

(1999) used a

included pers

scores on three

judgment scale

try to prevent faking in the first place. Three possibilities of faking control are the use of warning statements, the requirement of item elaboration, and the consideration of certain characteristics in item writing.

Warning

Warning statements may be a practical and effective way to manage inflation on **biodata** and situational judgment inventories. Dwight and Donovan (2003), in examining **faking** on noncognitive selection measures, found that warning statements were effective **in lowering predictor scores**. They demonstrated that by providing a warning that **included the risk of verification of dishonest responses as well as a negative consequence for dishonest responding, the least faking on the California Psychological Inventory (CPI) would result**. They showed an effect size of $d = .75$ on the CPI scale of well-being, and $d = .61$ for dominance, when comparing those who were not warned with those who **received an optimally effective warning**. Hough et al. (1990) also emphasize the **importance of the use of warning statements, stating that the use of warning statements “warrants greater attention as a method of reducing the amount and effect of intentional distortion” (p. 582)**. In a sample of employees, Becker and Colquitt (1992) included a **warning statement that responses to biodata questions may be verified with other sources**. **They** found significant mean differences between a test-taking group with no warning, a **faking** group with no warning, and an applicant group who was warned. Vasilopoulos (1999) used a warning of response verification in a study of a selection system that **included personality and situational judgment measures, with a resultant mean drop in scores on three of the personality scales for warned respondents, but not on the situational judgment scale**. Warning statements are likely to be most effective in limiting faking

that they include

among other

among other

and due to

and comp

Zi

personality

analysis, a

the indiv

the compus

findings with

with the type

either be born

the instrument

was falsifiable

related score

characteristic

Waller (1991)

objective di

relevant to th

Jacky Jones

the relative

when they include both a warning about the potential identification of faking, and a **warning** about the potential consequences of faking (Dwight & Donovan, 2003). While a **warning** may be effective in reducing score inflation, the effect may vary from item to **item** due to varying item transparency (Kluger & Colella, 1993); hence consideration of **item** composition is important.

Composition of Items

Zickar and Robie (1999) conducted an item-level analysis of faking good on **personality** items. They criticize the focus of the research literature on scale-level **analysis**, and identify the need for item-specific examination as people “respond to (and **fake**) individual items, not scales” (p. 552). Whether or not a test is fakeable depends on **the** composition of the items. Becker and Colquitt (1992) note the discrepancies in **findings** with regard to biodata items being faked and posit that the differences have to do **with** the type of items. They examined a group of employees who were instructed to **either** be honest, or fake good on a biodata instrument, and job applicants, who completed **the** instrument as part of the application process. The authors concluded that the form **was** fakeable, and that those who were faking good or were real applicants did have **inflated** scores compared to the honest condition. Items were examined for particular **characteristics** that may make them more or less fakeable, and using the framework of **Mael** (1991), they found that items that are more likely to be faked were less historical, **objective**, discrete, verifiable, and external than other items. They were also more **relev**ant to the job. (Mael’s taxonomy is shown in Appendix A.) Similarly, Elliot, **Lawty-Jones**, and Jackson (1996) found that responses to objective tests of personality **were** relatively unaffected by instructions to fake. Dwight and Alliger (1997) conducted

ایضاً

254

2251

—

NSW

252

1. *Chlorophyll a* and *Chlorophyll b* contents were determined by spectrophotometry using the method of Lichtenthaler and Whistler (1987). The total chlorophyll content was calculated as the sum of chlorophyll *a* and chlorophyll *b* contents.

2020

۱۴۰۰

Report

1255

2000

255

22

April

20

2000

422

2900

1992

52

22

25

a study of ratings of individual integrity test items, finding that the perception that an item would be easy to fake was related to the job relatedness ($r = .50, p < .001$) and invasiveness of the item ($r = .25, p < .05$). In their meta-analysis of the susceptibility of integrity tests to coaching and faking, Alliger and Dwight (2000) conclude that the overt tests were more susceptible to faking and coaching than were the covert tests. Biodata and situational judgment items should be less fakeable where the items are more objective, verifiable, and not clearly related to college performance. Mael's (1991) taxonomy of biodata items provides further ideas for item characteristics that may be used sensibly in biodata items, including equal access and being controllable.

Respondents without perceived access to the biographical experiences to which an item refers might be less likely to inflate their scores than individuals who have had such experiences. They may view such experiences as completely beyond the realm of possibility. Similarly, individuals may be unlikely to inflate responses when the issue is one in which they have little control. Graham, et al., (2002) examine biodata validity as explained by item characteristics using Mael's (1991) attributes. They concluded that item attributes associated with validity are different for faking and honest respondents, leaving the authors skeptical about the possibility of writing biodata items that are valid for both fakers and honest responders. For honest responders, item attributes most highly associated with validity were those that were controllable and concerned with the individual's feelings or attitudes ($r = .22$), verifiability through hard records ($r = .16$), and verifiable through supervisors or coworkers (.12). For fakers, items that were verifiable through friends ($r = .11$), controllable and concerned with actions that the individual chooses to perform ($r = .07$) were associated with validity. Schmitt, Gillespie, Kim,

Amosy

Advent

Amosy

R

Advent

Amosy

Advent

Amosy

T

S

T

n

It

E

—

—

—

Ramsay, Oswald, and Yoo (in press) found that biodata items that were more objective **and** verifiable were less correlated with the participants' BIDR self deception and **imp**ression management scores.

Requiring elaboration within biodata items is another method that may reduce the **like**lihood of individuals inflating their responses. In requiring elaboration, the test item **s**pecifies that the respondent should provide examples that reflect evidence of the level of **e**xperience that they indicated, rather than simply indicating the level of experience. An **e**xample of a biodata item with an elaboration requirement is shown in Table 1.

Table 1

Sample Elaborated Biodata Item

The number of high school clubs and organized activities (such as band, sports, newspapers, etc.) in which you took a leadership role was:

- a. I did not take a leadership role
- b. 1
- c. 2
- d. 3
- e. 4 or more

If you answered b, c, d, or e, briefly indicate up to 4 clubs or activities and the nature of your role.

1

est de

un po

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

est de

Schmitt and Kunce (2002) found that by requiring item elaboration in a biodata **test**, they reduced mean scores by .7 to .8 standard deviation units. They also found **carry-over** effects of score reduction in nonelaborated items in the same instrument. By **requiring** elaboration on biodata items, it should make the items less likely to be faked. **This** effect of elaboration on biodata items was found again in a more recent study by **Schmitt**, et al. (in press); however, the carryover effect was not found. Schmitt et al. (in **press**) showed a difference in mean scores of .8 standard deviations between elaborated **and** non-elaborated items, where elaborated items had lower scores. It may be valuable **to require** elaboration to biodata items as a means of limiting faking in elaborated items **and** possibly, through a carryover effect, in non-elaborated items. However, one should **bear** in mind one possible explanation; the lower scores as a result of elaboration could be **because** the individual involved simply can't remember the details of the biographical **event**, and they may limit their response to examples on which they can recall enough **detail** to elaborate.

While attending to specific item characteristics when generating items is an **important** way of discouraging inflation, it is likely that some individuals will still inflate **their** responses, and I will now discuss the identification of inflation.

Statistical Control using Social Desirability or Bogus Items

While I expect some score inflation to take place, it would be useful to be able **identify** those who tend to manipulate their responses. Inflation identification and control **possibilities** include the use of bogus items as a lie detection system, the use of scales that **capture** impression management, and the use of indices of improbable responding. Once **faking** has been identified, it is possible to statistically control for it by partialing out the

effects were

still or per

2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287, 288, 289, 290, 291, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 387, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435, 436, 437, 438, 439, 440, 441, 442, 443, 444, 445, 446, 447, 448, 449, 450, 451, 452, 453, 454, 455, 456, 457, 458, 459, 460, 461, 462, 463, 464, 465, 466, 467, 468, 469, 470, 471, 472, 473, 474, 475, 476, 477, 478, 479, 480, 481, 482, 483, 484, 485, 486, 487, 488, 489, 490, 491, 492, 493, 494, 495, 496, 497, 498, 499, 500, 501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 513, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527, 528, 529, 530, 531, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 557, 558, 559, 560, 561, 562, 563, 564, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 576, 577, 578, 579, 580, 581, 582, 583, 584, 585, 586, 587, 588, 589, 590, 591, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 621, 622, 623, 624, 625, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 639, 640, 641, 642, 643, 644, 645, 646, 647, 648, 649, 650, 651, 652, 653, 654, 655, 656, 657, 658, 659, 660, 661, 662, 663, 664, 665, 666, 667, 668, 669, 670, 671, 672, 673, 674, 675, 676, 677, 678, 679, 680, 681, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 694, 695, 696, 697, 698, 699, 700, 701, 702, 703, 704, 705, 706, 707, 708, 709, 710, 711, 712, 713, 714, 715, 716, 717, 718, 719, 720, 721, 722, 723, 724, 725, 726, 727, 728, 729, 730, 731, 732, 733, 734, 735, 736, 737, 738, 739, 740, 741, 742, 743, 744, 745, 746, 747, 748, 749, 750, 751, 752, 753, 754, 755, 756, 757, 758, 759, 760, 761, 762, 763, 764, 765, 766, 767, 768, 769, 770, 771, 772, 773, 774, 775, 776, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 787, 788, 789, 790, 791, 792, 793, 794, 795, 796, 797, 798, 799, 800, 801, 802, 803, 804, 805, 806, 807, 808, 809, 810, 811, 812, 813, 814, 815, 816, 817, 818, 819, 820, 821, 822, 823, 824, 825, 826, 827, 828, 829, 830, 831, 832, 833, 834, 835, 836, 837, 838, 839, 840, 841, 842, 843, 844, 845, 846, 847, 848, 849, 850, 851, 852, 853, 854, 855, 856, 857, 858, 859, 860, 861, 862, 863, 864, 865, 866, 867, 868, 869, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, 885, 886, 887, 888, 889, 890, 891, 892, 893, 894, 895, 896, 897, 898, 899, 900, 901, 902, 903, 904, 905, 906, 907, 908, 909, 910, 911, 912, 913, 914, 915, 916, 917, 918, 919, 920, 921, 922, 923, 924, 925, 926, 927, 928, 929, 930, 931, 932, 933, 934, 935, 936, 937, 938, 939, 940, 941, 942, 943, 944, 945, 946, 947, 948, 949, 950, 951, 952, 953, 954, 955, 956, 957, 958, 959, 960, 961, 962, 963, 964, 965, 966, 967, 968, 969, 970, 971, 972, 973, 974, 975, 976, 977, 978, 979, 980, 981, 982, 983, 984, 985, 986, 987, 988, 989, 990, 991, 992, 993, 994, 995, 996, 997, 998, 999, 1000

It

unlike

Using the

Using the

1941. A

as of a

for the

experience

side score

the inflation

they had

for the in

items were

were also

and type

and that th

this mea

the intent

Anderson

the material

effects using the various faking indices as covariates in estimating the validity of biodata, **SJI**, or personality measures (see Barrick & Mount, 1996).

Bogus items

It is apparent that individuals do inflate responses on certain tests in certain **contexts**, and that if one expects that faking will happen, it is practical to plan for it. **Using** items that capture the claim of impossible experiences is one way to identify **faking** (e.g., Alliger & Dwight, 2000; Anderson, Warner, & Spencer, 1984; Pannone, 1984). Anderson, et al. used such bogus items interspersed among real job task items in a **test** of a job applicant sample to establish whether applicants were faking. The level of **their** affirmations of experience in bogus items was used to adjust downward their **experience** score on real items. This was justified by the correlation between the inflation **scale** scores and examinations for the particular occupational classes. They concluded **that** inflated scores are pervasive. In their sample, 45% of the participants indicated that **they** had observed or performed at least one bogus task. Anderson, et al. also concluded **that** the inflation scales had high reliabilities (average alpha of .86) and that the bogus **items** were useful. In a secondary component of this study a sub-group of the applicants **were** also examined on a typing test after being asked how many words per minute they **could** type. The authors found that applicants did inflate their self-report of typing skill, **and** that this inflation correlated with the scores on the bogus items in the more general **skills** measure. Correcting for inflation increased the usefulness of the test in predicting **the** criterion. Alliger and Dwight (2000) generated bogus items similar to those used by **Anderson, et al.** (1984). They state that the bogus scale is “a measure of fabrication, the demonstrated willingness of the person to create information about themselves that has

a connection

suggesting

had the greatest

items are tapping

Social Desirability

Social

positive improvement

people can do

response distribution

trained social

The social desirability

those trying to

were education

base. They

response validity

Elling

desirability score

validity. In the

the SD scale

scores. How

components

stable trait

that is

no connection to actual experience, rather than a measure of subtle faking (e.g., **exaggerating** positive attributes)” p. 10. They found that the optimal warning condition **had** the greatest impact on scores on the bogus item scale ($d = .41$) and suggest that bogus **items** are tapping a specific aspect of faking – the tendency to fabricate information.

Social Desirability and Improbable Response Indices

Social desirability scales have been used to identify those more likely to present a **positive** impression. Hough, Eaton, Dunnette, Kamp, and McCloy (1990) agree that **people** can distort their responses, having examined personality constructs and the effect **of** response distortion on their validity, with four response validity scales that they **created**: social desirability, poor impression, self-knowledge and nonrandom response. **The** social desirability scale was patterned after other unlikely virtues scales, to detect **those** trying to appear more attractive, where the areas of experience being examined **were** education, training, job involvement, job proficiency, delinquency and substance **abuse**. They concluded that people can fake when instructed to do so, and that the **response** validity scales detected the distortion.

Ellingson, Sackett, and Hough (1999) attempted a correction using a social **desirability** scale score and found that the corrected score was ineffective at improving **validity**. In their within-subjects design across faking/non-faking conditions, they used **the** SD scale to correct faking conditions to see whether scores approximated non-faked **scores**. However, Paulhus (1984) emphasizes that the situation determines whether both **components** of social desirability should be controlled. As self-deception is viewed as a **stable** trait, there may be little value in controlling for it in examining the effectiveness of **a** test that is being used under specific situational constraints that are affecting the

inclination to provide socially desirable responses. In such a case, it may be more **relevant** to control for impression management to reduce the effect of conscious faking. **Controlling** for social desirability with a scale such as the Marlowe-Crowne may not **provide** useful information as the scale is made up of both factors in social desirability; **impression** management and self-deception (Paulhus, 1984).

Barrick and Mount (1996) examine the effect of self-deception and impression **management** on the predictive validity of the Big Five when looking at turnover and **supervisor** ratings. They found that distortion of personality constructs occurred through **self-deception** and impression management, but that the distortion did not negatively **affect** the validity. Christiansen, Goffin, Johnston and Rothstein (1994) attempted to **correct** for inflation on personality scales by partialing out the effects of the inflation **scale**, however, they were not able to demonstrate that such a correction affected the **validity** of the personality measure as a predictor of performance ratings.

Other researchers, however, have found that criterion-related validity and **selection** decisions can be improved by correcting for faking. Hough (1998) used results **of an** “Unlikely Virtues” scale to either adjust the individual’s score or to remove them **from** the applicant pool, finding both techniques to be effective in reducing the impact of **score** inflation on hiring decisions and increasing the predictive validity. Anderson, **Warner**, and Spencer (1984) attempted corrections after using bogus items that are true **lie** scores and found that corrections improved the test’s usefulness as a predictor.

Becker and Colquitt (1992) corrected the scores of their motivated group (the **applicants**) by using an inflation index. The index was created by calculating the mean **difference** between scores for applicants and non-applicants. This index was then used to

adjust the distr

tensions, the c

the candidates

incorrected sc

system with s

the same exten

scores to test v

found that with

in a selection m

if positions in

demonstrated

Althoug

may be a prac

related validit

items are to b

To su

scores are us

emotional ju

for blind dat

manipulation

be be imp

moderated +

adjust the distribution of scores. When corrected scores were used to make hiring **decisions**, the correction changed the rank order of candidates to the extent that 17% of **the** candidates were faced with a reversal of the hiring decision made under the **uncorrected** scores. This correction effectively neutralized the inflation, however, one **concern** with such an approach is that it assumes that everyone has inflated their scores to **the same** extent. Christiansen, et al. (1994) used the Krug (1978) approach to correcting **scores** to test whether there would be different hiring decisions after correction. They **found** that without correction, discrepant hiring decisions were made in 16% of the cases **at a** selection ratio of .15, and they also note that overall, candidates moved an average of **23** positions in rank ($SD = 36$) when corrections were used. Rosse, et al. (1998) also **demonstrated** that by inflating scores, applicants may be able to affect who is hired.

Although there are mixed findings about the effectiveness of corrections, they **may** be a practical way for test users to examine the effect of score inflation on criterion-**related** validity. Such examination will be important if biodata and situational judgment **items** are to be used in a college admissions context in predicting student performance.

Study Development

To summarize, traditional measures such as high school GPA and SAT/ACT **scores** are used to make college admissions decisions, and tests such as biodata and **situational** judgment could also be used (Oswald, et al., in press). However, it is apparent **that** biodata scores can be inflated, although they may not be as susceptible to **manipulation** as are personality scales (Sisco, 1999). Situational judgment scores can **also** be improved by faking (Nguyen, 2002). Performance on these measures is **moderated** by the situational factors of the conditions under which the tests are taken.

Being m

receiving

well all

hence a

though s

consider

in data

and the

with as

elaborate

where ce

W

Graham

negative

may be h

who resp

such a pu

direction

respond

up-down

By select

validate

method

Being motivated to perform well, experiencing coaching to be able to perform well, and **receiving** a warning statement about not exaggerating responses in an attempt to perform **well** all may contribute systematically to the score that an individual achieves. These **factors** are expected to interact, with some combinations providing a stronger likelihood **of high** scores than others. There are also individual differences such as **conscientiousness** and emotional stability that are expected to relate to performance on **biodata** and situational judgment. While cognitive ability may also affect faking, ability **is not** the focus of this study. Rather, I consider the individual differences of personality **traits** as they relate to performance and inflation. Item characteristics, such as biodata **elaboration** requirements and item response verifiability, also play an important role, **where** certain characteristics may make some items more vulnerable to inflation.

While there are mixed views on the effects of faking on validity, recent work by **Graham**, et al. (2002) demonstrates that criterion-related validity of biodata items can be **negatively** affected by faking. Assuming, therefore, that inflation could be problematic, it **may** be helpful to control for score inflation. This could be achieved by identifying those **who** respond to bogus items, those who score highly on an inflation index developed for **such** a purpose, and those who score unusually highly on impression management, and **correcting** for the inflation. Inflated scores may result in a change in the rank order of **respondents** (e.g., Frei, 1998), and subsequently, a change in who would be selected in a **top-down** selection system in admissions; a critical issue in a college admissions context. **By** selecting based on performance on these tests, we hope to be able to choose the best **candidates**, i.e., those with good actual college performance (college GPA and **attendance**).

2

Information for
Decision
Inflation

Individual
Differences

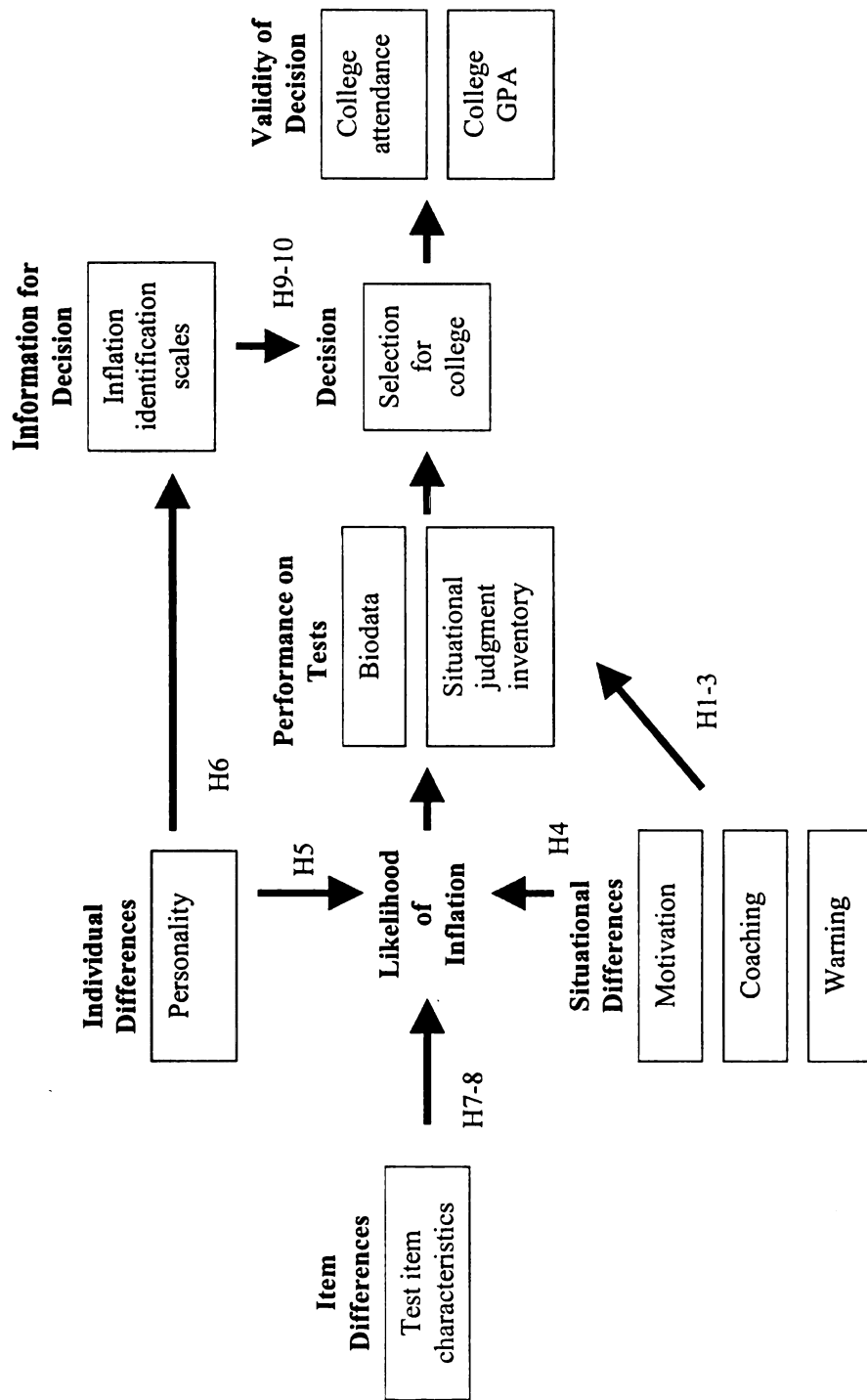


Figure 1
Conceptual Model

Figure

the major q

Firs

influence

addressed

gun admini

warnings

this study

sure infla

Sec

most effec

in general o

provide con

contribute t

some gener

emotional s

related, is ex

typically ide

stronger rela

testing hypo

Thir

inflation? A

important ro

Figure 1 provides a model linking the various components of the study, to address the major questions defined below.

First, how do different situational differences such as testing conditions influence how an individual performs on biodata and SJIs? Situational factors addressed in the preceding literature review included: incentive to perform well so as to gain admission to college, coaching as a way to know how to achieve the best score, and warnings to prevent score inflation. By examining these factors and their interactions, this study will enhance our understanding of the susceptibility of biodata and SJIs to score inflation. These factors are considered in examining hypotheses 1 – 4.

Second, what are the individual difference characteristics of people who are most effective in inflating their responses? We do know that those who perform well in general are conscientious and emotionally stable. The existing literature does not provide conclusive evidence on who is most likely to fake, and so this study will contribute to the development of theory on individual differences and inflation. Certainly some general predictions can be made. For one, it is expected that conscientiousness and emotional stability will be positively related to inflation. Self-deception, which is trait-related, is expected to be associated with inflation. While impression management is typically identified as a trait, it is also influenced by situation, and is expected to have a stronger relationship with inflation than self-deception. These questions are addressed in testing hypotheses 5 – 6.

Third, what are characteristics of the test items that seem most susceptible to inflation? As reflected in the work of many researchers, item characteristics can play an important role in developing a test invulnerable to faking. By examining the

character

electro

Still her

hyper-

predic

This is

in an ap

element

testing

effect t

in her

know th

Ones. I

Barbee

and Do

28, 1914

1914-1915

characteristics of biodata and SJI items under varying testing circumstances, I will be able to contribute to the theory of test construction, specifically for improved biodata and SJI items that are less vulnerable to faking. This question is addressed in examining hypotheses 7 – 8.

Fourth, does the susceptibility of biodata and SJIs to faking affect their predictive validity if they were to be used in college admissions decision-making?

This issue is of critical importance if the tests used here are put into practice for selection in an applied setting. I seek to identify differences in predictive validity and establish the extent to which inflation affects whom would be selected. This question is addressed in testing hypotheses 9 – 10.

To address the first question of the study, regarding the situational factors that affect faking, I propose several hypotheses based on past research. Different conditions influence how an individual performs on biodata and situational judgment tests. We know that people can and do inflate their scores when motivated to do so (Viswesvaran & Ones, 1999), that coaching is effective in improving scores (e.g., Cunningham, Wong, & Barbee, 1994), and that warning statements effectively suppress inflation (e.g., Dwight and Donovan, 2003). Specifically, our hypotheses are as follows:

H1: Motivation will improve scores on biodata and situational judgment.

H2: Coaching will improve scores on biodata and situational judgment.

H3: Warning will reduce scores on biodata and situational judgment.

Interactions between these situational factors as well as the main effects associated with each factor are also explored. For example, someone who is highly motivated to perform well, is provided with coaching, and is not given any warning about

1

and

since

the

but

the

These

the

of the

when

the

the

the

the

Table

Expect

Notes

Notes

Notes

inflating responses is most likely to score highest on the tests. Next highest would be someone who has had coaching and no warning, but no motivation. Individuals in these two groups would be more cautious in inflating responses when provided with a warning, but the warning is likely to have a more powerful effect for those who are motivated, resulting in those without an incentive having a higher score than those with an incentive. Those who are highly motivated to perform well have more to lose by being caught faking than those who are not particularly concerned about the test outcome. This effect of the interaction between warning and valence was demonstrated by McFarland (2000) when looking at the Big Five factor of openness. Performance is expected to be lowest in the group that has no coaching, no motivation. The warning statement is not expected to have any effect for this group, as they have nothing to lose nor gain by manipulating their scores. Table 2 shows these general expected score levels within the different situational conditions.

Table 2

Expected Score Levels

	Coached		Not coached	
	Warned	Not warned	Warned	Not warned
Motivated	Middle	High	Low	Middle
Not motivated	High	High	Low	Low

The following graph (Figure 2) provides a visual depiction of the expected interaction.

Figure 2

Expected In

Test Performance

H4:

The

know that co

and thus exp

H5. C

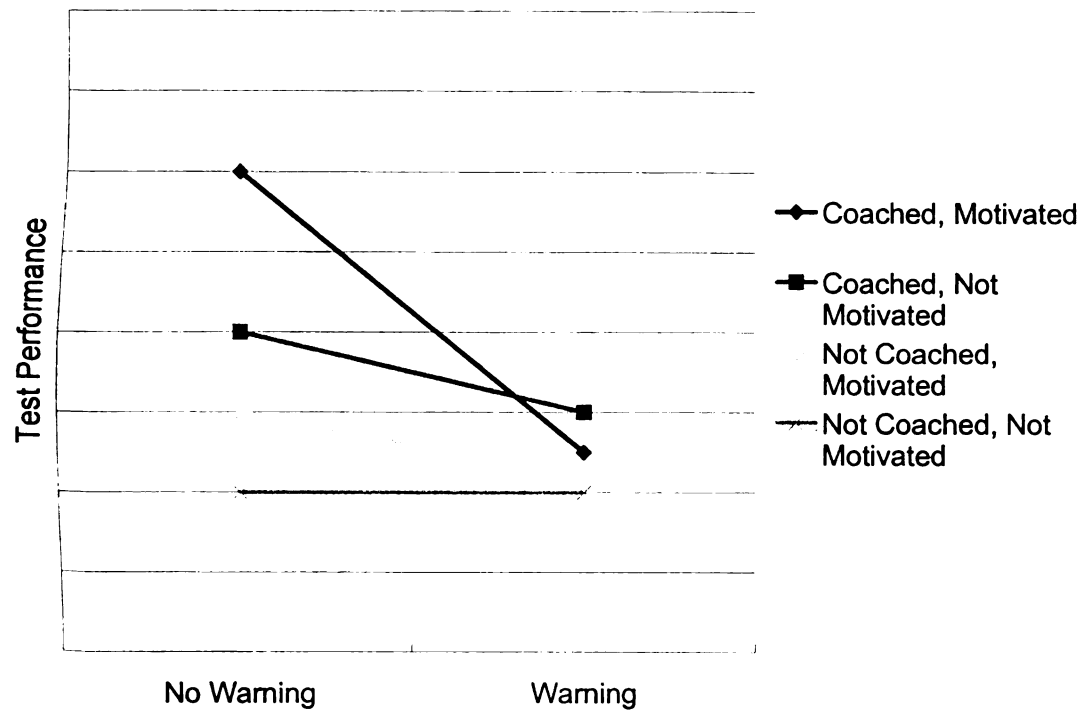
b

Based

and desired

Figure 2

Expected Interaction of Warning, Coaching, and Motivation on Test Performance



H4: Motivation, coaching, and warning will interact in affecting scores on biodata and situational judgment.

The second question of the study addresses individual differences in faking. We know that conscientiousness and emotional stability are generally related to performance, and thus expect these characteristics to relate to biodata and SJ performance.

H5: Conscientiousness and emotional stability will be related to performance on biodata and situational judgment.

Based on the findings that personality traits have been found to be associated with social desirability (Ones, et al., 1996) and faking (McFarland & Ryan, 2000), I posit that

consciousness

scores and

information. In

response to

performance

H6:

In the

as an individual

high items

possible

show extreme

described in

The

whether the

question is

who provide

information. It

objective and

relevant to the

more likely

interest of

conscientiousness and emotional stability will also be important correlates of inflated scores, and self-deception and impression management are also expected to be related to inflation. Inflation will be identified in three ways: by participants showing a positive response to bogus items, showing high scores on the inflation index or showing high performance on the BIDR impression management scale.

H6: Conscientiousness, emotional stability, and social desirability will be associated with inflation captured by performance on a bogus item scale, the inflation index, and impression management.

In this model, impression management is considered from two perspectives: first, as an individual difference that predicts biodata and SJI performance and inflation on bogus items and the inflation index, and second, as an outcome when used as an indicator of possible faking. (The empirical generation of the inflation index, where items that show extreme score differences between two manipulation groups in the study, is described in the *Method* section below.)

The third question of this study addresses characteristics of items that relate to whether the items are more or less likely to be susceptible to inflation. This research question is related to the work of Mael (1991), Becker and Colquitt (1992), and others, who provide insight into item characteristics that make items more vulnerable to score inflation. Items on which it is easier to provide an inflated response are items that are not objective and are difficult to verify (Becker & Colquitt, 1992). Items that are more relevant to the job (Becker & Colquitt, 1992) and overt (Alliger & Dwight, 2000) are more likely to be faked. I expect that items that are not viewed as relevant to the criterion of interest (here college performance) are less likely to be inflated as the link to desirable

modern

modern

modern

college

by and

children

receive

that are

see and

them. I

Mad. at

beyond

control

the you

that w

have" i

infected

known

H

academic behavior is difficult to make. Considering the examples of invasive and non-invasive items provided in Mael's (1991) taxonomy (p. 773), invasiveness is expected to reduce inflation. Consider Mael's noninvasive item "Were you on the tennis team in college?" Inflation is likely in an item such as this as there is little vulnerability created by answering positively. Now consider Mael's invasive item, "How many young children do you have at home?" This item, and others that are invasive, are likely to receive muted rather than extreme responses, and so are less likely to be inflated. Items that are perceived as unequal in access are less likely to be inflated as the individual may see inflated options as so far beyond the realm of possibility that they would not choose them. This argument is supported by the example of a nonequal access item provided by Mael, "Were you captain of the football team?" For most women, this item would be beyond what they would regard as possible. Similarly, an item that is considered controllable is more susceptible to inflation. Mael's example, "How many tries did it take you to pass the CPA exam?" is an example of a controllable item that is likely to be faked, while an uncontrollable item such as, "How many brothers and sisters do you have?" is less likely to be inflated. Logically, any item that is judged by people to be unfakeable is unlikely to prompt inflation, although this is a speculative hypothesis. The following hypotheses were tested.

H7: Biodata items that are rated as less objective, less verifiable, and more relevant to school work will show greater inflation than items that are rated as more objective, more verifiable, and less relevant to school work. Also, items that are rated as invasive, are outside the individual's control, are unequal in terms of access on the part of students, as well as items that are

1

A

relaxation

Shimizu

H

T

used in se

self-presen

performance

responses s

improved b

to predict pe

being indic

relation the

impression

encourage t

own selec

H9:

rated as unfakeable are less likely to show inflation that might be the result of some form of faking. (Greater inflation will be demonstrated by differences in the correlations between ratings of item types and mean responses to the items across study manipulation conditions.)

Another item type that has been shown to be relevant in reducing inflation is the requirement to elaborate on one's response to biodata items (Schmitt & Kuncze, 2002; Schmitt, et al., in press).

H8: Elaborated items will be less likely to be inflated than non-elaborated items.

The fourth question of the study addresses the validity of biodata and SJIs as tools used in selection. As inflation may not be an adaptive behavior, but rather an artificial self-presentation, it is likely that inflated responses are less effective predictors of performance than non-inflated responses (see Graham, et al., 2002). If non-inflated responses show criterion-related validity, then validity of inflated responses may be improved by correction. This possibility can be addressed by using regression analyses to predict performance, partialing out the effects of faking as measured by one or more faking indices (see White, et al., 2001). As mentioned earlier, I attempt to identify inflation through bogus items, an empirically constructed inflation index and the BIDR impression management scale. Some conditions are more likely than others to permit or encourage faking, and that will effect who would appear highest in a ranking in a top-down selection system (e.g. Dwight & Donovan, 2003; Frei, 1998).

H9: Controlling for responses to the three measures of faking will lead to a suppressor effect, with faking correlating with biodata and SJI but not with performance criteria. Statistically controlling for faking should thereby

increase the amount of variance in college performance outcomes (GPA and attendance) predicted by biodata and situational judgment.

More potential fakers (those marking bogus items, scoring high on impression management, and scoring high on the inflation index) will be identified among the top candidates if the faking goes uncorrected in top-down selection.

H10: Correction by removing those candidates who are identified as having inflated scores, and replacing them with the next-highest scoring individual not identified as having inflated scores, will result in a better choice of candidates, based on actual college performance (GPA and absenteeism).

inter-

series

upper

regard

to 35

of 11

This s

being

being

camp

of the

minor

Q. 10

and S

e H.

P. 1

made

the

paid

se t

METHOD

Sample

Michigan State University has a large student population (about 35,000 undergraduate students), and at the undergraduate level has relatively open admissions standards when compared with other universities. Because of this, I am provided the opportunity to sample a group of students in a situation in which I am not faced with significant range restriction in cognitive ability (ACT scores in this sample range from 8 to 35 with a mean of 23 ($N = 341$), and SAT scores range from 560 to 1480 with a mean of 1,100 ($N = 92$). National mean scores for 1999 were 21 for ACT and 1016 for SAT. This sample will allow us to consider more accurately the effects of general cognitive ability as they relate to our constructs of interest. Not only is there heterogeneity in ability in the student population, but also a diversity of ethnicities is represented on this campus. The ethnic breakdown of the student population is roughly comparable to that of the United States college applicant pool, although this campus underrepresents minority groups relative to the population of the US. At the university, 77.3 % are White, 9.8% are African American, 1.9 % are Hispanic American, 5.4 % are Asian American, and 5.6% Other. In the sample, 79.28 % are White, 6.08% are African American, 2.49 % are Hispanic American, 8.56 % are Asian American, and 5.6% Other. Of the sample 94% identified themselves as US citizens. Of the 5% who were non-citizens, 1% were Canadian, and 4% were of other citizenship. English was the primary language of 96% of the sample. Women account for 58.84% of the sample, and the campus has a student population that is 55% female. To ensure that we collected data from individuals who are close to the typical age of students applying for college, we recruited only a subgroup of

३२

۴۰۰

7

23

22

22

11

3

5

62

...

2

3.

2

2

1

the student population; to be able to participate, students were required to have been in their first year of college. The age of the participants ranged from 18 to 22 years, with 71% of the sample being 18, 20% being 19, and 6% being 20 years old. The mean age of the sample was 18. Participants were recruited from the university’s Psychology Department Subject Pool, had not participated in other studies using similar measures, and received extra course credit for their participation.

Study Design

This study is designed as a 2 X 2 X 2 orthogonal design, both for situational judgment and independently for biodata. The biodata part of the study has an additional factor with two levels: repeated measures on elaborated vs. non-elaborated items. The study manipulations are identified in Table 3 below, where there are eight possible variations in experience during the study, excluding elaboration, with equal numbers of participants in each cell.

Table 3

Study Manipulations and Participants per Cell

	Coached		Not coached	
	Warned	Not warned	Warned	Not warned
Motivated	43	45	45	45
Not motivated	44	48	46	46

There were approximately the same number of participants per cell in the research design, as shown in Table 3 above, and a total of 362 participants overall which provided sufficient statistical power to test for the main effects and interactive hypotheses



(Hypotheses 1 – 4). I am assuming a medium effect size ($d = .50$), and an alpha level of .05. With 40 participants per cell there will be a power level of .88. Samples within the cells will be too small to provide sufficient power to reject the null hypothesis for analysis of validity and reliability across conditions.

Procedure

The questionnaire for this study was split into two booklets, and four forms of the questionnaire were distributed. The four forms were essentially the same, apart from the instructional set that was used at the beginning of the second booklet. The instructional set provided the warning and motivation conditions. The first booklet of each form contained a Big Five personality measure, social desirability and impression management scales, college GPA, absenteeism and demographic questions. After completing the first booklet, those groups receiving coaching experienced a brief coaching session relating to the questions in the upcoming booklet, while those groups not targeted to receive coaching got nothing. There was no placebo treatment for the no coaching group; they were moved directly to the next phase of the experiment, which for both coached and non-coached groups was the second booklet. Instructional sets at the beginning of the second booklet varied across the four forms, providing the study manipulations. A sample of the questionnaire form is shown in Appendix B, although biodata and situational judgment items have been removed, due to the proprietary nature of the measures. Sample biodata and situational judgment items are shown in Appendix C and D, respectively. Apart from the instructional sets, forms were identical. The wording of all four of the instructional sets is shown in Appendix E.

Human

method

Aggen

collect

advent

Particip

rehab

psycho

Particip

entire

Impor

begin

Engel

ending

process

and sec

the pro

the p

three m

three o

the pro

This study was approved by the University Committee on Research Involving Human Subjects (UCRIHS). Samples of the two informed consent forms, one for motivated participants and one for participants who were not motivated, are shown in Appendix F and G respectively. The forms for this project were pilot tested, and data collection was completed in Fall 2002 and Spring 2003 semesters. The study was advertised through the web page of the Psychology Department subject pool.

Participation was restricted to freshman students who had not participated in other studies we have conducted using similar measures. Participants were offered extra credit in psychology for their participation, and the data collection sessions lasted 90 minutes. Participants, who signed up via the subject pool web site, were randomly assigned to different coaching conditions, and were randomly assigned to different forms. It is important in attempting to understand the effects of coaching through an experimental design that subjects be randomly assigned to the coaching condition (Messick & Jungeblut, 1981). Sessions were designed to seat up to 50 participants in a classroom setting, and were administered by research assistants working according to a written protocol (Appendix H). Informed consent forms, data release forms, questionnaire forms and scantrons were placed in envelopes for each participant and distributed according to the protocol. Half of the sessions were provided with a ten-minute coaching component after participants had completed the first booklet, containing all measures apart from three measures: biodata items, bogus items and situational judgment questions. These three components were completed immediately after coaching for the coached group, and the presentation of the manipulated instruction sets.

post

not

state

The

even

It

has

set

the

in

It

the

is

No

the

Co

the

the

Manipulations

Rather than conducting a directed faking study, where individuals are directed to provide the best response even if that means lying, I have preferred to create situational motivators in the form of an external cash incentive, provide constraints by using warning statements, and present ideas on what the tests are capturing through a coaching session. These are a more realistic and subsequently useful approach than directed faking in examining inflation on biodata and SJIs.

Motivation

To encourage students in the motivated conditions to do their very best on the test, and to make them most similar to motivated candidates in a college admission setting, a financial incentive was offered. Those who scored above the 50th percentile on the tests administered were promised and have been mailed \$10. Those in the non-motivated conditions received no incentive.

Warnings

To create a similar effect to warnings that may appear on college application materials, the materials for those in the warning conditions included warning statements as shown in the instruction sets in Appendix E. An example of a warning statement is: “Note that we may verify a subset of your responses, and if you respond dishonestly, that may invalidate this test as well as your chance to receive \$10 for high performance.”

Coaching

A ten-minute coaching component reviewed sample biodata and situational judgment items along with definitions of the performance dimensions that the questions were designed to measure. This coaching session was designed as an orientation to these

1

particular selection devices. This form of orientation is common in formal coaching for selection tests (Sackett, et al., 1989), and is similar in length to the coaching for biodata provided in Miller (2001). Also, it is far more comprehensive than the written coaching statements found by Cunningham, et al. (1994) to be effective. We expected that this brief coaching would have an effect because it was exercise-specific (Sackett, et al.), and was provided immediately prior to the completion of the biodata and situational judgment inventory questionnaire.

By reading aloud the directions (Appendix I) and handout material (Appendix J), the proctor of the session provided coaching for the participants. To ensure uniformity in coaching, the same proctor administered all coaching sessions, reading from a prepared script. Once the coaching component was complete, the proctor led immediately into administering the second form in the study, to avoid any discussion and questions during the testing session regarding the coaching.

Measures

College Grade Point Average

To be able to evaluate the predictive validity of the measures being tested in this study, two outcomes were collected. Participants were asked to release their actual college GPA. A sample of this information release form provided to the university registrar is shown in Appendix K.

Absenteeism

Participants were also asked to identify how frequently they have missed class, and the reasons for their absence.

Gen.

no inc.

High

score

relax

vari

mean

index

Person

PIP

Cons.

Extra

Stab

Social

meas.

1983

live s

negat

St. Co.

Gender and Ethnicity

To be able to examine subgroup differences in responses, participants were asked to indicate their ethnicity and gender.

High School GPA and SAT/ACT Scores

Participants were asked to provide access to their high school GPA and SAT/ACT scores as provided by the university admissions office. A sample of the information release form is shown in Appendix L. SAT and ACT scores were converted to new variables through linear transformation based on national normative information on means and standard deviations, and then combined to create a composite cognitive ability index consisting of an average of all available test scores for each person.

Personality

A Big Five personality inventory based on the International Personality Item Pool (IPIP) was used to measure personality (Goldberg, 1999). Scale alpha levels are Conscientiousness (.81), Openness (.77), Agreeableness (.82), Emotional Stability (.89), Extraversion (.87). I presented hypotheses only for the Conscientiousness and Emotional Stability; other measures were examined on an exploratory basis.

Social Desirability

Self-deception and impression management dimensions of social desirability were measured using the Balanced Inventory of Desirable Responding (BIDR) scale (Paulhus, 1988). Because of concerns about the intrusive nature of one item in each scale (i.e., “I have sometimes doubted my ability as a lover” and “I never read sexy books or magazines”), these two items were not used. Scale alphas are .67 for self-deception and .80 for impression management.

Biodata

Biodata items generated by Oswald, et al., (in press) were reviewed, and those that were empirically determined in their sample of first-year students to be the best predictors of college performance outcomes (GPA, absenteeism, and a self-assessment on a behaviorally anchored rating scale) were selected for this battery. While Oswald, et al. had written their items to tap 12 dimensions of student performance (see Appendix M), as the items were intercorrelated, they regarded their biodata scale as unidimensional. Nevertheless, to capture the breadth of student performance, items for this study were selected to ensure that the content of all twelve performance dimensions was addressed in the scale. Also included were all elaborated items used by Oswald et al., so that half of the items for this study were elaborated and half were not. Overall biodata scale alpha was .88. Alpha for elaborated items was .78 and .80 for non-elaborated items. Due to the proprietary nature of these items, a sample of the biodata items is provided in Appendix C.

Bogus Items

The biodata questions include four bogus items to assess faking. These items were based on bogus items used by Anderson, et al. (1984), and were interspersed with the real biodata items. The bogus items are identified in Appendix N. The bogus items scale alpha was .37. This was not unexpected as these items produced very little variance. If respondents were paying attention and were honest, I did not expect them to indicate any activity on these four items when there is no incentive to inflate responses.

Sum

and

per

and

per

on

the

Due

the

the

between

the

the

the

the

the

the

the

the

the

Situational Judgment

Situational judgment items generated by Oswald, et al., (in press) were reviewed, and those that were empirically determined to be the best predictors of college performance outcomes (GPA, absenteeism, and a self-assessment on a behaviorally anchored rating scale) were identified for this battery. On dimensions where the predictive validity of the SJI items was low, the best items were rationally selected based on content. Two items per dimension were selected to ensure that the content of all twelve performance dimensions was addressed in the overall scale. Scale alpha was .77. Due to the proprietary nature of these items, a sample of situational judgment items is shown in Appendix D.

Inflation Index

To be able to identify probable fakers, the mean score differences for items between the individuals in the manipulation condition where people are most apt to inflate their scores (coached, motivated, not warned) and in the manipulation condition where people would be least apt to inflate their scores (no coaching, no motivation, warned) were examined. Eight items with the largest mean difference in responses comprised the inflation index. These items are shown in Appendix O, with the difference scores in Appendix P.

Item Type

To generate assignments of item type to biodata items, two professors, four graduate students, and three undergraduate research assistants on the project provided ratings indicating the degree to which each biodata item was: objective, verifiable, controllable, equally accessible, relevant to college, invasive, and fakeable. Ratings were

made o

era S

dimens

assesse

Attitud

19... w

4 peop

consist

interch

agges

four re

ratings

college

dimens

dimens

Alfred

Items,

for gr

Item T

below

made on a 5-point scale ranging from “Not at all” to “Completely,” one dimension at a time. See Appendix Q for the rating form, which includes definitions of the item-type dimensions. The 126 items were distributed evenly among the raters so that each rater assessed a unique set of 56 different items. In the end, this resulted in 4 ratings per item. Additionally, four sets of 14 items were rated by the same set of 4 raters. (Items 1, 10, 19... were rated by the same people; items 2, 11... were rated by another set of the same 4 people, etc.)

In order to index the amount of agreement between these ratings, the internal consistency of the ratings across items was calculated, treating each rater as interchangeable. This analysis was conducted for each item-type dimension by aggregating ratings for the 126 items into four groups that consisted of ratings from the four, randomly assigned raters. The coefficient alpha estimates for the four groups of ratings were then examined for each dimension. Alpha coefficients were highest for college relevance, verifiability, objectivity, and controllability (see Table 4), and these dimensions were retained as relevant for further analysis. Low reliability on some dimensions was either a result of all the judges rating all the items the same way (e.g, fakeable, where all items were regarded as highly fakeable, with no variability across items), or inconsistency in how judges rated items (e.g., invasive).

For objectivity, verifiability, relevance, and controllability the ratings across the four groups of ratings were averaged to compute an overall dimension value for each item. These values were then used in item analyses, the results of which are described below.

Table 4

Control

Dimension

Objective

Variance

Control

Equality

College

Invasive

Fakeab

S

Non-type

Students

Response

College

Staff

Disturb

Time

Dimension

Objective

Dimension

1. We

Table 4

Coefficient Alpha and Descriptives for Ratings of Biodata Item Characteristics

Dimension	Alpha	<i>N</i>	Min.	Max.	Mean	<i>SD</i>
Objective	0.70	126	1.00	5.00	2.89	0.93
Verifiable	0.75	126	1.00	4.75	2.49	0.91
Controllable	0.67	126	1.75	5.00	3.97	0.77
Equal access	0.37	126	2.5	5.00	4.11	0.62
College relevant	0.80	126	1.25	5.00	3.28	0.94
Invasive	0.06	126	1.00	4.00	2.75	0.73
Fakeable	0.06	126	2.50	5.00	4.26	0.49

Similarly, analyses of SJI item type were conducted, and rating instructions and item-type definitions are shown in Appendix R. Two professors and five graduate students provided ratings indicating the degree to which each situational judgment item response option was: objective, verifiable, controllable, equally accessible, relevant to college, invasive, and fakeable. Ratings were made on a 5-point scale ranging from “Not at all” to “Completely,” one dimension at a time. The 24 situational judgment items were distributed to all raters, with each rater rating all items. To assess agreement between these raters, we measured the internal consistency of the ratings across items, for each dimension. The alpha coefficients were highest for college relevance, verifiability, objectivity, and controllability (see Table 5), and we retained the data for these dimensions that demonstrated traditionally accepted levels of internal consistency. Again, we discarded ratings on categories where the raters provided little variance across

Items (e.g.,

judges view

across the se

values were

inflation.

Table 5

Coefficient

Dimension

Objective

Verifiable

Controllable

Equal acc

College re

Invasive

Fakeable

Th

section of

items (e.g., fakeable, equal access) or were inconsistent in rating (e.g., invasive). The judges viewed biodata items as more fakeable than SJI items. Next, we averaged ratings across the seven individual raters' values to compute an index for each dimension. These values were then used in item analyses to identify items viewed as most vulnerable to inflation.

Table 5

Coefficient Alpha and Descriptives for Ratings of SJI Item Characteristics

Dimension	Alpha	<i>N</i>	Min.	Max.	Mean	<i>SD</i>
Objective	0.81	155	1.43	4.71	3.09	0.70
Verifiable	0.83	155	1.29	4.71	2.86	0.76
Controllable	0.78	155	1.71	4.86	3.67	0.67
Equal access	0.69	155	2.14	5.00	4.30	0.49
College relevant	0.86	155	1.86	5.00	3.77	0.91
Invasive	-0.50	155	2.00	3.43	2.76	0.31
Fakeable	0.17	155	2.86	4.71	3.66	0.37

The results of the use of these measures are described in the *Item Differences* section of the results below.

To
situational
motivation
biocata and
to elaborat
shown in T
various pos
white), sel
measures o
of respons
As
statistical
predicted
nonsignifi
factors w
For the an
Coaching
initiated
used
hing
from

RESULTS

Situational Differences

To address the first question of our study, and test Hypothesis 1-4, regarding the situational factors that affect faking, I conducted a 2 (coaching vs. no coaching) X 2 (motivation vs. no motivation) X 2 (warning vs. no warning) ANOVA separately for the biodata and SJI measures. The biodata ANOVA had an additional factor (elaboration vs. no elaboration on the items), which was a within-subjects factor. Results for biodata are shown in Table 6 for an analysis that included no covariates and an analysis that included various possible covariates of biodata responses including sex, race (minority versus white), self deception, impression management, cognitive ability, high school GPA, and measures of the Big Five constructs. Table 7 contains the means and standard deviations of responses to the biodata for all conditions.

As can be seen in Table 6, the motivational effect and the coaching effect are both statistically significant and the means in Table 7 indicate that the effect was in the predicted direction thus confirming our first two hypotheses. The Warning effect was nonsignificant indicating lack of support for Hypothesis 3. I hypothesized that these factors would interact, but did not observe the three way interaction depicted in Figure 2. For the analysis that did not include covariates, the interaction between Motivation and Coaching was marginally significant. Examination of the means for these conditions indicated that the combination of both Motivation and Coaching without Warning produced the largest biodata scores (Mean =3.41) as would be expected. Neither Coaching nor Motivation alone (Means =3.09 and 3.06) produced as large an increment *in* performance when neither Motivation nor Coaching were provided. Variance did

change across

not Warned

that people

When

factors inter

for both and

would be ve

monetary m

mediated c

compared t

mediated, i

One

Examination

elaborated

between Ex

elaboration

analyses.

change across manipulation conditions, with the greatest variability in the Coached and not Warned groups, suggesting that coaching does not necessarily standardize the way that people respond.

When covariates were included in the analyses, the Motivation and Warning factors interaction was also statistically significant. Examination of the pattern of means for both analyses (with and without covariates) indicated that a warning that responses would be verified did appear to erase the inflation of responses that occurred when a monetary motivation to get good scores was provided. With a warning, the means of the motivated (Mean=3.11) and nonmotivated groups (Mean=3.01) were not very different compared to the two conditions in which no warning was present (Means= 3.24 for motivated, and 2.99 for nonmotivated).

One of the covariates (Extraversion) interacted with the elaboration factor. Examination of this interaction indicated that the correlation between responses to elaborated biodata questions and Extraversion (.29) was higher than the correlation between Extraversion and the responses to nonelaborated items (.21). The impact of elaboration also appeared much smaller when all covariates were included in the analyses.

Table 6

Analysis of

Between Subjects

Coaching (C)

Motivation

Warning (W)

C x M

C x W

M x W

C x M x W

Error

Within Subjects

Elaboration

Ex C

Ex M

Ex W

Ex C x M

Ex C x W

Ex M x W

Ex C x M

Table 6

Analysis of Variance Results for Biodata with and without the Inclusion of Covariates

<u>Between Subjects Source</u>	<u>Without Covariates</u>		<u>With Covariates</u>	
	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>
Coaching (C)	1	24.95**	1	26.96**
Motivation (M)	1	16.73**	1	13.62**
Warning (W)	1	.39	1	.54
C x M	1	3.69	1	1.52
C x W	1	.02	1	1.92
M x W	1	2.11	1	6.28*
C x M x W	1	.03	1	1.22
Error	354	(.42) ^a	312	(.32) ^a
<u>Within Subjects</u>				
Elaboration (E)	1	726.11**	1	3.38
E x C	1	.60	1	.60
E x M	1	.07	1	.25
E x W	1	.03	1	.07
E x C x M	1	2.32	1	1.62
E x C x W	1	.72	1	.01
E x M x W	1	.08	1	.27
E x C x M x W	1	.63	1	1.33

Table 6 Co
Within Sub

Ex Self: De

Ex Impres

Ex Extrov

Ex Agreea

Ex Conser

Ex Emoti

Ex Openn

Ex High S

Ex Cogni

Ex Race

Ex Sex

Error

*Equals t

*p < .05

Table 6 Cont.

Within Subjects

<u>Source</u>	<u>Without Covariates</u>		<u>With Covariates</u>	
	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>
E x Self Deception			1	.89
E x Impression Management			1	.09
E x Extroversion			1	4.81*
E x Agreeableness			1	3.80
E x Conscientiousness			1	.95
E x Emotional Stability			1	.13
E x Openness			1	.81
E x High School GPA			1	2.14
E x Cognitive Ability			1	.56
E x Race			1	.21
E x Sex			1	.34
Error	354	(0.05) ^a	312	(0.05) ^a

^aEquals the Mean Square Error.

*p < .05. **p < .01.

Table 7

Means and Standard Deviations of Biodata across Conditions

	Elaborated Biodata		Nonelaborated Biodata		Combined Biodata Total		<i>N</i>
	Mean	<i>SD</i>	Mean	<i>SD</i>	Mean	<i>SD</i>	
Mot (0), Warn (0) Coach (0) ^a	2.67	0.37	3.13	0.45	2.88	0.37	46
Mot (0), Warn (0) Coach (1)	2.84	0.53	3.28	0.53	3.09	0.51	48
Mot (0), Warn (1) Coach (0)	2.71	0.45	3.20	0.49	2.94	0.45	46
Mot (0), Warn (1) Coach (1)	2.86	0.40	3.32	0.39	3.08	0.31	44
Mot (1), Warn (0) Coach (0)	2.88	0.43	3.28	0.40	3.06	0.37	45
Mot (1), Warn (0) Coach (1)	3.14	0.67	3.69	0.62	3.41	0.63	45
Mot (1), Warn (1) Coach (0)	2.75	0.42	3.21	0.47	2.95	0.40	45
Mot (1), Warn (1) Coach (1)	3.08	0.48	3.56	0.59	3.26	0.49	43

^aMot, Warn, Coach indicates Motivation, Warning, and Coaching. One indicates the manipulation occurred, 0 indicates there was no manipulation.

In 7

without th

standard d

was simil

effects of

Table 9) v

significant

significant

warning p

two cond

in Table

Warning

the warn

condition

Warned

included

way eve

agreeab

signific

paper.

In Table 8, I present the ANOVA results for the SJI measure both with and without the inclusion of the covariates. Table 9 provides the corresponding means and standard deviations for the conditions in our study, and the variance of the SJI responses was similar across all manipulation conditions. As was true for biodata responses, the effects of Motivation and Coaching were statistically significant and the means (see Table 9) were in the expected direction. The Warning effect was not statistically significant. One interaction, that of Motivation and Coaching, was statistically significant but the results were not consistent with expectations in that the presence of warning produced larger SJI scores than no warning, the difference being larger in the two conditions that did not receive the motivational manipulation. The means presented in Table 9 are consistent with our expectations in that the Motivated, Coached, No Warning condition produced the best SJI responses, over a standard deviation higher than the warned condition that was not coached or motivated. However, several other conditions did not fit our expected pattern of results. The Coached, Motivated and Warned group performed as well as did a similar group with no warning. Results that included the covariates did not change the impact of the manipulations in any substantial way even though a number of covariates (self deception, impression management, agreeableness, conscientiousness, high school GPA, sex and race) were related significantly to SJI responses. I will address these relationships in more detail later in the paper.

Table 8

Analysis

Between

Coach

Motiva

Warmi

C x M

C x W

M x W

C x M

Error

Self De

Impress

Extrave

Agreeab

Conscie

Emoti

Open

High S

Table 8

Analyses of Variance for SJI with and without Covariates

<u>Between Subjects</u>	<u>Without Covariates</u>		<u>With Covariates</u>	
	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>
Coaching (C)	1	40.60**	1	56.39**
Motivation (M)	1	14.65**	1	10.06*
Warning (W)	1	1.41	1	2.41
C x M	1	9.88*	1	6.25*
C x W	1	.22	1	1.36
M x W	1	.09	1	.91
C x M x W	1	.10	1	1.20
Error	352	(0.12) ^a		
Self Deception			1	5.45*
Impression Management			1	4.00*
Extraversion			1	2.41
Agreeableness			1	4.19*
Conscientiousness			1	4.71*
Emotional Stability			1	1.29
Openness			1	1.52
High School GPA			1	4.79*

Table 8

Cognitive

Race

Sex

Error

^aEquals

*p < .05

Table 9

Means

Mot (0).

Mot (0).

Mot (0).

Mot (0).

Mot (1).

Mot (1).

Mot (1).

Mot (1).

^aMot.

manip

Table 8 cont.	<u>Without Covariates</u>		<u>With Covariates</u>	
Cognitive Ability			1	.32
Race			1	3.96*
Sex			1	18.44**
Error			310	(0.09) ^a

^aEquals the Mean Square Error.

* $p < .05$. ** $p < .01$.

Table 9

Means and Standard Deviations of SJI Responses for Various Study Conditions

	Mean	SD	N
Mot (0), Warn (0) Coach (0) ^a	.50	.32	46
Mot (0), Warn (0) Coach (1)	.62	.37	48
Mot (0), Warn (1) Coach (0)	.56	.32	46
Mot (0), Warn (1) Coach (1)	.67	.37	44
Mot (1), Warn (0) Coach (0)	.52	.33	44
Mot (1), Warn (0) Coach (1)	.89	.32	45
Mot (1), Warn (1) Coach (0)	.58	.35	44
Mot (1), Warn (1) Coach (1)	.89	.33	43

^aMot, Warn, Coach indicates Motivation, Warning, and Coaching. One indicates the manipulation occurred, 0 indicates there was no manipulation.

and s

supp

with

in the

score

that i

colle

coach

admi

some

of res

most

who h

altho

word

losing

group

Dave

the e

Hypothesis 1, that motivation will increase scores, was supported for both biodata and situational judgment. Hypothesis 2, that coaching would increase scores, was also supported for biodata and situational judgment. Support for Hypothesis 3 was not found, with warning being ineffective at reducing scores. Hypothesis 4 was partially supported in the case of biodata, but not for situational judgment responses.

It is apparent from these findings that both biodata and situational judgment scores are susceptible to coaching, as well as to motivation. It is important to consider that in this study, very brief coaching was provided, and if such tests were to be used in college admissions decision-making, it is reasonable to expect that more comprehensive coaching would become available as a result of the high-stakes nature of college admissions decisions. The results do indicate, though, that a warning statement can have some impact in reducing inflation at least for the biodata. An examination of the pattern of results for biodata (Figure 3) shows that the suppressor effect of a warning statement is most powerful for those who are motivated, lending some support to the idea that those who have something to lose are more likely to take a warning statement seriously, although this would need to be verified with further research.

A limitation of the warning manipulation is the slight difference between the wording for motivated and not motivated groups, where the motivated groups risked losing the opportunity to earn extra cash if caught responding dishonestly, while the group without a cash incentive did not suffer this risk. According to the recent work of Dwight and Donovan (2003), potential consequences in faking are an important factor in the effectiveness of warning statements.

Figure

Inter

3.2

3.1

3

2.9

2.8

2.7

2.6

2.5

2.4

suspic

this m

includ

were

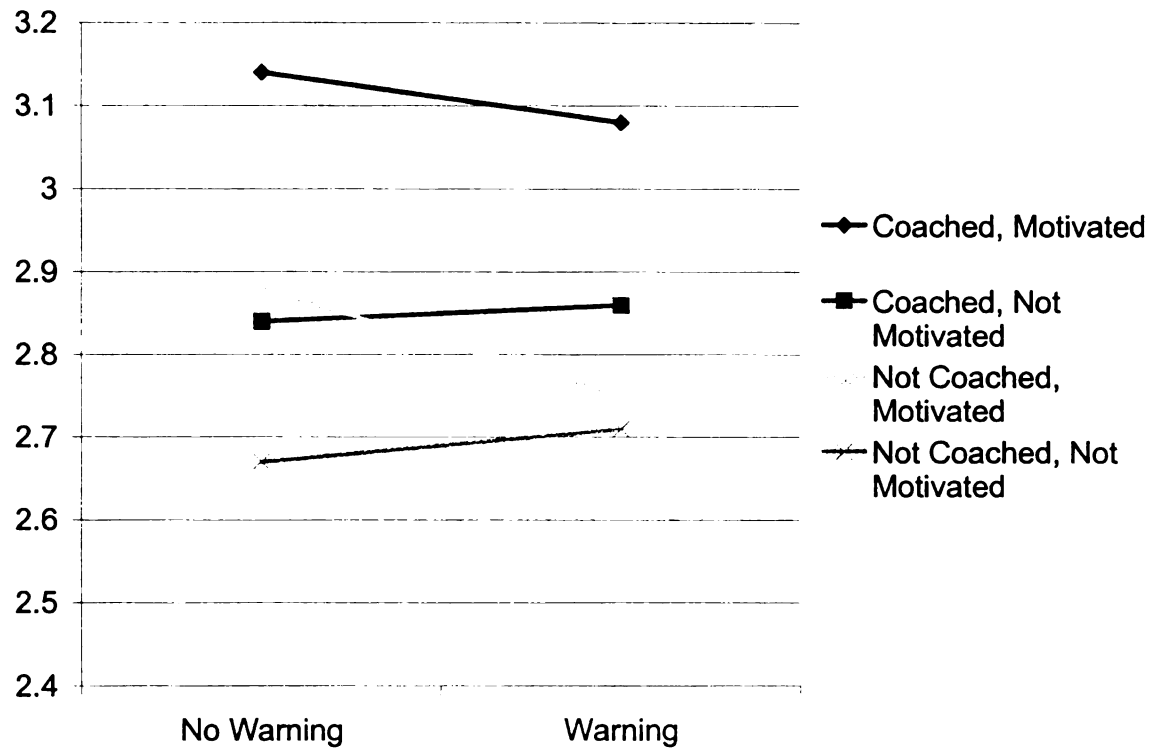
this m

were

of or

Figure 3

Interactions between Coaching, Motivation, and Warning for Biodata Performance



It is also possible that for someone who had not received a warning, their own suspicions about a bogus item may have raised the possibility that was a lie scale, and this may have effectively operated as a warning. It would be helpful in future research to include manipulation checks so that it is clear whether participants registered that they were being warned.

In this study, motivation was generated by an offer of a small amount of cash, and this incentive was sufficient to influence score inflation. If biodata and SJI performance were used to contribute to admissions ratings, the importance of admission to the college of one's preference would probably be a more powerful motivator, resulting in significant

score infl.

artificial c

may have

direction

T

examine

minority

(specific

analyses

showing

or 3) a h

identific

inflation

manage

$r=.36$

Person

openn.

$r=.22$

signifi

impro

stabi

score inflation. The risk of not gaining admission would be an important risk. The artificial context of the warnings in the study combined with the monetary motivation may have limited the effectiveness of warnings, but the effect was in the expected direction when the interaction of Motivation and Warning was examined for biodata.

Individual Differences

To address the second question of the study, test Hypotheses 5 and 6, and examine the impact of individual difference correlates of faking – race (white vs. minority), gender, ability (measured by ACT/SAT scores) and personality characteristics (specifically conscientiousness, emotional stability, and social desirability) – correlational analyses were conducted. Faking may be captured in three ways: by an individual showing 1) a positive response to the bogus items, 2) a high score on an inflation index, or 3) a high score on the BIDR impression management scale.

Table 10 shows the correlation matrix for the variables of interest. Of the faking identification methods, scoring on the bogus items was associated with scoring on the inflation index ($r=.38$), but neither of these was strongly related to impression management. The bogus items and inflation index were related to coaching ($r=.23$ and $r=.36$, respectively), and the inflation index was also related to motivation ($r=.24$). Personality was not highly associated with inflation as indexed by bogus items, with only openness having a significant relationship ($r=.16$). For the inflation index, extraversion ($r=.22$), agreeableness ($r=.21$), openness ($r=.21$), and self-deception ($r=.22$), showed significant relationships. However, personality was more strongly associated with impression management; conscientiousness ($r=.36$), agreeableness ($r=.33$), emotional stability ($r=.22$), and self deception ($r=.39$) were correlated significantly with impression

man.

abili

relate

outco

(r=

brod.

relati

(r=.3

relate

conso

other

conso

were

was re

to eith

to the

absen

relati

reg-

(r=

management. Methods of identifying inflation appeared to be unrelated to race, cognitive ability, and age, apart from impression management, for which age was negatively related ($r=-.12$). The only relationship between inflation identification scales and outcomes was for impression management, which was negatively related to absenteeism ($r=-.23$).

Biodata scores were associated with situational judgment scores ($r=.47$), and both biodata and SJI scores were related to multiple personality traits, however, the patterns of relationships were slightly different. Biodata scores were most highly related to openness ($r=.33$), self deception ($r=.27$), extraversion ($r=.26$) and agreeableness ($r=.25$). Also related to biodata were agreeableness ($r=.25$), impression management ($r=.16$), conscientiousness ($r=.15$) and emotional stability ($r=.13$). Situational judgment, on the other hand, was most closely related to impression management ($r=.29$), conscientiousness ($r=.24$), and self-deception ($r=.22$). Also related to SJI performance were agreeableness ($r=.21$), extraversion ($r=.16$) and openness ($r=.12$). Cognitive ability was related to biodata performance ($r=.18$), but not SJI performance. Race was unrelated to either biodata or SJI performance.

Both race ($r=.20$) and ability ($r=.37$) were related to GPA. Biodata was unrelated to the outcomes of absenteeism and GPA, and SJI performance was negatively related to absenteeism ($r=-.20$). Age was positively related to absenteeism ($r=.17$) and negatively related to GPA ($r=-.16$). Conscientiousness was positively related to GPA ($r=.13$) and negatively related to absenteeism ($r=.37$). Emotional stability ($r=-.11$) and self-deception ($r=-.16$) were also negatively related to absenteeism.

Table 10
Correlation Matrix

	Bogus Scale	Inflation Index ^a	IM	Biodata	SJI	Race	Cognitive Ability	Age	E	A
Inflation Index	0.38**									
IM	-0.03	0.19								
Biodata	0.48**	0.79(0.74)**	0.16**							
SJI	0.18**	0.60(0.55)**	0.29**	0.47**						
Race	0.04	0.06	-0.04	0.06	-0.04					
Ability	0.07	0.13	0.02	0.18**	0.03	0.21**				
Age	-0.02	-0.07	-0.12*	-0.07	-0.07	0.09	-0.16**			
E	0.04	0.22**	0.02	0.26**	0.16**	0.07	0.03	-0.13*		
A	0.07	0.21**	0.33**	0.25**	0.21**	0.05	0.05	-0.13*	0.30**	
C	0.04	0.10	0.36**	0.15**	0.24**	-0.01	-0.12*	-0.07	-0.05	0.12*
ES	0.03	0.12	0.22**	0.13*	0.07	0.01	0.02	-0.03	0.17**	0.18**
O	0.16*	0.21**	0.04	0.33**	0.12*	0.09	0.3**	-0.12*	0.29**	0.20**
SD	0.06	0.22**	0.39**	0.27**	0.22**	-0.13*	-0.06	-0.04	0.25**	0.14*
Motivated	0.08	0.24**	0.01	0.20**	0.19**	-0.03	-0.05	-0.04	-0.02	-0.03
Warned	-0.03	-0.01	0.08	-0.03	0.05	-0.00	-0.08	-0.08	-0.01	0.00
Coached	0.23**	0.36**	-0.03	0.25**	0.31**	0.09	0.04	0.12*	0.04	-0.05

Table 10 continued

	Bogus Scale	Inflation Index ^a	IM	Biodata	SJI	Race	Cognitive Ability	Age	E	A
Absenteeism	0.01	-0.03	-0.23**	-0.05	-0.20**	-0.01	0.07	0.17**	0.04	-0.05
GPA	0.06	0.07	0.08	0.09	0.06	0.20**	0.37**	-0.16**	-0.01	0.04
ES	0.11*									
O	-0.06	0.01								
SD	0.31**	0.45**	0.22**							
Motivated	0.00	-0.04	0.01	0.04						
Warned	-0.07	0.05	0.00	-0.03	0.01					
Coached	-0.02	0.07	0.09	0.02	-0.01	-0.02				
Absenteeism	-0.37**	-0.11*	0.03	-0.16**	-0.00	0.02	0.09			
GPA	0.13*	0.07	0.09	-0.03	-0.07	-0.07	-0.06	-0.20**		

^aBiodata and SJI correlations with inflation index shown in parentheses indicate coefficient when overlap items from inflation index have been removed from biodata and SJI scales.

*p < .05. **p < .01.

cons

whil

part

mar

to e

imp

and

dr

ex

po

bo

th

s-

th

st

or

ex

st

Hypothesis 5 was supported for biodata and partially supported for SJI. Both conscientiousness and emotional stability were associated with performance on biodata, while only conscientiousness was related to SJI performance. Hypothesis 6 was also partially supported. Conscientiousness and emotional stability were related to impression management, yet performance on the bogus items and the inflation index were not related to either of these personality traits. Self-deception was related to the inflation index and impression management.

These results provide support for the idea that positive responses to bogus items and inflating one's score on other items is less personality-related and more situation-driven, whereas the tendency toward impression management is more personality-related.

It is apparent from this study that individuals do claim experience with non-existent things, and that this scoring on bogus items is weakly related to openness. It is possible that people are being more "open" about how they interpret these particular bogus items, and may be drawing parallels between an experience that they have had and that which is captured by the bogus item, generously deciding to claim that they have suitable enough experience to identify it as such. However, as bogus item scores were more strongly associated with coaching, I expect that situational determinants have a more powerful effect on marking bogus items, where individuals who have had coaching on what dimensions are desirable may be overly enthusiastic about demonstrating experience on items that they see as matching those dimensions, rather than responding in an honest fashion. While it may be ethically questionable to include bogus items in a

ca

P

pe

su

N

re

co

in

no

the

usa

sel

net

on

ex

in

ad

de

s

w

college admissions application, this does provide useful information about the ways that people may fake.

As the inflation index used here is created empirically, based on the difference in performance between the optimal performance group and the reference group, it is not surprising that the index is very highly correlated with biodata and SJI performance. Nevertheless, once overlap items are removed from biodata and SJI scales, the index remains highly correlated with performance. It is interesting to note that conscientiousness, although significantly related to biodata and SJI performance, is unrelated to the index. As the inflation index is related to coaching, this reinforces the notion that situational factors are important in explaining faking. It should be noted that the inflation index created empirically for this study would require cross-validation for use elsewhere. (The items that made up the index can be viewed in Appendix O.)

That personality traits of conscientiousness and emotional stability were related to self-deception is consistent with research on social desirability, and it is not surprising to note that agreeableness is also related to self-deception – it is comforting to think of oneself as agreeable. These three personality traits appear to be the most important in explaining moderately inflated performance, however, they do not appear to be as useful in cases of extreme faking. This suggests that presenting oneself in a positive light is adaptive, and that items to which individuals respond probably include a level of social desirability and/or job desirability.

That cognitive ability was related to biodata but not SJI performance was surprising, but may be explained by the breadth of dimensionality captured by the SJI, which moves beyond typical college academic issues to include issues of social

co

w

ch

ch

ch

w

tr

co

Ta

le

T

O

B

D

C

M

S

st

consciousness, multicultural tolerance, artistic appreciation, etc. That cognitive ability was weakly related ($r = .18$) to biodata could be a result of socioeconomic status, where those with ability were also exposed to greater opportunity for suitable experiences.

Item Differences

To address the third question of the study, and examine the effects of certain item characteristics as they affect inflation, the following analyses were conducted. The item characteristic ratings made up of the mean rating provided by experts (described earlier) were correlated with the biodata item responses for each participant. The r was then transformed to a Fisher z . ANOVA was used to test for differences in Fisher z correlations across the study manipulation conditions. Overall mean levels are shown in Table 11, ANOVA results for the four item characteristics are shown in Table 12, and levels across conditions are shown in Table 13.

Table 11

Overall Means and Standard Deviations of Four Item Characteristic Fisher z for Biodata

Dimension	N	Mean	SD
Objectivity	362	-0.22	0.21
Verifiability	362	-0.18	0.21
Controllability	362	-0.15	0.16
Relevance	362	0.10	0.16

Table 12

Analysis of Variance Results for Fisher z for Four Item Characteristics for Biodata

<u>Between</u> <u>Subjects Source</u>	Objectivity		Verifiability		Controllability		Relevance	
	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>
Coaching (C)	1	7.08**	1	2.35	1	7.27**	1	0.99
Motivation (M)	1	5.52*	1	1.46	1	5.61*	1	0.52
Warning (W)	1	0.18	1	0.00	1	0.02	1	0.80
C x M	1	0.19	1	0.10	1	0.47	1	4.39*
C x W	1	0.50	1	0.21	1	0.08	1	2.09
M x W	1	0.21	1	0.27	1	0.14	1	0.00
C x M x W	1	0.11	1	0.12	1	0.00	1	0.35
Error	354	(.04) ^a	354	(.04) ^a	354	(.03) ^a	354	(.02) ^a

^aEquals the Mean Square Error.

*p < .05. **p < .01.

Table 13

Means and Standard Deviations of Four Item Characteristic Fisher z for Various Study Conditions for Biodata

	Objectivity		Verifiability		Controllability		Relevance	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Mot (0), Warn (0) Coach (0) ^a	-0.27	0.19	-0.22	0.21	-0.19	0.13	0.15	0.18
Mot (0), Warn (0) Coach (1)	-0.22	0.20	-0.19	0.21	-0.15	0.17	0.06	0.15
Mot (0), Warn (1) Coach (0)	-0.29	0.22	-0.21	0.20	-0.19	0.15	0.10	0.15
Mot (0), Warn (1) Coach (1)	-0.20	0.19	-0.17	0.22	-0.16	0.17	0.08	0.18
Mot (1), Warn (0) Coach (0)	-0.20	0.18	-0.17	0.18	-0.17	0.16	0.12	0.13
Mot (1), Warn (0) Coach (1)	-0.16	0.24	-0.16	0.23	-0.10	0.19	0.12	0.14
Mot (1), Warn (1) Coach (0)	-0.23	0.21	-0.20	0.22	-0.16	0.17	0.09	0.15
Mot (1), Warn (1) Coach (1)	-0.18	0.21	-0.15	0.21	-0.11	0.18	0.12	0.15

^aMot, Warn, Coach indicates Motivation, Warning, and Coaching. One indicates the manipulation occurred, 0

indicates there was no manipulation.

Sign

Tab

Tab

the

bio

ma

wo

Sign

Ve

dis

as

Fr

in

to

fr

of

o

fr

h

The correlation between item objectivity and item performance does vary significantly as a result of the manipulation conditions of Coaching and Motivation (see Table 12). Mean levels of the Fisher z for objectivity across conditions are shown in Table 13 below. It should be noted that all the correlations are negative indicating that the more objective the item was judged to be, the lower were students' scores on the biodata items. As would be expected the strongest negative correlations were in the two manipulation groups without motivation or warning, and the weakest were in the groups with both motivation and coaching.

The correlation between item verifiability and item performance does not vary significantly across manipulation conditions, and mean levels across conditions are not very different, as shown in Table 13. Again, all correlations are negative indicating different items that are more verifiable produce lower biodata scores.

For controllability, the correlation with item performance does vary significantly as a result of the manipulation conditions of coaching and motivation. Mean levels of the Fisher z for controllability across conditions are shown in Table 13, where the largest gap in means is between the groups without coaching or motivation, and the group that has both coaching and motivation, but no warning. Correlations between controllability and biodata scores are once again negative, but slightly lower than correlations between objectivity and biodata performance. This is contrary to the direction expected for controllability.

College relevance did not show significant differences in correlation with item performance across manipulation conditions, however, there is a significant Coaching by Motivation interaction. This suggests that those who are both motivated and coached

have a greater likelihood of scoring highly on biodata items, when those items appear relevant to academic performance, than those who only experience one or the other of those conditions. This also suggests that the coaching on dimensionality of biodata items is effective in helping individuals who are motivated to identify dimensions that are relevant to college performance. All correlations between item relevance and biodata item responses were positive, but lower than those with other item characteristics. These correlations indicate that the higher the perceived relevance of the item, the higher the response to biodata items.

Similar analyses were conducted for SJI correct response options, where the dimensions of objectivity, verifiability, controllability and college relevance of each correct response option for the question, “What would you be most likely do?” was correlated with SJI performance, and then converted to Fisher z .

Overall mean levels are shown in Table 14. ANOVA results for the four item characteristics are shown in Table 15, and mean levels across manipulation conditions are shown in Table 16.

Table 14

Overall Means and Standard Deviations of Four Item Characteristic Fisher z for SJI

Dimension	N	Mean	SD
Objectivity	362	0.02	0.22
Verifiability	362	-0.01	0.21
Controllability	362	0.13	0.18
Relevance	362	0.10	0.13

Table 15

Analysis of Variance Results for Fisher z for Four Item Characteristics for SII

<u>Between</u> <u>Subjects Source</u>	Objectivity		Verifiability		Controllability		Relevance	
	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>	<i>df</i>	<i>F</i>
Coaching (C)	1	47.83**	1	30.12**	1	1.11	1	0.00
Motivation (M)	1	23.23**	1	17.75**	1	2.64	1	0.24
Warning (W)	1	0.06	1	0.04	1	0.21	1	1.51
C x M	1	2.24	1	2.15	1	0.16	1	3.38
C x W	1	0.18	1	0.52	1	0.15	1	0.51
M x W	1	0.12	1	0.23	1	0.66	1	0.01
C x M x W	1	0.32	1	1.11	1	0.02	1	4.86**
Error	354	(.04) ^a	354	(.04) ^a	354	(.03) ^a	354	(.02) ^a

^aEquals the Mean Square Error.

*p < .05. **p < .01.

Table 16

Means and Standard Deviations of Four Item Characteristic Fisher z for Various Study Conditions for SJI

	Objectivity		Verifiability		Controllability		Relevance		N
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	
Mot (0), Warn (0) Coach (0) ^a	-0.10	0.18	-0.11	0.20	0.13	0.20	0.10	0.13	46
Mot (0), Warn (0) Coach (1)	0.04	0.19	0.01	0.20	0.13	0.16	0.10	0.14	48
Mot (0), Warn (1) Coach (0)	-0.08	0.21	-0.07	0.18	0.09	0.17	0.14	0.12	46
Mot (0), Warn (1) Coach (1)	0.02	0.19	-0.03	0.18	0.11	0.16	0.09	0.12	44
Mot (1), Warn (0) Coach (0)	-0.01	0.23	-0.03	0.21	0.13	0.15	0.10	0.12	45
Mot (1), Warn (0) Coach (1)	0.17	0.21	0.11	0.18	0.15	0.16	0.09	0.13	45
Mot (1), Warn (1) Coach (0)	-0.03	0.24	-0.05	0.22	0.13	0.25	0.08	0.14	45
Mot (1), Warn (1) Coach (1)	0.16	0.18	0.10	0.17	0.17	0.18	0.14	0.14	43

^aMot, Warn, Coach indicates Motivation, Warning, and Coaching. One indicates the manipulation occurred, 0 indicates there was no manipulation.

The correlation between item objectivity and item performance does vary significantly as a result of the manipulation conditions of Coaching and Motivation (see Table 15). Mean levels of the Fisher z for objectivity across conditions are shown in Table 16 below. Where the correlations are negative, this indicates that the more objective the item was judged to be, the lower were students' scores on the SJI items. As would be expected for the strongest negative correlations were in the two manipulation groups without motivation or warning, and the positive correlations are the groups with coaching. The largest mean correlations were for the coached and motivated groups.

The correlation between item verifiability and SJI item performance varies significantly across manipulation conditions of coaching and motivation, and mean levels across conditions vary, with negative correlations for the groups without coaching, as well as the group with coaching and a warning statement with no motivation. Again, negative correlations indicate different items that are more verifiable produce lower SJI scores under conditions that do not precipitate high performance. Coaching and motivation result in high performance, as well as positive correlations, suggesting that under certain conditions, verifiability may no longer suppress scores on SJIs.

For controllability, the correlation with item performance does not vary significantly as a result of the manipulation conditions of coaching, motivation or warning. Mean levels of the Fisher z for controllability across conditions are fairly stable. Correlations between controllability and SJI scores are positive, as expected for controllability.

College relevance did not show significant differences in correlation with item performance across manipulation conditions, however, there is a significant Coaching by

Motivation by Warning interaction. All correlations between item relevance and SJI item responses were positive. These correlations indicate that the higher the perceived relevance of the item, the higher the response to SJI items. One limitation of this analysis is the dichotomization of scores for SJI performance, where I have considered a correct response a score of 1, and an incorrect response a performance score of 0.

These results provide support for Hypothesis 7 on two dimensions; objectivity and college relevance, for biodata, and significant findings for controllability for biodata, but in a direction opposite to that hypothesized. This suggests that these item characteristics are related to the likelihood that the biodata item score may be inflated. Future work on biodata should take into account the characteristics of the item where they may be able to be formulated in a fashion that limits faking. A closer examination of the issue of controllability shows mixed results of the effect of controllability on biodata validity, depending on the type of controllability (see Graham, et al., 2002). This suggests that hypotheses about controllability must consider specific issues of whether the individual can choose to perform an action, has no control over the action, has shared control, or the individual's feelings or attitudes are the issue of interest. Each of these categories may provide conflicting results, and such specifications were not taken into account by the expert raters in this study.

The results also provide support for Hypothesis 7 on dimensions of objectivity, verifiability, and college relevance, for SJI. It appears from the change in correlation direction from negative to positive under high performance conditions that scores on SJIs are less likely to be suppressed as a result of item characteristics than are scores on biodata items under similar conditions.

The issue of an elaboration requirement as a means of reducing inflation on biodata items was tested, with results shown earlier in Table 6. While there was a significant main effect for elaboration, the impact of elaboration also appeared much smaller when all covariates were included in the analyses. Hypothesis 8 was supported. It is possible that there is an item type confound where the items that were chosen by Oswald, et al. (in press) for elaboration may have also been those that were more verifiable. Since they were verifiable, respondents were less likely to inflate responses.

Validity and Selection Decisions

The fourth question of the study addresses validity and the effects of selection decisions. Table 17 shows correlations between the predictors and outcomes, as well as their descriptive statistics.

Table 17

Correlations and Descriptive Statistics for Predictors and Criteria

	Year 1 GPA	Absentee- ism	SJI total	<i>N</i>	Mean	<i>SD</i>
Year 1 GPA				353	2.88	0.73
Absenteeism	-0.20**			362	3.11	1.11
SJI total	0.06	-0.20**		360	0.65	0.37
Biodata total	0.09	-0.05	0.47**	362	3.09	0.48

** $p < .01$.

To test the validity of these predictors of student performance, and to test Hypothesis 9, the difference between the zero-order validity coefficients and the validity

coefficients with the faking identification scales partialled out were calculated. Results are shown in Table 18.

Table 18

Zero Order and Partial Correlations between Situational Judgment and Biodata and Two Criteria Controlling for Measures of Faking

Situational Judgment

	Zero order	Bogus	Inflation	Impr.Mgt.	All three
	<i>r</i>	Partial	Partial	Partial	Partialled
GPA	0.06	0.05	0.02	0.03	0.01
Absenteeism	-0.20*	-0.20*	-0.23*	-0.14*	-0.19*

Biodata

	Zero order	Bogus	Inflation	Impr.Mgt.	All three
	<i>r</i>	Partial	Partial	Partial	Partialled
GPA	0.09	0.07	0.05	0.08	0.03
Absenteeism	-0.05	-0.06	-0.05	-0.01	-0.05

* $p < .05$.

For the GPA outcome, neither biodata nor SJI proved to be useful predictors of academic performance, whether or not they were statistically adjusted using the faking measures. SJI did predict absenteeism, but the zero-order correlations were comparable to all four partial correlations. The partial correlation when Impression Management was controlled was somewhat lower, though not statistically different from the other correlations. In any event, there is no evidence that inflation, as measured by our three



indices, is attenuating the validity of these measures to predict grade point average and absenteeism.

As McFarland (2000) noted that faking may affect reliability, the alpha coefficients across manipulation conditions were examined for biodata and SJI, and are shown in Table 19. While they do not differ dramatically across conditions, it is interesting to note that the manipulation condition which would be expected to display the greatest degree of inflation (Motivated, Coached, not Warned) does have the highest reliability for biodata. However, the same group has the lowest reliability for SJI, but none of the reliabilities are very different across conditions, and all are relatively high.

Table 19

Coefficient Alpha of Biodata and SJI Scales for Various Study Conditions

	Biodata	SJI	N
Mot (0), Warn (0) Coach (0) ^a	.80	.70	46
Mot (0), Warn (0) Coach (1)	.89	.74	48
Mot (0), Warn (1) Coach (0)	.87	.71	46
Mot (0), Warn (1) Coach (1)	.78	.77	44
Mot (1), Warn (0) Coach (0)	.82	.70	44
Mot (1), Warn (0) Coach (1)	.94	.69	45
Mot (1), Warn (1) Coach (0)	.84	.77	44
Mot (1), Warn (1) Coach (1)	.90	.66	43

^aMot, Warn, Coach indicates Motivation, Warning, and Coaching. One indicates the manipulation occurred, 0 indicates there was no manipulation.

Table 20

Adjusted r for Biodata and SJI Scales in Predicting GPA and Absenteeism across Various Study Conditions

	Biodata			SJI		
	GPA	Absenteeism	GPA	Absenteeism	N	
Mot (0), Warn (0) Coach (0) ^a	-0.41**	-0.00	-0.10	-0.27*	46	
Mot (0), Warn (0) Coach (1)	0.25	-0.10	0.17	-0.39**	48	
Mot (0), Warn (1) Coach (0)	0.10	-0.00	0.00	-0.00	46	
Mot (0), Warn (1) Coach (1)	-0.00	-0.17	0.00	-0.42**	44	
Mot (1), Warn (0) Coach (0)	0.00	0.00	0.27	-0.25	44	
Mot (1), Warn (0) Coach (1)	0.00	-0.00	-0.00	-0.10	45	
Mot (1), Warn (1) Coach (0)	0.32*	-0.00	0.00	-0.14	44	
Mot (1), Warn (1) Coach (1)	0.25	-0.00	0.00	0.00	43	

^aMot, Warn, Coach indicates Motivation, Warning, and Coaching. One indicates the manipulation occurred,

0 indicates there was no manipulation.

*p < .05. **p < .01.

The amount of variance in GPA and absenteeism, predicted by biodata and SJI, across conditions is shown in Table 20. There are no distinct patterns to validity across conditions, and validity comparisons across groups can not be clearly interpreted as they are limited by the small sample size and low overall criterion-related validity of the biodata and SJI measures.

To test Hypothesis 10, I examined the responses of two groups of respondents. Respondents in a group that should be most apt to inflate their scores (the Coaching, Motivation, no Warning group, which I refer to as the *high performance* group) were compared with those in the group in which faking should be minimized as the individuals would be least apt to inflate their scores (the no Coaching, no Motivation and a Warning group, which I refer to as the *reference* group). Table 21 below presents descriptive data for each of the two groups. The high performance group shows a higher mean score on all three of the faking identification scales, although the largest gap between the two is on the scales comprising bogus items and the inflation scale on which the mean differences are over .5 standard deviation units. Recall that the Inflation Index was generated using these two conditions, so this large difference is artificially inflated, and this index would need to be cross-validated. A much smaller mean difference was obtained on the Impression Management scale, suggesting that this is more stable and personality-related, rather than being prompted by situational factors, which seems to be the case for the Bogus items and items making up the Inflation Index.

Table 21

Descriptives for High Performance and Reference Groups

	High Performance			Reference		
	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>
Bogus Scale (4-20)	45	5.84	2.12	46	4.87	1.34
Inflation Index (0-23)	44	15.86	3.99	46	10.04	3.49
Impression Management (19-105)	45	50.91	11.76	46	52.70	10.43
Year 1 GPA (0-4.0)	45	2.74	0.75	45	2.90	0.78
Absenteeism (1-5)	45	3.20	1.04	46	2.89	0.99
						<i>d</i>
						-0.72
						-1.67
						0.17
						0.21
						-0.31

Given these data, a large number of positive responses to the Bogus items definitely seems to indicate that the respondents' answers are generally suspect. The same might be true of the Inflation index, but, as I noted above, results for that scale should be cross-validated.

To examine how the use of these scales to correct for faking might affect who is selected or admitted to college, I examined possible admissions decisions based on a selection of the top 10%, 25% and then 50% of the participants, based on their biodata and SJI scores. To control for faking, I identified those who scored above a cut-point on the faking scales. The Bogus scale is presented as an example; similar analyses were conducted with the inflation index and impression management, with similar results. A cut-point of 4 on the Bogus items scale was chosen as this reflected any positive response on the Bogus scale items. As the Bogus items are made up of experiences that are impossible, I can presume that anyone who claims experience on these items has not been accurate about their experience, and it may be reasonable to exclude them. Having excluded these individuals, I then proceeded down the list, selecting the next best candidates in terms of biodata and SJI performance, who do not meet the cut score on the Bogus items scale. The actual college performance (GPA and absenteeism) of those who were excluded was then compared to the actual college performance of those who were selected as alternatives.

Looking first at the .10 selection ratio, those chosen for best performance included 26 individuals with scores above 4 on the Bogus scale. These individuals were removed, and replaced with the next-highest scoring individuals who did not have scores above the

cut on the Bogus scale. The differences in college performance for these two groups is shown in Table 22 below.

Table 22

Descriptive Statistics for Selection Ratio of .10

	Removed from top			Replacements for top			
	10% selection			10% selection			
	N	Mean	SD	N	Mean	SD	<i>d</i>
Absenteeism	26	3.08	1.06	26	3.08	0.84	0.00
Year 1 GPA	26	2.77	0.90	26	2.85	0.82	-0.09

The same process was followed for a selection ratio of .25 and the results are shown below, demonstrating little difference between those excluded and those pulled in as replacements.

Table 23

Descriptive Statistics for Selection Ratio of .25

	Removed from top			Replacements for top			
	25% selection			25% selection			
	N	Mean	SD	N	Mean	SD	<i>d</i>
Absenteeism	58	3.02	1.03	58	2.95	0.96	0.07
Year 1 GPA	58	2.89	0.78	55	2.87	0.73	0.03

The same process was followed again for a selection ratio of .50, with similar findings.

Table 24

Descriptive Statistics for Selection Ratio of .50

	Removed from top			Replacements for top			
	50% selection			50% selection			
	N	Mean	SD	N	Mean	SD	<i>d</i>
Absenteeism	109	3.02	1.02	109	3.05	1.15	-0.03
Year 1 GPA	108	2.93	0.70	103	2.89	0.79	0.06

It appears that, regardless of the level of selectivity, using Bogus items as a means of removals does not change the quality of the student population, when considering actual college performance.

This study demonstrates those who are coached and motivated to achieve can produce high scores on biodata and situational judgment, and those scores may be a result of inflation due to faking. Inflation scales can be a useful method of identifying response manipulation but in this case are less useful in correcting for it. Further research into inflation scales and corrections would be necessary to develop protocols for their use. It may be helpful to conduct further analyses with a broader criterion measure that is conceptually more closely related to the biodata and situational judgment dimensions. Results here are probably affected by the very low criterion-related validity of the biodata and SJI measures, and Graham, et al. (2002) have linked faking in biodata to reduced criterion-related validity.

CONCLUSIONS

Discussion

This study has addressed four major questions. First, how do different situational differences such as testing conditions influence how an individual performs on biodata and SJIs? Second, what are the individual difference characteristics of people who are most effective in inflating their responses? Third, what are characteristics of the test items that seem most susceptible to inflation? Fourth, does the susceptibility of biodata and SJIs to faking affect their predictive validity if they were to be used in college admissions decision-making?

Situational Factors

Situational factors can be very important determinants of performance on biodata and situational judgment items. Brief coaching was shown to improve scores on both of these non-cognitive measures, an important issue in the context of test application. According to Sackett, et al. (1989), "In the typical performance domain, the examinee may adopt an explicit self-presentation strategy in responding to the selection device based on a hypothesis about what the employer is looking for in an applicant." (p. 148). Coaching appears effective in aiding examinees in generating hypotheses and presenting themselves appropriately. Were biodata and SJIs to be used in a college admissions selection process, those who have access to coaching would be better able to improve their scores. Such susceptibility to coaching is of great practical significance to the College Board, the owners of these measures, as they would need to be able to defend the reliability and validity of the measures, or ensure that examinees have received equal access to coaching (see *Standards for Educational and Psychological Tests and*

Manuals). Considering the effectiveness of the written coaching provided in Cunningham, et al. (1994), it may be appropriate for researchers to consider future studies that examine the difference in effectiveness between face-to-face coaching programs, and materials that can be provided in hard copy or via electronic media. Also, it may be useful to conduct an examination of the relationship between warning statements and coaching, as coaching may effectively negate the power of the warning statement.

If these tests were to be used to contribute to college admissions decisions, it is realistic to assume that college applicants would be highly motivated to perform well. Motivation was shown to have a significant effect on performance on these measures, and the interaction of coaching and motivation for SJIs in this study suggests that maximal performance in these tests in an applied setting will be best facilitated by the combination of personal motivation and coaching.

Warning statements, although not found in this study to be a significant factor in suppressing inflation overall, did operate in the direction expected when individuals are motivated to perform well. Warning statements are not equally relevant for biodata and SJI items. While objective and verifiable biodata items, such as the number of leadership positions that a person held in high school, could be verified by contacting the high school, verification would be more difficult to conduct for SJIs. While peers or teachers could be contacted and asked whether an individual is actually likely to behave in a certain way in a certain situation, the possibility of this verification is clearly more remote, suggesting that warning statements about verification of responses are less useful

for SJIs, and may even make matters worse by actually planting the idea that one could answer dishonestly.

To be able to implement the biodata and SJI in an admissions context, these measures would need further examination, ideally with a sample of individuals who really have a personal desire to score well on the tests and a great deal to lose if they do not, to fully understand the power of a warning statement and the interaction effects of motivation and warning. To ensure the most powerful warning effect it may be best to include both the warning that faking could be identified, as well as the warning that there would be potential negative consequences for faking (see Dwight & Donovan, 2003). According to Dwight and Donovan, both of these characteristics of a warning statement play a part in limiting faking. A study using a sample of real college applicants who believe that these tests are being used in their admissions decision-making process would provide an excellent framework for building our understanding of the effects of motivation and warning on biodata and SJI score inflation.

Identification of Inflation

Three inflation identification methods were used, providing valuable information about individual differences and inflation. Inflation captured by these scales was unrelated to race, gender, or cognitive ability.

First, Bogus items were considered. Bogus items were successfully used to identify those who had claimed experiences that were impossible, and scoring on Bogus items was more situational than personality-related. Bogus item scoring was weakly related to openness to experience, which might suggest that those who are more creative in framing their own personal experiences are able to warrant marking experience on

Bogus items, perhaps viewing the experience being addressed as in some way related to an experience they have had. Future research with bogus items should include a broader selection of bogus items, with varying degrees of obviousness. There may be particular bogus item characteristics that make some more effective as flags of inflation than others.

Second, an Inflation Index was created. Inflation captured with an empirically-derived Inflation Index was more personality-driven than the Bogus items, and was weakly related to extraversion, agreeableness, openness and self-deception. The Inflation Index proved to be a useful technique in identifying those with unusually high scores. Nevertheless, using such an index as a way to flag problematic responders could cause headaches for the College Board, who would need to be able to defend their practices against those arguing that they did not cheat, but were responding honestly, but simply have unusual experiences that result in their being flagged by the index. It should be noted that while the empirical Inflation Index proved useful, results should be cross-validated.

Third, the BIDR Impression Management scale was used as an indicator of inflation. Impression management appeared to be the most personality-related of the identification scales. Impression management was more strongly related to conscientiousness, agreeableness, emotional stability, and self-deception. This suggests that inflation related to impression management may be more adaptive than inflation related to Bogus items, or the Inflation Index. That impression management appears to be trait-related lends support to the argument that it should not be used for controlling for socially desirable responding. It is apparent from this study that inflation is largely a

function of the situation in which the test is taken, rather than the characteristics of the individual, and this study lends support to the view of impression management as a trait.

Item Characteristics

Item characteristics can play an important role in creating a test that is less vulnerable to inflation. This study found that biodata items that were judged as being less objective, less controllable, and more college relevant were more susceptible to inflation under different manipulation conditions. SJI items that were objective or verifiable tended to have lower scores, unless the individual was coached and motivated to do well, in which case, the suppressor effect faded. The suppressor effect of item characteristics was not weakened for biodata to the extent it was for SJI items, when coaching and motivation to perform were provided.

To limit inflation, it may be important for biodata test generators to choose items that are objective and verifiable where possible. However, one disadvantage of trying to ensure that items are objective and verifiable is that biodata inventory builders may lose the valuable multidimensionality captured from a broader pool of items. Also, there may be resistance to removing items that are fakeable when those items are perceived as the most valid. While college relevance was shown to increase inflation, there is an inherent tradeoff in seeking to minimize faking by trying to disguise college relevance – tests that are viewed as relevant are more likely to be perceived as fair (see Schmitt, Oswald, Kim, Gillespie, & Ramsay, in press), and trying to reduce college relevance would not be recommended.

That less controllable items were more susceptible to inflation suggests that further research will be necessary to resolve the different categories of control, and the

consequences of having items that tap those categories. Graham, et al., (2002) indicate four categories of controllable items, and these may all produce differing results in terms of inflation and validity. Judges in this study were not given multiple categories for rating controllability of items, and the work of Graham, et al., provides guidance for improved evaluation of controllability in future research.

An elaboration requirement in biodata items appears to be effective in suppressing scores, adding to existing literature in this area (e.g., Schmitt & Kuncze, 2002). However, not all items are equally suitable for an elaboration requirement. It may be that more verifiable items are more easily written with an elaboration requirement. Also, there is an issue of memory that may be affecting responses to elaborated biodata items.

Respondents may limit their responses when elaboration is required only because they can't remember the details and so would be unable to record them. Ramsay, Kim, Gillespie, and Friede (2003) have considered memory as a factor in a biodata elaboration study, but used a memory test that has questionable construct validity, and may be more a test of general knowledge. Further research in this area may provide answers to the questions surrounding elaboration requirements and the factors that contribute to their effectiveness.

These are issues that test developers will need to balance as they attempt to create inventories that are reliable and useful. Considering the relationship of item characteristics to inflation, and that validity may also be affected by vulnerability to inflation linked to item characteristics (Graham, et al., 2002) item characteristics as a means of understanding inflation may be a fruitful area for future research.

Validity

The issue of validity is of critical importance if the tests used here are put into practice for selection in an applied setting. It appears from this study that inflation on biodata and SJI does not attenuate the criterion-related validity, however, we are examining a set of tests that show limited criterion-related validity without inflation. To reach a definitive conclusion about the relationship of inflation and validity, it may be useful to conduct further research with performance measures that are multidimensional and more closely linked to the dimensions captured by the biodata and SJI questions.

Identifying participants who appeared suspicious, based on the three inflation identification methods, and then replacing them with participants who were not suspected of inflation did not prove effective in selecting a higher performing group in terms of college performance. However, criterion-related validity was low, and such corrections may have different effects if one were examining a broader criterion space.

Limitations

One limitation of this study is the slightly different warning manipulations for the motivated and not motivated groups. While both groups received the same statement about the possibility that responses would be verified, the consequences of being found providing dishonest answers had a greater negative consequence for the motivated group, who would lose a cash payout for high performers. These differences mean that the manipulation was not equivalent across groups. It would have been helpful to have included a manipulation check regarding the warning statements to establish that they had been attended to. This sample was not completing these measures under the real expectation that responses would contribute to college admission decisions, limiting the

power of the warning statement that there would be negative consequences for responding dishonestly. In a real world scenario, the risk that one's college application may be rejected based on dishonest responding would be of great consequence. Therefore, future research will need to examine warning statements in a sample of real applicants. As the inflation index was generated in the study sample, it would need to be cross-validated as similar results may not be found in a different sample.

Another possible limitation is that elaborated items in the study may have been items that were more verifiable. This verifiability, rather than the elaboration requirement itself, may have limited the respondents' inclination to inflate responses. In addition, it is difficult to draw conclusions about any changes in validity, or the effectiveness of correcting for inflation, when there is very limited criterion-related validity for the biodata and SJI measures in this sample.

Practical Implications

The use of biodata and situational judgment measures have shown promise as predictors of student performance (Oswald, et al., in press). However, in this sample, these measures did not prove effective as performance predictors. For these tests to be implemented in a college admissions context, greater criterion-related validity would need to be demonstrated. Considering the vulnerability of these measures to inflation prompted by motivation and coaching, it may be most practical to reserve measures such as these for informational or developmental purposes, rather than for admissions decision-making.

Bearing in mind that respondents can improve their scores when coached very briefly on how to do so, it is problematic to promote this test as one that could be used in

admissions decision-making. If the tests were to be used in such a way, test owners would need to provide access to coaching materials, so that all examinees have an equal opportunity to improve their scores through knowledge about the test. While written or web-based materials could easily be generated by the test owners, it is reasonable to assume that the implementation of tests such as these would prompt market forces to produce tests preparation training classes that could provide greater advantages to those with the resources to attend such training. This could precipitate sub-group differences in test performance, one of the issues that these tests are designed to avoid.

Test owners would be well-advised to implement a warning statement in the test administration, once further research has been conducted that clarifies the most effective type of warning statement in a college admissions context. It is expected that a warning statement, claiming that dishonest responses may be verified and that there may be real negative consequences for inflating responses, may be effective. However, the power of the warning statement is expected to vary based on the purpose for which the testing tools are used. High-stakes decisions based on test performance may result in greater responsiveness to a warning statement that includes negative consequences for dishonest responding.

In compiling an inventory of biodata items, test developers should consider items that are less likely to be inflated. This may be achieved by focusing on items that are written in a way that maximizes their content verifiability and objectivity. Controllability of item content should also be considered, but further research is necessary regarding the dimensions of controllability.

There may be items that have high criterion-related validity, yet are vulnerable to inflation. Test developers will need to balance the issue of transparency and possible inflation with prediction of performance. It would also be advisable to include an elaboration requirement for biodata items, where possible, to limit the item susceptibility to inflation.

While high scores on the BIDR's impression management scale are personality-related, and may be adaptive, scores on bogus items are not. The inclusion of bogus items is one method of identifying individuals who may be responding in a dishonest manner. However, some members of the public may find it inappropriate to ask "trick questions", and could generate publicity that makes the items themselves less useful. An inflation index is another suitable means of identifying responses that are suspicious, and the index used here will need to be validated with another sample. While use of one or more of these scales to identify inflation is possible in an applied setting, their implementation may be difficult. While conceptually, bogus items identifying those who have lied, and the empirical soundness of an inflation index, should be defensible reasons to exclude the scores of a respondent, the public relations problems that could ensue with implementation of such measures could make use of these scales infeasible.

One problem that may arise with the use of corrections using inflation identification scales is the partialing out of variance that may be an important predictor of performance. Also, people do not necessarily inflate their responses in the same way, and standard correction method may not account for this (see discussion in Christiansen, et al., 1994). Also, some individuals may honestly have marked responses that indicate

their real but very unusual levels of experience, and such candidates may find themselves discriminated against, if such methods were put into practice.

While we can use these techniques to identify inflation, what to do about respondents suspected of faking in a college admissions context is debatable. Decisions to dismiss dishonest applicants in such a high stakes setting will need to be based on far more extensive research than has been provided in this study, to withstand public scrutiny and possible litigation. If these tests were simply to be used for developmental purposes, bogus items and the inflation index would be useful markers for a counselor reviewing test results.

Further research on these biodata and situational judgment measures would allow them to be refined so that they effectively capture the dimensionality of the criterion space that they are designed to predict. To be of practical value they should have construct and criterion-related validity, and this will need to be examined in different samples, while considering different item types. Research should also seek to establish the feasibility of using these measures in an admissions context and should investigate other possible uses for these tests.

APPENDICES

APPENDIX A

Mael's Taxonomy

Mael's Taxonomy of Biodata Items

Historical

How old were you when you got your first paying job?

External

Have you ever been fired from a job?

Objective

How many hours did you study for your real-estate license test?

First-hand

How punctual are you about coming to work?

Discrete

At what age did you get your driver's license?

Verifiable

What was your grade point average in college?

Were you ever suspended from Little League?

Controllable

How many tries did it take you to pass the CPA exam?

Equal access

Were you ever class president?

Job relevant

How many units of cereal did you sell during the last calendar year?

Noninvasive

Were you on the tennis team in college?

Future or hypothetical

What position do you think you will be holding in 10 years?

What would you do if another person screamed at you in public?

Internal

What is your attitude towards friends who smoke marijuana?

Subjective

Would you describe yourself as shy?

How adventurous are you compared to your coworkers?

Second-hand

How would your teachers describe your punctuality?

Summative

How many hours do you study during an average week?

Nonverifiable

How many servings of fresh vegetables do you eat every day?

Noncontrollable

How many brothers and sisters do you have?

Nonequal access

Were you captain of the football team in high school?

Not job relevant

Are you proficient at crossword puzzles?

Invasive

How many young children do you have at home?

Note. Adapted from "A Conceptual Rationale for the Domain and Attributes of Biodata Items," by F. A. Mael, 1991, *Personnel Psychology*, p. 773.

APPENDIX B

Sample Questionnaire

PID:A_____

FORM 1

The first booklet asks questions pertaining to how you approach other people or life in general. You will also complete some demographic questions. You will use the first scantron for these questions.

The second booklet contains questions that ask you about your history and life experiences. You will also be presented with descriptions of problem situations, and you will indicate which action you would be most likely to take and which action you would be least likely to take. These are situational judgment tasks. You will use the second scantron for these questions.

As you are answering these questions, please record your answers on the scantron form. For each question, please fill in completely the circle you choose. Where you are asked to elaborate on your answer, please write your response on the lines provided in your exercise booklet.

First, please take a moment to complete the following areas of your first scantron:

PID – Please write in your PID, and then fill in the corresponding circles.

Form – Please indicate Form 1 A

Also, please indicate your PID on the cover of this booklet, at the top right hand corner.

You will have 90 minutes to complete this study.

The following pages contain phrases describing people's behaviors. Use the rating scale below to describe how accurately each statement describes you and please provide answers that describe yourself as you generally are now, not how you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully.

Please use the five-point scale below:

- 1 = Very Accurate
- 2 = Moderately Accurate
- 3 = Neither Accurate nor Inaccurate
- 4 = Moderately Inaccurate
- 5 = Very Inaccurate

- 90. Make people feel at ease.
- 91. Am not interested in abstract ideas.
- 92. Change my mood a lot.
- 93. Don't like to draw attention to myself.
- 94. Talk to a lot of different people at parties.
- 95. Have excellent ideas.
- 96. Insult people.
- 97. Follow a schedule.
- 98. Am exacting in my work.
- 99. Get stressed out easily.
- 100. Seldom feel blue.
- 101. Don't mind being the center of attention.
- 102. Worry about things.

Please use the five-point scale below:

- 1 = Very Accurate
- 2 = Moderately Accurate
- 3 = Neither Accurate nor Inaccurate
- 4 = Moderately Inaccurate
- 5 = Very Inaccurate

- 103. Have little to say.
- 104. Don't talk a lot.
- 105. Use difficult words.
- 106. Keep in the background.
- 107. Have difficulty understanding abstract ideas.
- 108. Make a mess of things.
- 109. Pay attention to details.
- 110. Am always prepared.
- 111. Feel little concern for others.
- 112. Have a rich vocabulary.
- 113. Like order.
- 114. Often feel blue.
- 115. Am full of ideas.
- 116. Spend time reflecting on things.
- 117. Take time out for others.
- 118. Have frequent mood swings.
- 119. Have a soft heart.
- 120. Am quick to understand things.

Please use the five-point scale below:

- 1 = Very Accurate
- 2 = Moderately Accurate
- 3 = Neither Accurate nor Inaccurate
- 4 = Moderately Inaccurate
- 5 = Very Inaccurate

- 121. Am interested in people.
- 122. Start conversations.
- 123. Am the life of the party.
- 124. Neglect my duties.
- 125. Am relaxed most of the time.
- 126. Am not interested in other people's problems.
- 127. Often forget to put things back in their proper place.
- 128. Feel others' emotions.
- 129. Sympathize with others' feelings.
- 130. Do not have a good imagination.
- 131. Get irritated easily.
- 132. Am easily disturbed.
- 133. Get chores done right away.
- 134. Am not really interested in others.
- 135. Am quiet around strangers.
- 136. Feel comfortable around people.
- 137. Leave my belongings around.
- 138. Have a vivid imagination.
- 139. Get upset easily.

Please answer this next set of questions using the five-point scale below.

- 1 = Very true
- 2 = Mostly true
- 3 = Somewhat true
- 4 = Mostly untrue
- 5 = Very untrue

- 140. My first impressions of people usually turn out to be right.
- 141. It would be hard for me to break any of my bad habits.
- 142. I don't care to know what other people really think of me.
- 143. I have not always been honest with myself.
- 144. I always know why I like things.
- 145. When my emotions are aroused, it biases my thinking.
- 146. Once I've made up my mind, other people can seldom change my opinion.
- 147. I am not a safe driver when I exceed the speed limit.
- 148. I am fully in control of my own fate.
- 149. It's hard for me to shut off a disturbing thought.
- 150. I never regret my decisions.
- 151. I sometimes lose out on things because I can't make up my mind soon enough.
- 152. The reason I vote is because my vote can make a difference.
- 153. My parents were not always fair when they punished me.
- 154. I am a completely rational person.
- 155. I rarely appreciate criticism.

Please stop and check that you have just completed the first side of the first scantron.

- 156. I am very confident of my judgments.
- 157. It's all right with me if some people happen to dislike me.

Please continue using the five-point scale below.

- 1 = Very true
- 2 = Mostly true
- 3 = Somewhat true
- 4 = Mostly untrue
- 5 = Very untrue

- 158. I don't always know the reasons why I do the things I do.
- 159. I sometimes tell lies if I have to.
- 160. I never cover up my mistakes.
- 161. There have been occasions when I have taken advantage of someone.
- 162. I never swear.
- 163. I sometimes try to get even rather than forgive and forget.
- 164. I always obey laws, even if I'm unlikely to get caught.
- 165. I have said something bad about a friend behind his or her back.
- 166. When I hear people talking privately, I avoid listening.
- 167. I have received too much change from a salesperson without telling him or her.
- 168. I always declare everything at customs.
- 169. When I was young I sometimes stole things.
- 170. I have never dropped litter on the street.
- 171. I sometimes drive faster than the speed limit.
- 172. I have done things that I don't tell other people about.
- 173. I never take things that don't belong to me.
- 174. I have taken sick-leave from work or school even though I wasn't really sick.
- 175. I have never damaged a library book or store merchandise without reporting it.
- 176. I have some pretty awful habits.
- 177. I don't gossip about other people's business.

178. Please enter 8 on your scantron for this item.

Please fill in the appropriate answer on your form according to the responses provided.

For questions 90 and 91, use the following scale:

- a. = less than 1.00
- b. = 1.00 to 1.49
- c. = 1.50 to 1.79
- d. = 1.80 to 2.09
- e. = 2.10 to 2.39
- f. = 2.40 to 2.69
- g. = 2.70 to 2.99
- h. = 3.00 to 3.39
- i. = 3.40 to 3.59
- j. = 3.60 or greater

179. Cumulative college GPA:

180. GPA for this past semester:

181. Indicate the extent to which you have missed regularly scheduled class(es) in the past six months.

- a. I have never missed class
- b. I missed 1-3 classes.
- c. I missed 4-8 classes.
- d. I missed 9-15 classes.
- e. I missed more than 15 classes.

If you have missed class in the past six months, indicate the reasons you missed class.

Please mark:

a = Yes

b = No

- 182. You were faced with an emergency.
- 183. You were sick.
- 184. You partied too much the night before.
- 185. You were tired or you failed to get up in time.
- 186. You were talking or socializing with friends.
- 187. You were involved with another university event and couldn't go.
- 188. You found the class boring.
- 189. You did not believe the instructor would cover anything new or important.

Demographics – Please respond to the following questions to the best of your ability.

- 190. What is your age?
 - a. 18
 - b. 19
 - c. 20
 - d. 21
 - e. 22
 - f. 23
 - g. 24
 - h. 25
 - i. 26
 - j. 27 +

- 191. What is your gender?
 - a. male
 - b. female

- 192. What is your year in school?
 - a. freshman
 - b. sophomore
 - c. junior
 - d. senior
 - e. 5th year +

193. Which of the following best characterizes you?

- a. U.S. Citizen
- b. Non-citizen – Canadian
- c. Non-citizen – other

194. Is English your primary language?

- a. yes
- b. no

195. What ethnicity do you consider yourself to be?

- a. Mexican American
- b. Puerto Rican
- c. Other Hispanic
- d. American Indian or Alaskan native
- e. Asian
- f. Black/African American
- g. Caucasian/White/Not of Hispanic origin
- h. Native Hawaiian or other Pacific Islander
- i. Other

**You have now completed the first section of your form.
Please raise your hand to let the proctor know that you have
finished this section.
Now please wait quietly until everyone has completed this section.**

PID:A

You are now beginning the second section.

Please take a moment now to prepare the second scantron:

PID – Please write in your PID, and then fill in the corresponding circles.

Form – Please indicate Form 1 B

Also, please indicate your PID on the cover of this booklet, at the top right hand corner.

Please read these instructions very carefully:

Imagine that you are applying for admission to Michigan State University, and your responses to these questions could influence the decision on whether or not to accept you for admission. In other words, imagine that this questionnaire is part of the test requirements for college admissions, and admission here is very important to you. Complete this questionnaire in a way that presents yourself honestly but in the best light possible so that you are most likely to get admitted to the university.

As an added incentive to do well, participants in this study who score above the 50th percentile on this questionnaire will receive \$10.

Now please proceed, answering the remainder of the questions in this study.

Request for Payment Information

Please provide the following information to allow us to mail a check to you if you meet the required score. Those participants who score above the 50th percentile on this questionnaire will be mailed \$10.

Name: _____

Address line 1: _____

Address line 2: _____

City: _____ **State:** _____ **Zip:** _____

Following are some biographical data questions:

(Biodata items are inserted here)

You will be presented with descriptions of problem situations. Each problem has between four and seven alternative actions that might be taken to deal with the problem. You are to make two judgments for each problem. First, decide which alternative you would be MOST LIKELY to take in response to the problem. It might not be exactly what you would do in the situation, but it should be the alternative that comes closest to what you would actually do. Record your answer on the scantron form. Second, decide which alternative you would be LEAST LIKELY to take in the situation, and record your answer on the scantron. Please read all of the alternatives before deciding.

(Situational judgment items are inserted here)

**You have now completed the study.
Please wait quietly until everyone has finished their work.**

Thank you for your participation.

APPENDIX C

Sample Biodata Items

1. During the past year, how many times out of self-interest have you searched for information about other regions, countries, or cultures (at the library or on the Internet)?

- f. 0
- g. 1-3
- h. 4-7
- i. 8-12
- j. more than 12

If you answered b, c, d, or e, briefly describe up to 5 countries or cultures and the topic that you investigated.

2. How many times in the past year have you tried to get someone to join an activity in which you were involved or leading?

- a. never
- b. once
- c. twice
- d. three or four times
- e. five times or more

3. In the past six months, how often did you read a book just to learn something?

- a. never
- b. once
- c. twice
- d. three or four times
- e. five times or more

If you answered b, c, d, or e, briefly describe up to 4 books you read and what you wanted to learn.

Note. For a further information on the complete set of items, please contact the College Board, New York, New York.

APPENDIX D

Sample Situational Judgment Items

You will be presented with descriptions of problem situations. Each problem has between four and seven alternative actions that might be taken to deal with the problem. You are to make two judgments for each problem. First, decide which alternative you would be MOST LIKELY to take in response to the problem. It might not be exactly what you would do in the situation, but it should be the alternative that comes closest to what you would actually do. Record your answer on the scantron form. Second, decide which alternative you would be LEAST LIKELY to take in the situation, and record your answer on the scantron. Please read all of the alternatives before deciding.

Your grade for a particular class is based on three exams, with no class attendance requirement. All of the homework requirements for the class are posted on the professor's web site. What would you do?

- a. Attend class for as long as you feel that it is helping your grades.
- b. Do all the homework but only go to some of the lectures. It's the exams that count.
- c. Go to all the classes anyway. The professor may say something important.
- d. Skip classes, but if you did poorly on the first exam, start going to classes.
- e. There is no need to go to classes. Just get the homework done, and pass the exams.

4. What would you be most likely to do?

5. What would you be least likely to do?

You are finding your freshman year very difficult. The courses are hard, and you feel your grades are not satisfactory. Material in class seems to be covered very quickly. What would you do?

- a. Talk with the professors and TAs to get help on how to study.
- b. Find a study partner and work on homework and class material together.
- c. Talk to your parents and an advisor.
- d. Study hard, try your best, and don't worry about it.
- e. Talk to my advisor and teachers; see if there are study groups or review sessions I can attend.
- f. Hire a tutor for the difficult classes.

6. What would you be most likely to do?

7. What would you be least likely to do?

Note. For a further information on the complete set of items, please contact the College Board, New York, New York.

APPENDIX E

Wording of Instruction Sets

Motivated Warned Group:

“Please read these instructions very carefully:

Imagine that you are applying for admission to Michigan State University, and your responses to these questions could influence the decision on whether or not to accept you for admission. In other words, imagine that this questionnaire is part of the test requirements for college admissions, and admission here is very important to you. Complete this questionnaire in a way that presents yourself honestly but in the best light possible so that you are most likely to get admitted to the university.

As an added incentive to do well, participants in this study who score above the 50th percentile on this questionnaire will receive \$10.

Note that we may verify a subset of your responses, and if you respond dishonestly, that may invalidate this test as well as your chance to receive \$10 for high performance.

Now please proceed, answering the remainder of the questions in this study.”

Motivated Not Warned Group:

“Please read these instructions very carefully:

Imagine that you are applying for admission to Michigan State University, and your responses to these questions could influence the decision on whether or not to accept you for admission. In other words, imagine that this questionnaire is part of the test requirements for college admissions, and admission here is very important to you. Complete this questionnaire in a way that presents yourself honestly but in the best light possible so that you are most likely to get admitted to the university.

As an added incentive to do well, participants in this study who score above the 50th percentile on this questionnaire will receive \$10.

Now please proceed, answering the remainder of the questions in this study.”

Not Motivated Warned Group:

“Please read these instructions very carefully:

The following questionnaire is being tested as a way to collect information about high-school students who are applying to go to college. We would like your straightforward, honest answers to these questions. Your responses are strictly confidential, and they will not be used to evaluate you in any way, so please provide answers that are as honest and accurate as possible.

Note that we may verify a subset of your responses, and if you respond dishonestly, that may invalidate this test.

Now please proceed, answering the remainder of the questions in this study.”

Not Motivated Not Warned Group:

“Please read these instructions very carefully:

The following questionnaire is being tested as a way to collect information about high-school students who are applying to go to college. We would like your straightforward, honest answers to these questions. Your responses are strictly confidential, and they will not be used to evaluate you in any way, so please provide answers that are as honest and accurate as possible.

Now please proceed, answering the remainder of the questions in this study.”

APPENDIX F

Informed Consent Form – Motivated Group

Predictors of Student Success - Informed Consent

Please read and sign below:

In the project in which you are participating, we will be asking you to respond to a series of questions. The first two sets of questions are measures of judgment and of background experiences and preferences; they are experimental measures designed to be related to outcomes of students attending a college or university. The major purpose of this project is to investigate how well students do on these measures, given the instructions to take them, and whether the measures of judgment and background are related to your MSU grades. We are also asking you to respond to some commonly used personality measures which will help us interpret the meaning of your responses to the judgment and background measures. Because a major purpose of our study is to determine if your responses to the judgment and background measures are related to your performance as a student at MSU, we will also be asking your permission to allow the registrar to give us access to your grades and to the Office of Admissions to allow access to your high school grades and ACT/SAT scores.

In order to link your responses to the measures with your college and high school grades, and ACT/SAT scores, we will be asking you to provide your PID. All information you provide will be completely confidential. Only the project team (two faculty members and three graduate students) will have access to the password-protected data file with the original PID attached, and all data will be reported at the group level so that no one will be able to identify a particular person. As soon as we link your responses to the data from Admissions and the Registrar's Office, your PID will be deleted from our data file. Your privacy will be protected to the maximum extent allowable by law.

We expect that it will take 90 minutes for you to complete this study, for which you will earn extra credit in Psychology 101. Participation in this study is voluntary. As an alternative to participation in this study, you may do other work, such as a paper, that is coordinated with your instructor. As an incentive to do well on this questionnaire, those participants who score above the 50th percentile on this questionnaire will be given \$10. By signing below you indicate that you are free to refuse to participate in this project or any part of the project. You may refuse to answer some of the questions and may discontinue your participation at any time without penalty.

If you have any questions or concerns about your participation in this project, you can call Neal Schmitt (517-355-8305) or send an email message to Schmitt@msu.edu. If you have questions or concerns regarding your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact – anonymously if you wish - Ashir Kumar, M.D., Chair of the University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, e-mail: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824.

Your signature below indicates your voluntary agreement to participate in this study.

Signature

Date

APPENDIX G

Informed Consent Form – Not Motivated Group

Predictors of Student Success - Informed Consent

Please read and sign below:

In the project in which you are participating, we will be asking you to respond to a series of questions. These questions are measures of background experiences and preferences, experimental measures designed to be related to outcomes of students attending a college or university. The major purpose of this project is to investigate how well students do on these measures, given the instructions to take them, and whether the measures of judgment and background are related to your MSU grades. Because a major purpose of our study is to determine if your responses to the judgment and background measures are related to your performance as a student at MSU, we will also be asking your permission to allow the registrar to give us access to your grades and to the Office of Admissions to allow access to your high school grades and ACT/SAT scores.

In order to link your responses to the measures with your college and high school grades, and ACT/SAT scores, we will be asking you to provide your PID. All information you provide will be completely confidential. Only the project team (two faculty members and three graduate students) will have access to the password-protected data file with the original PID attached, and all data will be reported at the group level so that no one will be able to identify a particular person. As soon as we link your responses to the data from Admissions and the Registrar's Office, your PID will be deleted from our data file. Your privacy will be protected to the maximum extent allowable by law.

We expect that it will take 90 minutes for you to complete this study, for which you will earn extra credit in Psychology 101. Participation in this study is voluntary. As an alternative to participation in this study, you may do other work, such as a paper, that is coordinated with your instructor. By signing below you indicate that you are free to refuse to participate in this project or any part of the project. You may refuse to answer some of the questions and may discontinue your participation at any time without penalty.

If you have any questions or concerns about your participation in this project, you can call Neal Schmitt (517-355-8305) or send an email message to Schmitt@msu.edu. If you have questions or concerns regarding your rights as a study participant, or are dissatisfied at any time with any aspect of this study, you may contact – anonymously if you wish - Ashir Kumar, M.D., Chair of the University Committee on Research Involving Human Subjects (UCRIHS) by phone: (517) 355-2180, fax: (517) 432-4503, e-mail: ucrihs@msu.edu, or regular mail: 202 Olds Hall, East Lansing, MI 48824.

Your signature below indicates your voluntary agreement to participate in this study.

Signature

Date

1

APPENDIX H

Faking Study Protocol

PROTOCOL – Faking Study

College Board – Fall 2002 data collection

Before test administration:

Subjects:

1.) Find out how many subjects will be participating and bring enough of all supplies for each subject plus a few extras. (check subject pool for # of subjects signed up 30 minutes prior to proctoring).

How to check for number of subjects:

<http://psychology.msu.edu/SubjectPool/Welcome.asp> Sign in, view/modify experiment sessions on left side, then choose “Predictors of Student Success”, and look at “subjects signed up”.

2.) Print off the list of names.

Materials:

There are four slightly different forms for this study. Materials have been prepared and placed in envelopes numbered one through four. As forms have slightly different consent and personal information collection forms, all of these are already in the envelopes.

You will need to take enough materials so that you have (the number of participants divided by four) of each packet, plus a few extras in case of problems.

Please have these materials *organized* so that materials can be distributed fairly quickly.

- ☒ Sign-in sheet (fill in date, location, and proctor name). Take a new sign-in sheet each time.
- ☒ Sufficient copies of questionnaire envelopes
- ☒ Extra 10-option scantrons
- ☒ Debriefing forms
- ☒ Pencils -- please **sharpen** them all beforehand.
- ☒ Make sure you have a watch or some way to tell time with you.
- ☒ Stamp and stamp pad
- ☒ Green Sheets for giving credit

During the test administration:

Procedure:

- Arrive at least 15 minutes before the start of the session.
- Place the envelopes on the desks, every other seat, or spaced apart further than that if there are not many people signed up. Distribute the four different forms systematically, one through four, as you lay them out, so that the different forms are evenly distributed around the room.
- As people enter room, ask them to sign the sign-in sheet. Check ID to make sure that face, name, and what they write on the sign in sheet all match.
- If an individual shows up who has not signed up through the website, thank them for coming and ask them to sign up online for a different session. If subjects forget their ID but they are signed up, let them participate.
- As they come in, tell them that they may open their envelopes, and review the informed consent and date release forms while they wait for the session to begin. It is okay for subjects to sign forms in pencil.
- Start about 5 minutes after designated start time, or earlier if all participants have arrived. Those who are late may stay late.
- Read the script verbatim.
- Collect Informed Consent and Data Release.
- Time for test: 90 minutes.
- Participants will stop working at the end of the first section, and raise their hands to let you know they are done. When everyone has completed section one, you will tell everyone to begin section two at the same time. They should leave their envelopes containing section 2 under their seats until it is time to move on to section 2. *(This is done because the room having a coaching session will also be stopped at this point to begin coaching before moving on.)*
- As you administer the test, (a) at 7:00 PM post the time remaining as 30 minutes, and at 7:15 PM as 15 minutes,
(b) announce the time aloud as you post the time.
- If they finish early, you may start to collect their forms ~10 minutes before the end of the session.
- As you collecting the materials, check for the following, with the student:

- ☒ PID (on both scantron sheets)
 - ☒ Form # and A or B (on both scantron sheets)
 - ☒ Completed scantron sheets (should be completed through #106 on first scantron sheet and #97 on the second sheet)
 - ☒ Booklets with elaborations, with PID and name on the booklets
 - ☒ If a participant has incomplete information, ask them to stay to complete the form.
- When you collect materials, give the student the debriefing form.
 - Stamp the white cards that the students brought and fill in 3 half-hour credits. If they forgot their card, fill out a green sheet for them. Name of experiment is Predictors of Student Success. Experiment stamp # is MSU/PSYCH/101, Experimenter is Neal Schmitt.
 - If a participant decides not to participate, just let them leave and thank them for their time.
 - Do not answer any content related questions

Script

<Have envelopes on desks>

<As students enter>

“Hi. Please sign-in and may I check your ID? You may take a seat, open your packet and look at the consent forms, but do not begin.”

<check ID when they're signing in>

<Begin speech 5 minutes after designated starting time>

“Hello,

Thanks for participating in our study about the characteristics and experiences of college students. In order to ensure that the instructions are identical each time we run this study, I'm required to read these instructions verbatim. My name is *<say your name>* and I will be proctoring this data-collection session (“along with *<other person's name>*” – *add this if you have two people proctoring*). This session lasts about an hour and a half, and so let me say in advance that we appreciate your time and help with this project.

It is important that this is your first year of college, and that you have not previously participated in this study. If either of these criteria render you ineligible, please let me know now.

What I(we) have provided is an envelope containing two booklets containing questions for you to read over and complete. You'll fill out the scantrons, and record some answers on your booklets. I(we) will post and announce the time when you have only 30 minutes left to complete it. The time is posted just to make sure everything runs smoothly – you should have plenty of time, so don't worry about having to rush.

Please answer everything thoughtfully. It is important that you answer these questions seriously and follow the instructions closely. When you have completed the first section, please stop, and raise your hand. Do not move on to the second section until I ask you to do so.

When you are through with the entire study, please wait patiently until everyone else is finished with theirs. If you know now that you will not be able to stay for the entire hour and a half, please sign-up for a different session of the study where you will be able to stay for the entire duration.

Now, please read and sign the Informed Consent form, which provides more information on the nature of the task. With the Informed Consent form are two Data Release forms we also need you to sign. When you are finished we will collect those forms.”

<you collect the Informed Consent and Data Release forms >

“Please bubble-in your PID and form number and letter A on the first scantron immediately. Please write your PID at the top right corner of the question booklet. Keep the questionnaire for the first section on your desk, and place the envelope containing the second questionnaire packet under your seat.”

<Wait for subjects to fill in scantron and place envelope under seat>

“Please raise your hand any time if you have any questions or problems, for example, if there is a page missing in your packet or you don’t understand something you’re reading. Do you have any questions?”

You may now begin.”

<Note start time>

<When all students have raised their hands at the end of the first section, collect their forms and check scantrons for proper completion:>

“Please bubble-in your PID and form number and letter B on the second scantron. Please write your PID at the top right corner of the second section question booklet. **It is very important that you follow all instructions in this section carefully.** You may now begin the second section.”

<Time is up>

“Okay, time is now up. I will come by to collect your scantrons and packets. Please remain seated until all packets have been picked up and checked over quickly and you have received a debriefing form.

<Forms and scantrons are collected>

<Check to make sure they are completed properly>

<Pass out debriefing form>

“Thank you very much for participating! I can now stamp your white credit cards, if your teacher requires them. If you forgot yours, I have a slip I can fill out for you with instructions on how to exchange the slip for credit. We will also be giving you credit online through the subject pool. We hope you found this study rewarding or at least worth the extra credit. If you know anyone who may be interested and eligible, please have them sign up with the Subject Pool.”

===== **THINGS TO DO AFTER FINISHED** =====

1. Enter info. from sign-in sheet into designated folder in Fred’s lab.
2. Return materials to Fred’s lab in Baker. Place items in the appropriate location.

APPENDIX I

Coaching Directions

Coached Group:

“Shortly you will be given a questionnaire that is pretty different from the SAT or ACT, but it is a type of test that a high school student might take as a college admissions test. We would like to help you do your very best on this questionnaire. In order to do that, we will take a few minutes to discuss the types of questions that you may see in the questionnaire. We will explain the factors that will be important in selecting the best response options and answering the questions effectively.

Colleges are interested in several key performance dimensions. They are shown on this handout and we will read a definition for each dimension. We will look at an example of the type of question that you might see related to each dimension.

Biodata questions are scored on a scale of more or less positive. On the handout, the plus and minus sign show you which answers are more or less desirable on a given dimension.

The situational judgment questions have answer options that have been reviewed by a panel of judges. The judges have determined which are the best and worst answers for the dimension. The best and worst answer options are indicated for you.

Let’s look at the handout.

(Read dimensions and question examples on the handout.)

Now that you have an idea of what colleges might be looking for, you are better equipped to answer the questions that are coming up.”

APPENDIX J

Coaching Handout

Dimensions	Explanation										
Knowledge, learning, mastery of general principles (incl. GPA)	Gaining knowledge and attaining mastery of general principles to which one is exposed. This is not necessarily equivalent to grades, though grades can be an indicator.										
An example of a biodata question is: Think about those courses in high school you were most interested in. Generally how determined were you to learn the facts and concepts from the class? <table border="0"> <tr> <td>a. extremely determined</td><td>+</td></tr> <tr> <td>b. very determined</td><td></td></tr> <tr> <td>c. rather determined</td><td></td></tr> <tr> <td>d. sort of determined</td><td></td></tr> <tr> <td>e. not very determined</td><td>-</td></tr> </table> An example of a situational judgment question is: Your grade for a particular class is based on three exams, with no class attendance requirement. All of the homework requirements for the class are posted on the professor's web site. What would you do? - Attend class for as long as you feel that it is helping your grades. - Do all the homework but only go to some of the lectures. It's the exams that count. - Go to all the classes anyway. The professor may say something important. - Skip classes, but if you did poorly on the first exam, start going to classes. - There is no need to go to classes. Just get the homework done, and pass the exams. <i>Best Response?</i> c <i>Worst Response?</i> e		a. extremely determined	+	b. very determined		c. rather determined		d. sort of determined		e. not very determined	-
a. extremely determined	+										
b. very determined											
c. rather determined											
d. sort of determined											
e. not very determined	-										

Continuous learning, intellectual interest and curiosity	Being intellectually curious and actively interested in continuous learning both in core areas of study as well as in peripheral or novel areas.
--	--

An example of a biodata question is:

In the last two months, how many times have you watched or listened to a documentary (e.g., The History Channel, The Learning Channel, PBS, NPR)?

- a. never -
- b. 1-3
- c. 4-6
- d. 7-10
- e. more than 10 +

An example of a situational judgment question is:

You are enjoying a particular class, and find yourself frustrated that only limited material can be covered in class time. What would you do?

- a. Nothing. You would be frustrated but it's not a big deal.
- b. Ask the professor about obtaining some extra materials.
- c. Take the initiative and study outside of class. Find related materials on your own.
- d. Take a more advanced class on the topic. Either switch to one or take it in the future.
- e. Drop the class and just learn the material on your own, maybe by reading the same textbook.
- f. Ask the professor to cover more material in class.

Best Response? d

Worst Response? e

Multicultural tolerance and appreciation	Being open and tolerant to a multicultural environment. Positively impacting, contributing to, or actively seeking the benefits of a multicultural environment.
--	---

An example of a biodata question is:

In the last six months, how many times have you tried to talk to someone from a different country or culture just to learn about their background?

- a. never -
- b. once
- c. twice
- d. three or four times
- e. five times or more +

An example of a situational judgment question is:

You grew up in a small farming community and moved into a dorm area in which all students were from an urban background. They seem to have different concerns and interests and often just stare when you talk about your background and experiences. How would you react?

- a. Ask them questions about their experiences in the hopes that they will develop some interest in your background.
- b. Find other places to make friends with people who also come from farming communities.
- c. Try to talk to just one person, on their own, about what life was like for you growing up.
- d. Ask others about their experiences and ask if they have any questions of you.
- e. Voice your feelings about the staring, and limit your talking about your background.

Best Response? a or d

Worst Response? e

Leadership	Demonstrating leadership skills such as motivating others and coordinating groups, or otherwise performing a leadership role.
An example of a biodata question is: In high school, how often were you seen as the person that organized weekend activities with your friends? a. very often + b. quite often c. often d. not that often e. not very often - An example of a situational judgment question is: You notice that people are leaving soda cans and trash lying around in the common areas of the dorm. What would you do? a. Tell them to stop, and then tell the resident assistant. b. Collect the cans and throw away the trash, turning in the cans for money. c. Don't do anything; the maintenance crew will clean it up, as it is their job to do that. d. Post a sign to remind people not to litter, then bring it up at a floor meeting. e. Comment on the mess to your friends in the dorm, and not do anything about it. <i>Best Response? b or d</i> <i>Worst Response? e</i>	

Interpersonal skills	Exhibiting proficient oral communication skills, working well with others and possessing an awareness of and appropriate reaction to the social dynamics of a situation.
An example of a biodata question is: What is your attitude about working in a team or group (e.g., on work or classroom projects)? a. Very positive, I prefer to work in teams + b. Positive, I usually like to work in teams c. I am indifferent, it makes no difference whether or not I work in teams d. I usually prefer to work alone on individually assigned tasks that contribute to the team e. I always try to work alone, I dislike working on teams - An example of a situational judgment question is: One of your roommates has an embarrassing problem. He/she has a strong body odor and it is offensive to everyone who must interact with him/her. What would you do? a. Provide lots of hints such as making a joke or referring to the unpleasantness of your own body odor. Hopefully the person would realize the point. b. Tactfully tell the roommate about the problem and suggest a solution. One way to start could be to ask the roommate if he/she would like you to be honest and direct about any problems. c. Take actions to deal with the odor without confronting the roommate directly. You might open the windows or buy some air fresheners. d. Ask someone else to talk to your roommate about the problem or leave an anonymous message. e. Switch your room. f. Leave them alone. They probably know about it and cannot help the odor. <i>Best Response? b</i> <i>Worst Response? e</i>	

Physical and psychological health	Engaging in physically healthy behaviors (e.g. eating properly and exercising regularly) and avoiding physically unhealthy behaviors (e.g. alcohol and drug abuse and unprotected sex). Being mentally alert, stable, and engaged in a scholastic environment.
An example of a biodata question is: How much energy do you have compared with other people? a. More than anyone I know + b. More than most c. About the same amount d. Less than most e. Less than anyone I know - An example of a situational judgment question is: All of the people who live near you seem to party, drink and use drugs on weekend nights. You like most of these people, but do not want to engage in some of the behavior in which they engage. You have no one else to hang out with. How would you deal with this situation? a. Move away from that group, so that you are not getting into trouble. b. Continue to be friends and hang out with them, but do not engage in their activities. c. Join a club or find other friends, and tell your current friends they should understand. d. Try to teach them the problems associated with partying and using drugs. e. Start hanging out by yourself, and refer your friends to some professional help. <i>Best Response? b</i> <i>Worst Response? e</i>	

Perseverance	Following through on intended courses of action, "sticking to it", and being committed and dedicating time and effort to established goals and priorities. Taking personal pride in one's commitments.
--------------	---

An example of a biodata question is:

When encountering problems that take a long time to solve, how impatient have you become?

- | | |
|-------------------------|---|
| a. extremely impatient | - |
| b. very impatient | |
| c. somewhat impatient | |
| d. slightly impatient | |
| e. not at all impatient | + |

An example of a situational judgment question is:

An accounting class is a required course for the degree you are pursuing, and you have no interest in accounting. Your grades in the class are beginning to reflect your lack of interest. What would you do?

- a. Try your best to get through it, understanding that you really need the course, and don't want to have to take it a second time.
- b. Meet with a study group to get your grades up.
- c. Focus on asking questions to make it more interesting.
- d. Spend time with a tutor to boost your grades.
- e. See if you can find a way to relate accounting to your area of interest, to keep your enthusiasm up.

Best Response? b or d

Worst Response? c

Ethics and integrity	Exercising good judgment and adhering to a set of ethics. Demonstrating integrity by being honest and lawful, not cheating, and maintaining a sense of honor in one's personal and academic life.
----------------------	---

An example of a biodata question is:

How many times have you cheated on an exam?

- a. never +
- b. once
- c. twice
- d. three or four times
- e. five times or more -

An example of a situational judgment question is:

You have been having trouble in a course. On the day of the exam someone offers to sell you a copy of the exam. You have heard rumors that a large number of students in the class have purchased this exam and are afraid that if many students do well, your grade will look even worse. What would you do?

- a. Purchase the exam so that you do not look worse than everybody else.
- b. Take the exam with the knowledge that you have from studying.
- c. Approach the instructor and explain what you have been offered.
- d. Buy the exam, but try to miss some of the questions.
- e. Inform the instructor anonymously in the hopes that he will change the exam.

Best Response? e

Worst Response? d

APPENDIX K

College GPA Release Form

Authorization of Release of Cumulative GPA for Freshman and Sophomore Years

As part of a research project on the prediction of undergraduate student performance, I authorize the Office of the Registrar to release my GPA for each of the first four semesters of my academic work at MSU to a research team directed by Drs. Fred Oswald and Neal Schmitt. As soon as these investigators link these data to other test information I am providing them, any information identifying me personally will be removed from the data file. This information will not be released to any other research teams or individuals and all reports on this project will involve aggregated data that do not allow personal identification.

Signed _____

PID _____

Date _____

APPENDIX L

High School GPA and ACT/SAT Score Release Form

Authorization of Release of cumulative high school GPA and ACT/SAT Scores

As part of a research project on the prediction of undergraduate student performance, I authorize the MSU Office of Admissions to release my ACT and/or SAT scores to a research team directed by Drs. Fred Oswald and Neal Schmitt. As soon as these investigators link these scores to other test information I am providing them, any information identifying me personally will be removed from the data file. This information will not be released to any other research teams or individuals and all reports on this project will involve aggregated data that do not allow personal identification.

Signed _____

PID _____

Date _____

APPENDIX M

Twelve Dimensions of College Performance

Intellectual Behaviors

Knowledge, learning, and mastery of general principles (Knowledge)

Gaining knowledge and mastering facts, ideas and theories and how they interrelate, and understanding the relevant contexts in which knowledge is developed and applied.

Grades or GPA can indicate, but not guarantee, success on this dimension.

Continuous learning, and intellectual interest and curiosity (Learning)

Being intellectually curious and interested in continuous learning. Actively seeking new ideas and new skills, both in core areas of study as well as in peripheral or novel areas.

Artistic cultural appreciation and curiosity (Artistic)

Appreciating art and culture, either at an expert level or simply at the level of one who is interested.

Interpersonal Behaviors

Multicultural tolerance and appreciation (Multicultural)

Showing openness, tolerance, and interest in a diversity of individuals (e.g., by culture, ethnicity, or gender). Actively participating in, contributing to, and influencing a multicultural environment.

Leadership (Leadership)

Demonstrating skills in a group, such as motivating others, coordinating groups and tasks, serving as a representative for the group, or otherwise performing a managing role in a group.

Interpersonal skills (Interpersonal)

Communicating and dealing well with others, whether in informal social situations or more formal school-related situations. Being aware of the social dynamics of a situation and responding appropriately.

Social responsibility, citizenship and involvement (Citizenship)

Being responsible to society and the community, and demonstrating good citizenship.

Being actively involved in the events in one's surrounding community, which can be at the neighborhood, town/city, state, national, or college/university level. Activities may include volunteer work for the community, attending city council meetings, and voting.

Intrapersonal Behaviors

Physical and psychological health (Health)

Possessing the physical and psychological health required to engage actively in a scholastic environment. This would include participating in healthy behaviors, such as eating properly, exercising regularly, and maintaining healthy personal and academic relations with others, as well as avoiding unhealthy behaviors, such as alcohol/drug abuse, unprotected sex, and ineffective or counterproductive coping behaviors.

Career orientation (Career)

Having a clear sense of career one aspires to enter into, which may happen before entry into college, or at any time while in college. Establishing, prioritizing, and following a set of general and specific career-related goals.

Adaptability and life skills (Adaptability)

Adapting to a changing environment (at school or home), dealing well with gradual or sudden and expected or unexpected changes. Being effective in planning one's everyday activities and dealing with novel problems and challenges in life.

Perseverance (Perseverance)

Committing oneself to goals and priorities set, regardless of the difficulties that stand in the way. Goals range from long-term goals (e.g., graduating from college) to short-term goals (e.g., showing up for class every day even when the class isn't interesting).

Ethics and integrity (Ethics)

Having a well-developed set of values, and behaving in ways consistent with those values. In everyday life, this probably means being honest, not cheating (on exams or in committed relationships), and having respect for others.

Note. From "Developing a Biodata Measure and Situational Judgment Inventory as Predictors of College Student Performance," by F. L. Oswald, N. Schmitt, L. J. Ramsay, B. H. Kim, and M. A. Gillespie, *Journal of Applied Psychology*, in press.

APPENDIX N

Bogus Biodata Items

1. In the past six months, how often have you resolved disputes by isometric analysis?
 - a. never
 - b. once
 - c. twice
 - d. three or four times
 - e. five times or more
2. How often in the past year have you programmed in AJMR?
 - a. never
 - b. once
 - c. twice
 - d. three or four times
 - e. five times or more
3. How often, in the past three years, have you operated a rhetaguard?
 - a. never
 - b. once
 - c. twice
 - d. three or four times
 - e. five times or more
4. In the past six months, how often have you matrixed solvency files according to publication standards?
 - a. never
 - b. once
 - c. twice
 - d. three or four times
 - e. five times or more

Note. Items 1, 3 and 4 adapted from "Inflation Bias in Self-Assessment Examinations: Implications for Valid Employee Selection," by C. D. Anderson, J. L. Warner, and C. C. Spencer, 1984, *Journal of Applied Psychology*, 69, 577. Item 2 was written for this study.

APPENDIX O

Inflation Index Items

- 1 In the past year, in how many fundraisers have you participated?
 - a. none
 - b. 1
 - c. 2
 - d. 3
 - e. 4 or more

- 2 In the past month, how many times did you go out and learn more about something simply because it seemed interesting?
 - a. never
 - b. once
 - c. twice
 - d. three or four times
 - e. five times or more

- 3 If a fellow student offered you a copy of upcoming exam questions that he had retrieved from the teacher's recycling bin, how likely would you be to accept a copy?
 - a. extremely likely
 - b. quite likely
 - c. somewhat likely
 - d. not at all likely

- 4 How often in the past year have you volunteered to be the spokesperson for a group project you did at school or work?
 - a. much more often than most people
 - b. somewhat more often than most people
 - c. about as often as most people
 - d. somewhat less often than most people
 - e. a good bit less often than most people

Briefly describe up to 4 examples and the nature of the projects.

5 You are finding your freshman year very difficult. The courses are hard, and you feel your grades are not satisfactory. Material in class seems to be covered very quickly. What would you do?

- a. Talk with the professors and TAs to get help on how to study.
- b. Find a study partner and work on homework and class material together.
- c. Talk to your parents and an advisor.
- d. Study hard, try your best, and don't worry about it.
- e. Talk to my advisor and teachers; see if there are study groups or review sessions I can attend.
- f. Hire a tutor for the difficult classes.

1. What would you be most likely to do?

2. What would you be least likely to do?

6 You grew up in an environment that did not expose you to people from other ethnic groups. When you first get to college, you find that your roommate is of a different race, and you find some of their habits strange. What would you do?

- a. Seek out information about that person's culture.
- b. Try to accept the person as is and not judge her/him.
- c. Talk about your cultural differences and learn a new perspective.
- d. Be patient. Try not to ask her/him questions since that might make her/him feel uncomfortable.
- e. Ask the person questions about her/his habits and understand the situation.
- f. Politely ask her/him to stop the habits or just put up with it.

1. What would you be most likely to do?

2. What would you be least likely to do?

APPENDIX P

Calculation for Inflation Index Selection

Calculation Inflation Index Item Selection

	Optimal group (coached, motivated, not warned)			Reference group (not coached, not motivated, with warning)			Reference group mean minus optimal group mean	
	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>		
BIO1	45	2.31	1.35	46	1.78	1.11	-0.53	
BIO2	45	3.71	1.27	46	3.13	1.44	-0.58	
BIO3	44	2.64	1.48	46	1.93	1.39	-0.70	
BIO4	45	3.47	1.36	46	2.72	1.28	-0.75	
BIO5	45	2.78	0.90	46	2.39	0.61	-0.39	
BIO6	45	3.84	1.36	46	3.37	1.48	-0.47	
BIO7	44	3.50	1.53	46	3.07	1.62	-0.43	
BIO8	44	3.18	1.32	46	2.15	1.25	-1.03	**
BIO9	45	2.24	1.17	46	1.87	0.91	-0.37	
BIO10	45	3.24	1.32	46	2.04	1.11	-1.20	**
BIO11	45	2.96	1.49	46	2.59	1.60	-0.37	
BIO12	45	2.44	1.49	46	1.80	1.17	-0.64	
BIO13	45	3.22	1.59	46	2.76	1.55	-0.46	
BIO14	45	3.22	1.43	46	2.52	1.28	-0.70	
BIO15	45	2.31	1.35	46	1.74	1.14	-0.57	
BIO16	45	3.02	1.48	46	2.41	1.27	-0.61	
BIO17	45	1.62	0.98	46	1.50	0.89	-0.12	
BIO18	45	4.07	1.12	46	3.89	1.20	-0.18	
BIO19	45	3.33	1.55	46	2.65	1.42	-0.68	
BIO20	45	3.24	1.49	46	2.72	1.66	-0.53	
BIO21	45	4.29	1.14	46	4.26	1.24	-0.03	
BIO22	45	3.60	1.07	46	3.02	1.06	-0.58	
BIO23	45	3.60	0.99	46	3.20	0.83	-0.40	
BIO24	45	3.82	1.13	46	3.15	1.17	-0.67	
BIO25	45	4.24	0.71	46	4.09	0.69	-0.16	
BIO26	45	3.02	0.89	46	2.41	0.78	-0.61	
BIO27	45	3.91	1.02	46	3.65	0.99	-0.26	
BIO28	45	3.56	0.99	46	3.30	0.79	-0.25	
BIO29	45	4.38	0.72	46	4.30	0.84	-0.07	
BIO30	45	3.73	1.23	46	3.04	1.11	-0.69	
BIO31	45	4.33	1.02	46	4.30	0.96	-0.03	
BIO32	45	2.93	1.39	46	2.39	1.16	-0.54	
BIO33	45	4.00	0.98	46	3.85	0.82	-0.15	
BIO34	45	2.58	0.81	46	3.04	0.76	0.47	
BIO35	45	4.38	0.78	46	4.00	0.84	-0.38	
BIO36	45	3.93	0.86	46	3.74	0.91	-0.19	
BIO37	45	4.40	0.72	46	4.15	0.82	-0.25	
BIO38	45	3.53	1.14	46	3.13	1.36	-0.40	
BIO39	45	3.53	0.79	46	2.39	1.11	-1.14	**
BIO40	45	3.78	1.22	46	3.46	1.03	-0.32	

BIO41	45	3.47	0.81	46	3.22	0.94	-0.25	
BIO42	45	3.47	1.39	46	2.41	1.36	-1.05	**
SJI1C	45	1.76	0.57	46	1.26	0.74	-0.49	
SJI2C	45	0.91	0.95	46	-0.02	1.14	-0.93	**
SJI3C	45	1.00	0.71	46	0.98	0.71	-0.02	
SJI4C	45	0.09	0.73	46	0.02	0.83	-0.07	
SJI5C	45	0.71	1.12	46	0.26	1.00	-0.45	
SJI6C	45	0.53	1.06	46	0.39	0.88	-0.14	
SJI7C	45	1.47	0.76	46	1.07	1.00	-0.40	
SJI8C	45	1.24	0.86	46	0.85	1.15	-0.40	
SJI9C	45	1.29	1.22	46	1.30	0.89	0.02	
SJI10C	45	1.20	1.16	46	0.76	1.49	-0.44	
SJI11C	45	0.47	0.79	46	0.26	0.77	-0.21	
SJI12C	45	0.73	0.72	46	0.39	0.98	-0.34	
SJI13C	45	0.89	0.93	46	0.39	1.16	-0.50	
SJI14C	45	0.22	0.74	45	-0.07	0.75	-0.29	
SJI15C	45	0.44	0.97	46	0.15	0.84	-0.29	
SJI16C	45	1.27	0.81	45	1.18	1.03	-0.09	
SJI17C	45	1.13	1.14	46	-0.37	1.08	-1.50	**
SJI18C	45	1.42	0.84	46	1.26	0.74	-0.16	
SJI19C	45	0.96	0.82	46	0.43	0.78	-0.52	
SJI20C	44	0.86	0.82	46	0.83	0.90	-0.04	
SJI21C	43	1.05	0.65	46	1.02	0.86	-0.02	
SJI22C	43	1.07	1.33	46	0.74	1.36	-0.33	
SJI23C	44	0.91	0.74	46	0.37	0.93	-0.54	
SJI24C	44	1.20	1.00	46	0.48	1.13	-0.73	

**Inflation index items

APPENDIX Q

Biodata Item Types Coding Instructions

Biodata Item Type Coding

- 1=Not at all
- 2=Very small extent
- 3=Moderate extent
- 4=Great extent
- 5=Completely

Provided below are dimensions by which a biodata item can be described, along with their definitions. Please choose one dimension and rate all items on that dimension. Then, proceed to the next dimension.

Objective	The factual parts of an event that are not tied to one's interpretation or evaluation.
------------------	--

To what extent does this item refer to **objective** events?

Verifiable	Events that could be corroborated by an independent source; someone unrelated could agree with every statement that was made.
-------------------	---

To what extent does this item refer to **verifiable** events?

Controllable	Events where one has a choice about whether to perform or not.
---------------------	--

To what extent does this item refer to **controllable** events?

Equal access	Experiences or skills that everyone has an equal chance of obtaining.
---------------------	---

To what extent does this item refer to events with **equal access**?

College relevant	Events that seem to be related to understanding why one will do well in college.
-------------------------	--

To what extent does this item refer to events **relevant to college life**?

Invasive	Events that are too personal and/or invade one's privacy.
-----------------	---

To what extent does this item refer to **noninvasive** events?

Fakeableness	The extent to which respondents can produce false (but realistic) responses without being caught. This is an overall rating and not a single aspect of the item.
---------------------	--

To what extent does this item appear to be **fakeable**?

APPENDIX R

SJI Item Types Coding Instructions

SJI Item Types Scoring

Provided below are dimensions by which an SJI stem and response options can be described, along with their definitions. Please choose one dimension and rate all items (stem and response options) on that dimension. Then, proceed to the next dimension.

<i>Dimension Definition</i>	<i>Rating Instruction</i>
Objective: The factual parts of an event that are not tied to one's interpretation or evaluation.	To what extent does this item refer to objective events?
Verifiable: Events that could be corroborated by an independent source; someone unrelated could agree with every statement that was made.	To what extent does this item refer to verifiable events?
Controllable: Events where one has a choice about whether to perform or not.	To what extent does this item refer to controllable events?
Equal Access: Experiences or skills that everyone has an equal chance of obtaining.	To what extent does this item refer to events with equal access?
College Relevant: Events that seem to be related to understanding why one will do well in college.	To what extent does this item refer to events relevant to college life?
Invasive: Events that are too personal and/or invade one's privacy.	To what extent does this item refer to noninvasive events?
Fakeableness: The extent to which respondents can produce false (but realistic) responses without being caught. This is an overall rating and not a single aspect of the item.	To what extent does this item appear to be fakeable?

Recall that the questions are framed as:
 What would you be most likely to do?
 What would you be least likely to do?

Please use the following rating scale when determining the extent to which an item stem or response fits a dimension:

- 1=Not at all
- 2=Very small extent
- 3=Moderate extent
- 4=Great extent
- 5=Completely

REFERENCES

REFERENCES

- Alliger, G. M., & Dwight, S. A. (2000). A meta-analytic investigation of the susceptibility of integrity tests to faking and coaching. Educational and Psychological Measurement, 60(1), 59-72.
- Anderson, C. D., Warner, J. L., & Spencer, C. C. (1984). Inflation bias in self-assessment examinations: Implications for valid employee selection. Journal of Applied Psychology, 69(4), 574-580.
- Becker, T. E., & Colquitt, A. L. (1992). Potential versus actual faking of a biodata form: An analysis along several dimensions of item type. Personnel Psychology, 45, 389-406.
- Bunting, B. P., & Mooney, E. (2001). The effects of practice and coaching on test results for educational selection at eleven years of age. Educational Psychology, 21(3), 243-253.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd Ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Corr, P. J., & Gray, J. A. (1994). Relationship between attributional style and lie scores in an occupational sample motivated to fake good. Personality and Individual Differences, 18(3), 433-435.
- Cunningham, M. R., Wong, D. T., & Barbee, A. P. (1994). Self-presentation dynamics on overt integrity tests: Experimental studies of the Reid Report. Journal of Applied Psychology, 79, 643-658.
- Clevenger, J., Pereira, G. M., Wiechmann, D., Schmitt, N., & Schmidt-Harvey, V. (2001). Incremental validity of situational judgment tests. Journal of Applied Psychology, 86(3), 410-417.
- Dweck, C. S., & Leggett, E. L., (1988). A social-cognitive approach to motivation and personality. Psychological Review, 95, 256-273.
- Dwight, S. A., & Alliger, G. M. (1997). Reactions to overt integrity test items. Educational and Psychological Measurement, 57(6), 937-948.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? Human Performance, 16(1), 1-23.

- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. Journal of Applied Psychology, 84(2), 155-166.
- Elliot, S., Lawty-Jones, M. & Jackson, C. (1996). Effect of dissimulation on self-report and objective measures of personality. Personality and Individual Differences, 21(3), 335-343.
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. Journal of Personality and Social Psychology, 70, 461-475.
- Frei, R. L. (1998). Fake this test! Do you have the ability to raise your score on a service orientation inventory? Dissertation Abstracts International, 58, 5681.
- Friede, A., Gillespie, M., Kim, B., Oswald, F., Ramsay, L. & Schmitt, N. (2002). Development and Validation of Alternative Measures of College Student Success (Technical Report). East Lansing: Michigan State University.
- Goldberg, L. R. (1993). The structure of phenotypic personality traits. American Psychologist, 48, 26-34.
- Gore, B. A. (2001). Reducing and detecting faking on a computer-administered biodata questionnaire. Dissertation Abstracts International, 61, 4457.
- Graham, K. E., McDaniel, M. A., Douglas, E. F., & Snell, A. F. (2002). Biodata validity decay and score inflation with faking: Do item attributes explain variance across items? Journal of Business and Psychology, 16(4), 573-592.
- Harackiewicz, J. M., Barron, K. E., Carter, S. M., Lehto, A. T., & Elliot, A. J. (1997). Determinants and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. Journal of Personality and Social Psychology, 73, 1284-1295.
- Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of Achievement Goal Theory: Necessary and Illuminating. Journal of Educational Psychology, 94, 638-645.
- Harackiewicz, J. M., Barron, K. E., Tauer, J. M., & Elliot, A. J. (2002). Predicting success in college: A longitudinal study of achievement goals and ability measures as predictors of interest and performance from freshman year through graduation. Journal of Educational Psychology, 94, 562-575.
- Holden, L.M. (1996). The effectiveness of training in improving test performance. Presented at 11th Annual Conference for the Society of Industrial and Organizational Psychology, Inc., San Diego, CA.

- Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. Human Performance, 11(2-3), 209-244.
- Hough, L. M., Eaton, N. K., Dunnette, M.D., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. Journal of Applied Psychology, 75(5), 581-595.
- Jones, R. F. (1987). The effect of commercial coaching courses on performance on the MCAT. Journal of Medical Education, 61(4), 273-284.
- Jones, R. F., & Vanyur, S. (1987). A comparison of the predictive validity of the MCAT for coached and uncoached students. Journal of Medical Education, 61(4), 335-338.
- Klubeck, S., & Bass, B. M. (1954). Differential effects of training on persons of different leadership status. Human Relations, 7, 59-72.
- Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: are option-keyed instruments more resistant? Journal of Applied Psychology, 76(6), 889-896.
- Kluger, A. N., & Colella, A. (1993). Beyond the mean bias: The effect of warning against faking on biodata item variances. Personnel Psychology, 46(4), 763-780.
- Kulik, J. A., Bangert-Downs, R. L., & Kulik, C. C. (1984). Effectiveness of coaching for aptitude tests. Psychological Bulletin, 95, 175-188.
- Kulik, J. A., Kulik, C. C., & Bangert, R. L. (1984). Effects of practice on aptitude and achievement test scores. American Educational Research Journal, 21, 435-447.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. Psychological-Bulletin, 127(1), 162-181.
- Lautenschlager, G. J. (1994). Accuracy and faking of background data. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), Biodata handbook (pp. 391-419). Palo Alto, CA: CPP Books.
- Mabe, P. A. III & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta analysis. Journal of Applied Psychology, 67, 280-296.
- Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. Personnel Psychology, 44, 763-792.

- McCrae, R. R., & Costa, P. T. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), Handbook of personality: Theory and research (pp. 139-153). New York: Guilford.
- McDaniel, M. A., Morgeson, F. P., Bruhn Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Predicting job performance using situational judgment tests. Journal of Applied Psychology, 86(4), 730-740.
- McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. International Journal of Selection and Assessment, 9(1-2), 103-113.
- McFarland, L. A. (2000). Toward an integrated model of applicant faking (Doctoral dissertation, Michigan State University, 2000). Dissertation Abstracts International, 61, 2805.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. Journal of Applied Psychology, 85(5), 812-821.
- McManus, M. A. & Kelly, M. L. (1999). Personality measures and biodata: Evidence regarding their incremental predictive value in the life insurance industry. Personnel Psychology, 52, 137-148.
- Messick, S., & Jungeblut, A. (1981). Time and method in coaching for the SAT. Psychological Bulletin, 89(2), 191-216.
- Midgley, C., Kaplan, A., Middleton, M., Maehr, M., Urdan, T., Anderman, L. H., Anderman, E., & Roeser, R. (1998). The development and validation of scales assessing students' achievement goal orientations. Contemporary Educational Psychology, 23, 113-131.
- Miller, C. E. (2001). The susceptibility of personality selection tests to coaching and faking. (Doctoral dissertation, University of Akron, 2000). Dissertation Abstracts International, 61, 3888.
- Motowidlo, S. J., Russell, T. L., Carter, G. W., & Dunnette, M. D. (1988). Revision of the Management Selection Interview: Final report. Minneapolis, MN: Personnel Decisions Research Institute.
- Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. Journal of Applied Psychology, 75(6), 640-647.
- Mount, M., Witt, L. A. & Barrick, M. R., (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. Personnel Psychology, 299-323.

- Mumford, M. D., & Stokes, G. S. (1992). Developmental determinants of individual actions: Theory and practice in applying background measures. In M. D. Dunnette & L. M. Hough (Eds), Handbook of industrial and organizational psychology (Vol. 3, pp. 61-138). Palo Alto, CA: Consulting Psychologists Press.
- Nguyen, N. T. (2002). Faking in situational judgment tests: An empirical investigation of the work judgment survey. Dissertation Abstracts International, 63, 3109.
- Ones, D. S., Viswesvaran, C., & Reiss, A.D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. Journal of Applied Psychology, 81(6), 660-679.
- Oswald, F. L., Schmitt, N., Ramsay, L. J., Kim, B. H., & Gillespie, M. A. (in press). Developing a biodata measure and situational judgment inventory as predictors of college student performance. Journal of Applied Psychology.
- Pannone, R. D. (1984). Predicting test performance: A content valid approach to screening applicants. Personnel Psychology, 37, 507-514.
- Paulhus, D. L. (1988). Assessing self-deception and impression management in self-reports: The Balanced Inventory of Desirable Responding. Unpublished manual. Vancouver, British Columbia: University of British Columbia.
- Paulhus, D. L., & Reid, D. B. (1991). Enhancement and denial in socially desirable responding. Journal of Personality and Social Psychology, 60, 307-317.
- Petty, M. M. (1974). A multivariate analysis of the effects of experience and training upon performance in a leaderless group discussion. Personnel Psychology, 27, 271-282.
- Powers, D. E., & Rock, D. A. (1998). Effects of coaching on SAT I: Reasoning scores. New York, NY: The College Board.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. Journal of Applied Psychology, 83(4), 634-644.
- Ryan, A. M., Ployhart, R. E., Gregarus, G. J., & Schmit, M. J. (1998). Test preparation programs in selection contexts: Self-selection and program effectiveness. Personnel Psychology, 51(3), 599-621.
- Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection. In C. L. Cooper and I. Robertson (Eds.), International Review of Industrial and Organizational Psychology, (pp. 145-183). Oxford, England: John Wiley & Sons Ltd.

- Schmit, M. J., & Ryan, A. M. (1993). The Big Five in personnel selection: Factor structure in applicant and nonapplicant populations. Journal of Applied Psychology, 78(6), 966-974.
- Schmitt, N., Gillespie, M. A., Kim, B. H., Ramsay, L. J., Oswald, F. L., & Yoo, T. Y. (in press). Impact of elaboration on social desirability and the validity of biodata measures. Journal of Applied Psychology.
- Schmitt, N. & Kuncze, C. (2002). The effect of required elaboration of answers to biodata questions. Personnel Psychology, 55(3), 569-587.
- Schmitt, N., Oswald, F. L., Kim, B. H., Gillespie, M. A., & Ramsay, L. J. (in press). The impact of justice and self-serving bias explanation of the perceived fairness of different types of selection tests. International Journal of Selection and Assessment.
- Sisco, H. C. (1999). Biodata predictors of the five factor model. Dissertation Abstracts International, 59, 4520.
- Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. Journal of Applied Psychology, 87(2), 211-219.
- Stanush, P. L. (1997). Factors that influence the susceptibility of self-report inventories to distortion: A meta-analytic investigation. (Doctoral dissertation, Texas A and M University, 1997). Dissertation Abstracts International, 58, 2167.
- Stark, S., Chernyshenko, O. S., Chan, K., 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.
- Stokes, G. S. (1994). Introduction and history. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), Biodata handbook (pp. xv-xix). Palo Alto, CA: CPP Books.
- Stokes, G. S., & Searcy, C. A. (1999). Specification of scales in biodata form development: Rational vs. empirical and global vs. specific. International Journal of Selection and Assessment, 7(2), 72-85.
- Vasilopoulos, N. L. (1999). The impact of job familiarity and warning of response verification on the relationship between response latency and impression management: A field investigation. Dissertation Abstracts International, 59, 4521.

- Viswesvaran, C. & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. Educational and Psychological Measurement, 59(2), 197-210.
- Wiggins, J. S., & Trapnell, P. D. (1997). Personality structure: The return of the Big Five. In R. Hogan, J. Johnson, & S. Briggs (Eds.), Handbook of personality psychology (pp. 737-765). San Diego: Academic Press.
- White, L. A., Young, M. C., & Rumsey, M. G. (2001). ABLE implementation issues and related research. In J. P. Campbell & D. J. Knapp (Eds.), Exploring the Limits of Personnel Selection and Classification (pp. 525-558). Mahwah, NJ: Lawrence Erlbaum Associates, Inc., Publishers.
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. Applied Psychological Measurement, 20, 71-87.
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: an item-level analysis. Journal of Applied Psychology, 84(4), 551-563.

