Psychology Science, Volume 48, 2006 (3), p. 268-287

The impact of situational context variables on responses to biodata and situational judgment inventory items

LAUREN J. RAMSAY[1], NEAL SCHMITT, FREDERICK L. OSWALD,
BRIAN H. KIM & MICHAEL A. GILLESPIE

Abstract

Biodata measures and situational judgment inventories (SJIs) have been shown to be useful supplements to traditional selection tests in a variety of employment and educational settings. However, scores on both measures may be systematically biased when applicants are motivated and know how to perform well on the tests. This study examines the independent and joint effects of motivation, coaching, and warning not to fake on scores on biodata and SJI measures. Generally, coaching and motivation improved scores on these measures, and warning statements did not decrease scores. Item characteristics including objectivity, controllability, verifiability, and relevance were all shown to be related to biodata scores, as was the requirement to provide written elaboration on multiple-choice item responses. Based on our findings, we offer practical advice regarding the use of biodata and SJIs.

Key words: biodata, situational judgment, inflation, faking, coaching, motivation, warning, item type, elaboration

[1] Lauren J. Ramsay, Neal Schmitt, Frederick L. Oswald, and Brian H. Kim, Department of Psychology, Michigan State University; Michael A. Gillespie, Denison Consulting, Ann Arbor, MI.
We acknowledge the financial support of the College Board, who sponsored this research.
Correspondence concerning this article should be addressed to Lauren Ramsay, Department of Psychology, Michigan State University, 302 Psychology Building, East Lansing, Michigan 48824-1116; E-mail: ramsayl@msu.edu.

Biographical data (biodata) questions and situational judgment inventories (SJIs) have demonstrated promise as measures that predict performance across a broad array of criteria (Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004). Given the aim of improving the intellectual and cultural diversity in many academic and employment settings, the prospect of using selection instruments that cover both cognitive and non-cognitive domains is attractive for gathering a wealth of applicant information. However, compared with cognitive measures, applicants who are motivated to "fake" or distort their responses on non-cognitive measures are more able to do so (Hough, Eaton, Dunnette, Kamp, & McCloy, 1990; McFarland & Ryan, 2000; Viswesvaran & Ones, 1999), and thus such measures require thorough development and validation efforts. More specific to the selection process, applicants who fake to create a positive impression on a test may enhance their likelihood of success in being selected (Hough, et al., 1990), thereby increasing the error rates related to the selection decisions made on the basis of the test. Therefore, examining the susceptibility of biodata and SJI measures to faking increases our understanding of how these measures may function when implemented in practice.

To this end, the present study investigates the impact of different situational context variables on biodata and situational judgment inventory items. Specifically, the effects of being motivated to perform well, receiving coaching, and receiving a warning statement are simultaneously evaluated. While these three situational context factors have independently demonstrated some effect on scores (e.g., Cullen & Sackett, 2004; Dwight & Donovan, 2003; Viswesvaran & Ones, 1999), we are aware of no effort to investigate their additive and interactive effects at the same time. In addition, we explore an alternative item format for biodata proposed by Schmitt and Kunce (2002) that may affect respondents' motivation or ability to fake: respondents are required to provide written elaboration on their answers to certain biodata items. Further, we explore biodata item characteristics that may make some items more susceptible to faking than others (Becker & Colquitt, 1992; Graham, McDaniel, Douglas, & Snell, 2002); for example, items with responses that are more verifiable may be less likely to be faked. Finally, we consider how these factors affect the utility of biodata and SJIs designed for selection. While we fully recognize that individual differences also play a role in the inflation of responses to noncognitive measures (McFarland and Ryan, 2000; Ones, Viswesvaran, & Reiss, 1996), an examination of their contribution is not the objective of this research.

This paper will first discuss the literature relevant to situational context factors that may have an effect on non-cognitive measures such as biodata and SJI: coaching and motivation. Next we review research on ways to control inflation: warning statements, biodata item elaboration, and specific biodata item characteristics. We then build our specific hypotheses regarding biodata and SJI.

Although biodata and SJIs primarily measure non-cognitive characteristics, there are some fundamental differences. Appendix A provides a brief overview of biodata and SJI for the benefit of those readers who are less familiar with these instruments. Because biodata items are based on fact (typically past experiences), observed increases in biodata scores that are due to the testing context can be attributed primarily to faking where respondents distort their self-report of life experiences. This may compromise the construct validity of the measure. SJI questions, on the other hand, provide hypothetical scenarios, and respondents speculate about what they would do under such circumstances. The response options that each SJI item provides may increase the range of responses that the respondent would ordi-

narily consider, compared to the responses that would be generated if respondents had to provide the solutions themselves. Increases in SJI scores are likely to reflect faking based on improved knowledge of the dimensions underlying what each SJI item (and its response alternatives) measures, assuming respondents undergo test conditions where they feel motivated to provide the best responses possible. Given these differences in biodata and SJIs, we anticipate that three characteristics of testing – coaching to improve job knowledge, providing extrinsic motivation for respondents to do well, and including warning statements about dishonest responding – each may function differently for biodata and SJIs.

Situational context factors

Coaching. Test coaching may facilitate faking or it may lead to actual improvement in the domain or domains a test is designed to measure; either way, we assume that coaching will almost certainly occur when any measure is used to make high-stakes decisions. An examination of the shelves of any large bookstore provides ample evidence that there is a market for coaching manuals on high stakes tests. Coaching can simply be defined as getting external guidance on how to take tests (White, Young, & Rumsey, 2001). Such interventions have been shown to improve scores on selection tests (see discussion in Sackett, Burris, & Ryan, 1989), and non-cognitive test scores may be especially responsive to coaching compared with cognitive test scores. For instance, the Alliger and Dwight (2000) meta-analysis showed that both specific coaching and general instructions to "fake good" tended to improve scores on integrity tests. Cullen and Sackett (2004) demonstrated that a very short coaching session could raise scores dramatically on SJIs where responses could be arranged from least extreme to most extreme (versus our example, where each response is complex and cannot be ordered).

Dwight and Alliger (1997) showed that the job-relatedness of integrity items was positively related to their fakeability. This may be because job-relevant items are more transparent, which makes both coaching and job-desirable responding easier. Job-desirable responding reflects the recognition that individuals understand that, for particular jobs, different job behaviors or characteristics will be appropriate and desirable and can thus respond in a job-desirable way. Vasilopoulos, Reilly and Leaman (2000) found that the difference in mean SJI scores between those low on impression management and high on impression management grew as job familiarity grew. For job-relevant items, the process of coaching could render the job relevance of items in a biodata or situational judgment inventory, along with the test's dimensions, more overt (see Miller, 2001). Overt tests have been shown to be more easily faked than are covert tests (Alliger & Dwight, 2000).

Sackett et al. (1989) note that exercise-specific training, which is effectively 'training to the test', increases the effectiveness of coaching. Miller (2001) demonstrated that the Conscientiousness Biodata Questionnaire was readily faked in a study where participants were provided with a brief exercise-specific training program that provided trait-related information. Even coaching consisting simply of written statements about the content of the test, given prior to the test administration, has been effective in improving performance (Cunningham, Wong, & Barbee, 1994).

Brief coaching that describes the dimensions that biodata and SJI items intend to measure may be effective in raising scores on these tests in two ways; generating improved job

knowledge and generating increases due to more sophisticated faking based on this knowledge. As our biodata items are based on historical fact, we anticipate that score increases due to coaching will be primarily because of improved understanding of how to fake. As SJIs are thought to be based in part on job knowledge, we anticipate that increases due to coaching will be because of improved job knowledge, as well as faking.

Motivation. We assume that individuals who have an opportunity for personal gain as a result of a high score on a test will be motivated to do well. That gain may be a job offer, an offer of college admission, or a financial reward. Such motivation may then lead applicants to manipulate their responses in ways that enhance the likelihood of their success in a selection process (Hough, et al., 1990). When motivated to perform well, those responding to biodata items may be inclined to fake, and/or they may be more conscientious in thinking of historical events tapped by biodata questions. For SJIs, motivated respondents may improve their performance by reading more carefully and in general more effectively applying themselves to the task.

It is important to be able to understand the direct effects of motivation as well as how motivation interacts with other situational context variables in influencing biodata and SJI test performance. It may be the case that those who are highly motivated and receive coaching are relatively better-equipped to score highly on a selection test, given the same ability to fake. Motivation on its own may be an important factor in prompting higher scores, and coaching on its own may facilitate improved performance. However, we expect that the combined effect of these two factors may be especially powerful in that the individual now has both the capacity and the will to do their very best.

Inflation controls

Given the concerns regarding the inflation of measures that are not purely cognitive, and its potential impact on decision-making, there have been various attempts to control inflation. These include using warning statements to limit dishonest responding, presenting an elaboration requirement in biodata items, and using biodata items with certain characteristics (such as response verifiability) that make them less susceptible to inflation.

Warning. Dwight and Donovan (2003) found that warning statements were effective in lowering scores on noncognitive selection measures. Using the California Psychological Inventory, they demonstrated that when presenting a warning that included the risk that items may be verified to ensure honest responding as well as a negative consequence for dishonest responding, faking was minimized. Becker and Colquitt (1992) tested a warning statement that responses to biodata questions would be verified with other sources and found significant mean differences between a test-taking group with no warning, a faking group with no warning, and an applicant group that 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. This ineffectiveness of a warning statement in the case of an SJI could be because SJIs are presented and scored in a less straightforward manner (i.e., through empirical keying than through simple more-is-better scoring in a Likert format). Nevertheless, we posit that there may be some level of faking on SJIs when respondents are motivated to do

well, and a warning statement should precipitate more honest responses. We anticipate that scores on SJI will be lowered as a result of a warning statement, although warning may not have as powerful an effect on SJIs as it would for biodata. As our biodata items are based on fact and are susceptible to dishonest responding, we expect the warning statement to be effective in lowering scores on biodata.

We expect that the effectiveness of the warning statement may interact with the level of motivation in predicting biodata and SJI performance. That is, those who are motivated will be more susceptible to the effects of a warning statement in their responses to these non-cognitive measures. A similar interaction between warning and valence was demonstrated by McFarland (2000) when looking at openness to experience. Those with the greatest motivation to succeed may be those who have the most to lose by being caught faking, and may therefore be particularly responsive to inflation control attempts such as warning statements.

Biodata item elaboration. One score inflation control method that has shown promise is requiring that respondents elaborate on their responses to individual biodata items. In other words, respondents select a multiple-choice answer and then provide additional supporting information (e.g., naming the specific leadership positions held after indicating how many leadership roles have been held). A sample elaborated biodata item is shown in Appendix A. Schmitt and Kunce (2002) found that requiring elaboration for biodata reduces mean test scores by .7 to .8 standard deviations as compared to nonelaborated items. For tests containing only some items requiring elaboration, there were carry-over effects to nonelaborated items within the same instrument, where scores on those items were reduced by .25 to .40 standard deviations, although more recent work by Schmitt et al. (2003) did not replicate this carry-over effect.

Biodata item characteristics. Items may be more or less susceptible to faking based on their particular characteristics. Becker and Colquitt (1992) examined items in a biodata instrument using Mael's item type framework (1991, p.773). Specifically, items that were more likely to be faked were less historical, objective, discrete, verifiable, and external than other items. Such items were also more relevant to the job. Similarly, Elliot, Lawty-Jones, and Jackson (1996) found that responses to objective tests of personality were relatively unaffected by instructions to fake. Schmitt et al. (2003) found that biodata items that were more objective and verifiable were less correlated with the participants' BIDR self-deception and impression management scores. Dwight and Alliger (1997) conducted a study of ratings of individual integrity test items, finding that the perception that an item would be easy to fake was tied to the job relatedness and invasiveness of the item. In their meta-analysis of the susceptibility of integrity tests to coaching and faking, Alliger and Dwight (2000) concluded that overt tests were more susceptible to inflation and coaching than were covert tests. Similarly, Kluger and Colella (1993) suggest that the effect on overall scores may differ as a function of item transparency. Mael's (1991) taxonomy provides further hypotheses regarding the item characteristics that may be used sensibly in biodata items, including items that tap experiences for which respondents would have had equal access and personal control.

As item characteristics play an important role in making biodata items more or less resistant to inflation, we rated each of our biodata items on six characteristics, so that we could evaluate the interaction of item type with inflation under different situational constraints. These characteristics are the extent to which the item is: objective, verifiable, controllable, relevant to college performance, invasive, and fakeable. Since SJI items could not be charac-

terized along these dimensions, they were not rated and the effects of SJI item characteristics were not investigated.

Goals of the current study

We seek to address the following questions: First, how do different testing conditions, independently and in concert, influence how an individual performs on biodata and SJIs? Second, how does an elaboration requirement affect biodata scores across different conditions? Third, do certain biodata item characteristics make some items more susceptible to score changes than others?

While a laboratory setting in many ways does not parallel circumstances in an actual selection context, we would be remiss in implementing a new measure in a selection context without first evaluating the susceptibility of that measure to score inflation under different controlled conditions. 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 et al., 1994), and that warning statements can effectively suppress inflation (e.g., Dwight & Donovan, 2003). Possible interactions between these situational factors as well as the main effects associated with each factor are explored. Specifically, our hypotheses, based on the preceding discussion, are as follows:

H1a: Manipulations designed to increase motivation to achieve high scores will be positively related to scores on biodata.

H1b: Manipulations designed to increase motivation to achieve high scores will be positively related to scores on situational judgment.

H2a: Coaching will be related to higher scores on biodata.

H2b: Coaching will be related to higher scores on situational judgment.

H3a: Warning individuals that the accuracy of their statements may be verified and faking, if discovered, will have negative consequences, will be related to lower scores on biodata.

H3b: Warning individuals that the accuracy of their statements may be verified and faking, if discovered, will have negative consequences will be related to lower scores on situational judgment.

Interactions between these three manipulations are also expected. Specifically, those who are motivated and who receive coaching have a particularly strong advantage over those who experience only one or the other of these situational factors. They experience the synergy of having both the capacity and the incentive to do well.

H4a: Motivation and coaching will interact in their effect on biodata scores. That is, those who are motivated and coached will have higher scores on biodata than would be predicted by the two independent main effects.

H4b: Motivation and coaching will interact in their effect on situational judgment scores. That is, those who are motivated and coached will have higher scores on situational judgment than would be predicted by the two independent main effects.

Motivation and warning will interact such that those who are motivated will be more strongly influenced by a warning statement. That is, we anticipate that the warning will have a more powerful effect for those who are motivated, as those respondents would be the ones who feel they have more to lose by being caught faking. Those who

are not particularly concerned about the test outcome should not be as strongly affected by a warning as they have less to lose.

H5a: Motivation and warning will interact in affecting scores on biodata such that the warning will be more effective at lowering biodata scores for those in the motivated condition.

H5b: Motivation and warning will interact in affecting scores on situational judgment such that the warning will be more effective at lowering situational judgment scores for those in the motivated condition.

We have seen that biodata scores can be suppressed by an elaboration requirement and expect that this technique will limit faking.

H6: Requirements that biodata item responses be elaborated will lower participants' scores on the biodata scale.

Items that are less objective and are difficult to verify are more susceptible to inflated responses (Becker & Colquitt, 1992). Items that are more relevant to a job or academic situation for which a person is applying (Becker & Colquitt, 1992) and overt in nature (Alliger & Dwight, 2000) are also more likely to be faked. Accordingly, we expected that items viewed as less relevant to college student performance would be less likely to be inflated. Further, Mael (1991) has provided a taxonomy that suggests that responses to more invasive items may be more likely to be inflated. Finally, any item that addresses behavior or outcomes that are directly controlled by the respondent are seen as more likely to be faked than are those not under respondent control.

H7: Item objectivity and verifiability will be negatively correlated with biodata responses and relevance to college performance, invasiveness, controllability, and fakability will be positively related to biodata scores.

Method

Participants

The sample included 362 first-year undergraduate students at a large Midwestern university. The age of the participants ranged from 18 to 22 years, with a mean age of 18.4 (SD = .74). The mean ACT score was 23.5 (SD = 3.3). Of the sample, 79.3% were White; 6.1%, African American; 2.5%, Hispanic American; 8.6%, Asian American; 2.5%, other, and 1.1% did not declare their ethnicity. Women accounted for 58.8% of the sample.

Study design

The experiment was a 2 (coaching vs. no coaching) × 2 (motivation vs. no motivation) × 2 (warning vs. no warning) × 2 (elaborated vs. nonelaborated biodata items) design, with the first three factors as between subjects and the last factor as within subjects. Because there was no elaboration requirement on the SJI, only the between subjects factors were relevant in the analysis of SJI responses.

The study was advertised through the web page of the Psychology Department subject pool, where students were offered extra credit in psychology for their participation. Partici-

pation was restricted to freshman students who had not participated in similar studies we had conducted previously. Sessions were conducted in a classroom setting and were administered by trained proctors who followed a written protocol. Participants were randomly assigned to the eight experimental conditions. The questionnaire for this study was split into two booklets. 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 assigned to receive coaching experienced a brief coaching session (10 minutes), while those groups not targeted to receive coaching received nothing. There was no placebo treatment for the non-coached group. Both coached and non-coached groups then were administered the second booklet, which contained biodata and SJI questions. Instructional material at the beginning of the second test booklet provided the warning and motivational manipulations.

Measures

Biodata. Biodata items by Oswald et al. (2004) had been selected if they were empirically determined to be the best cross-validated predictors of college performance criteria. Criteria included first-year college GPA, absenteeism, and a self-assessment on behaviorally anchored rating scales. The 42 items selected represented the content of all 12 college performance dimensions (see Table 1). Coefficient alpha was .88.

Table 1:
Performance Dimensions Captured by the Biodata and Situational Judgment Inventory Measures

Intellectual Skills
- Knowledge and mastery of general principles
- Continuous learning, and intellectual interest and curiosity
- Artistic and cultural appreciation

Interpersonal Skills
- Appreciation for diversity
- Leadership
- Interpersonal skills
- Social responsibility and citizenship

Intrapersonal Skills
- Physical and psychological health
- Career orientation
- Adaptability and life skills
- Perseverance
- Ethics and integrity

Note. Adapted from "Developing a Biodata Measure and Situational Judgment Inventory as Predictors of College Student Performance," by F. L. Oswald, N. Schmitt, B. H. Kim, L. J. Ramsay, and M. A. Gillespie, 2004, Journal of Applied Psychology, 89, p. 189.

Situational judgment items. Situational judgment items generated by Oswald et al. (2004) were reviewed, and as with biodata, we selected those that had been empirically determined to be the best cross-validated predictors of the aforementioned college performance criteria. On dimensions where the criterion-related validity of the SJI items had been low, the best items were selected rationally based on content while ensuring that all 12 performance dimensions (see Table 1) were addressed in the overall scale. Coefficient alpha for the 24-item measure was .77.

Manipulations

Motivation. To encourage participants in the motivated conditions to do their best, we gave instructions that indicated that they should respond in a way that would present themselves favorably. We also offered a financial incentive for good performance. Specifically, the instructions in the motivated condition were: "Imagine that you are applying for admission to [your university], and your responses to these questions could influence the decision on whether or not you were accepted 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." Those in the motivation condition were also told that if they scored above the 50th percentile on the tests administered, they would receive $10. Note that both the message framing and the financial incentive together served as sources of motivation in the high-motivation condition. Our goal was not to investigate the influence of these sources independently.

Those in the nonmotivated conditions received no financial incentive. The instructions for those in the non-motivated condition removed the instruction to present oneself favorably: "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."

Warnings. To create an effect similar to that of warnings that may appear on college application materials, the materials for those in the motivation and warning conditions included the following warning statement, "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." Those in the warning condition without motivation received the following warning statement: "Note that we may verify a subset of your responses, and if you respond dishonestly, that may invalidate this test." The statements thus include the warnings of potential verification as well as negative consequences. Warning statements used in college admissions and other selection contexts tend to include either or both of these components.

Coaching. The proctor provided a 10-minute coaching session for the participants by reading aloud directions and providing written material that reviewed sample biodata and SJI items and provided definitions of the 12 performance dimensions that the items were intended to measure (see Table 1). The proctor also noted how the biodata and SJI items are scored. To ensure uniformity in coaching, the same proctor administered all coaching ses-

sions. Once the coaching session was complete, the proctor immediately began administering the second booklet in the study to avoid any discussion regarding the coaching.

Elaboration. Of the 42 multiple-choice biodata items, 22 required further elaboration in requesting a written response with more specific information (e.g., if you indicated that you were the leader of 3 clubs, you would then be asked to name the clubs). Items requiring elaboration here were those items that required elaboration in Oswald et al. (2004). The group of elaborated items and the group of non-elaborated items were parallel to the extent that each group tapped the same performance dimensions, although the specific items differed in wording.

Data analyses

To address the first question of our study and to test Hypotheses 1-5 regarding the situational factors that affect inflation, we conducted a 2 (coaching vs. no coaching) × 2 (motivation vs. no motivation) × 2 (warning vs. no warning) ANOVA for the SJI measures. The ANOVA on biodata responses included an additional factor (elaboration vs. no elaboration on the items) as a within-subjects factor to evaluate Hypothesis 6. Hypothesis 7 was evaluated by correlating ratings of item characteristics with mean item responses.

Results

Biodata

Analyses of variance results are shown in Table 2. Table 3 contains the means and standard deviations of responses to the biodata and SJI for all conditions along with *d*-values showing the effect size when scores are compared to those of the reference group (the group receiving no motivation, no coaching, and with a warning). As can be seen in Table 2, the motivational effect and the coaching effect are both statistically significant, and the means in Table 3 indicate that the effect was in the predicted direction, with coaching alone ($d = 0.33$) and motivation alone ($d = 0.27$) both producing small effect sizes, thus confirming our first two hypotheses for biodata, that motivation (Hypothesis 1a) and coaching (Hypothesis 2a) would tend to result in score increases. The warning effect was not statistically significant, indicating lack of support for Hypothesis 3a for biodata, although an examination of mean levels suggests that the warning statement did operate in the direction expected in the motivated conditions, though the interaction between warning and motivation was not statistically significant ($p > .05$). Those who received both coaching and motivation but no warning produced the largest effect size ($d = 1.04$) but when a warning statement was applied for those who received both coaching and motivation, the effect size was reduced ($d = 0.71$). Although the interaction between motivation and coaching (Hypothesis 4a) was marginally significant ($p < .06$) for biodata, an examination of the means for these conditions indicated that the combination of both motivation and coaching produced the largest biodata scores (Mean = 3.41) and the greatest effect size, as would be expected. Neither coaching nor motivation alone (Means = 3.09 and 3.06, respectively) produced as large an increment in biodata scores as that which occurred when both motivation and coaching were provided together.

Table 2:
Analysis of Variance Results for Biodata and Situational Judgment Inventory (SJI)

	Biodata		SJI	
Source	*df*	*F*	*df*	*F*
Between Subjects				
Coaching (C)	1	25.13*	1	40.88*
Motivation (M)	1	16.80*	1	14.75*
Warning (W)	1	0.39	1	1.41
C × M	1	3.70	1	9.95*
M × W	1	2.12	1	0.09
Error	356	0.42	354	
Within Subjects				
Elaboration (E)	1	728.23*		
E × C	1	0.61		
E × M	1	0.08		
E × W	1	0.03		
E × C × M	1	2.35		
E × M × W	1	0.08		
Error	356	0.05		

Note. $*p < .01$.

The interaction between motivation and warning (Hypothesis 5a) was not significant, but an examination of the pattern of means indicated that a warning that responses would be verified did appear to erase the inflation of responses that occurred in the motivated conditions, where money and instructions to do well were both provided. The largest effect of the warning statement was apparent by comparing the coached, motivated group without a warning (Mean = 3.41) and the coached, motivated group with a warning (Mean = 3.26). As noted above, the effect size dropped from $d = 1.04$ to $d = 0.71$ when a warning statement was applied.

The *F* value for the elaboration main effect was statistically significant, and means for elaboration conditions were much lower than similar means for non-elaborated conditions (*d*s are approximately 1.0 across conditions), providing support for Hypothesis 6. There was no interaction between the presence or absence of the requirement to elaborate on biodata items with any of the other manipulations.

Situational judgment inventory

In Table 2, we also present the ANOVA results for the SJI measure. Table 3 provides the corresponding means, standard deviations and effect sizes. As was true for biodata responses, the effects of motivation and coaching (Hypotheses 1b and 2b) were statistically significant and the means were in the expected direction. Coaching alone produced a *d*-value of 0.19, and motivation alone, $d = .13$. There was no effect of warning for SJI (counter to Hypothesis 3b); in fact, the means in the Warning conditions were slightly higher than those in the No Warning conditions. The interaction of motivation and coaching (Hypothesis 4b)

Table 3:
Descriptive Statistics for Biodata and Situational Judgment Inventory (SJI) across Study Conditions

	Biodata					
	Elaborated	Non-Elaborated	Combined	*d*	SJI	*d*
Motivation						
Warning						
Coaching	3.08	3.56	3.26	0.71	0.89	1.03
(*N* = 43)	(0.48)	(0.59)	(0.49)		(0.33)	
No Coaching	2.75	3.21	2.95	0.02	.58	0.06
(*N* = 45)	(0.42)	(0.47)	(0.40)		(0.35)	
No Warning						
Coaching	3.14	3.69	3.41	1.04	0.89	1.03
(*N* = 45)	(0.67)	(0.62)	(0.63)		(0.32)	
No Coaching	2.88	3.28	3.06	0.27	0.52	-0 .13
(*N* = 45)	(0.43)	(0.40)	(0.37)		(0.33)	
No Motivation						
Warning						
Coaching	2.86	3.32	3.08	0.31	0.67	0.34
(*N* = 44)	(0.40)	(0.32)	(0.31)		(0.37)	
No Coaching	2.71	3.20	2.94	REF	0.56	REF
(*N* = 46)	(0.45)	(0.49)	(0.45)		(0.32)	
No Warning						
Coaching	2.84	3.28	3.09	0.33	0.62	0.19
(*N* = 48)	(0.53)	(0.53)	(0.51)		(0.37)	
No Coaching	2.67	3.13	2.88	-0.13	0.50	-0.19
(*N* = 46)	(0.37)	(0.45)	(0.37)		(0.32)	

Note. SJI items are scored from -2 to +2, and biodata items from 1 to 5. REF refers to the reference group upon which d values (standardized mean differences) are calculated; this is the group receiving no motivation, no coaching, and with a warning. Means are presented for each condition; values in parentheses are standard deviations of responses in a given condition.

was statistically significant, and mean levels showed that those with both coaching and motivation (Mean = .89) fared better than those who received one or the other of these manipulations. Those experiencing both coaching and motivation had the highest SJI scores compared to the reference group (d = 1.03). Motivation and warning did not interact in predicting SJI performance (lack of support for Hypothesis 5b).

Summary. Hypotheses 1 and 2, that motivation and coaching would tend to increase both the biodata and situational judgment scores, were supported. Support for Hypothesis 3 was not found in the case of biodata or SJI, with the warning manipulation having no significant effect on scores. Hypothesis 4, that coaching and motivation would interact, was not supported in the case of biodata, but was supported for situational judgment responses. Hypothesis 5, that motivation and warning would interact, was not supported. Hypothesis 6, that elaboration would lower biodata scores, was supported.

Biodata item characteristics

We also conducted analyses to understand better the relationship between item characteristics and biodata scores as well as changes in this relationship under the varied situational constraints.

To provide an assessment of item type for biodata items using Mael's (1991) taxonomy, 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, relevant to college performance, invasive, and fakeable. On each biodata item, four of the judges made ratings on each of these six dimensions; ratings on a five-point scale ranging from "Not at all" to "Completely". Acceptable coefficient alpha estimates for the sums of four ratings were found: objectivity (.70), verifiability (.75), controllability (.67) and college relevance (.80), and these dimensions were retained for further analysis. Low reliability on the remaining 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, and thus there was no variability across items), or as a result of inconsistency in how judges rated items (e.g., there was little agreement on how invasive the items were). It would not be appropriate to retrain the judges in an attempt to increase the reliability of invasiveness, as our goal was to evaluate whether such a dimension was a useful mental framework as participants were evaluating items. In other words, if our judges were inconsistent in generating clear hypotheses about the invasiveness of an item, we did not expect our participants to do so.

The four judges' ratings of objectivity, verifiability, relevance, and controllability were averaged for each item. The correlations between the ratings for these four item types are shown in Table 4. These correlations indicate that verifiability and objectivity were very nearly identical as judged by the four raters; but items were judged differently when relevance and controllability were considered. For each item type, the average value of the ratings of the judges was then correlated with the response to each item for each respondent. These correlations were an indication of the degree to which item responses were related to the four item characteristics. Across all participants, the average correlations were -.22, -.18, -.15 and .11 for objectivity, verifiability, controllability, and relevance respectively. That is, participant responses to more objective, verifiable and controllable items were likely to be lower, and responses to items judged to be more relevant to academic performance were likely to be higher. The negative correlation between the judged controllability of items and

Table 4:
Biodata Item Characteristics: Correlation of Judges' Ratings between Item Types

	Objectivity	Verifiability	Controllability	Relevance
Objectivity	1.00			
Verifiability	.82*	1.00		
Controllability	.26*	.03	1.00	
Relevance	-.47*	-.47*	-.07	1.00

*$p < .01$

item responses is inconsistent with our hypothesis and Mael (1990) that the relationship would be positive.

Further analyses of the correlations between item characteristics and biodata responses were conducted to determine if these correlations were affected by any of the situational factors. Following Cohen, Cohen, West and Aiken (2003, p. 240), ANOVA was used to test for differences in Fisher-*z* transformed values of the correlations presented in Table 5 across the eight experimental conditions. Separate ANOVAs were conducted for each of the four item characteristics.

Table 5:
Biodata Item Characteristics: Mean Correlation with Biodata Item Responses across Conditions

	Objectivity		Verifiability		Controllability		Relevance	
Motivation		*d*		*d*		*d*		*d*
Warning								
Coaching	-0.18	0.55	-0.15	0.30	-0.11	0.53	0.12	0.13
(*N* = 43)	(0.21)		(0.21)		(0.18)		(0.15)	
No Coaching	-0.23	0.32	-0.20	0.05	-0.16	0.20	0.09	-0.07
(*N* = 45)	(0.21)		(0.22)		(0.17)		(0.15)	
No Warning								
Coaching	-0.16	0.64	-0.16	0.25	-0.10	0.60	0.12	0.13
(*N* = 45)	(0.24)		(0.23)		(0.19)		(0.14)	
No Coaching	-0.20	0.45	-0.17	0.20	-0.17	0.13	0.12	0.13
(*N* = 45)	(0.18)		(0.18)		(0.16)		(0.13)	
No Motivation								
Warning								
Coaching	-0.20	0.45	-0.17	0.20	-0.16	0.20	0.08	-0.13
(*N* = 44)	(0.19)		(0.22)		(0.17)		(0.18)	
No Coaching	-0.30	REF	-0.21	REF	-0.19	REF	0.10	REF
(*N* = 46)	(0.22)		(0.20)		(0.15)		(0.15)	
No Warning								
Coaching	-0.22	0.36	-0.19	0.10	-0.15	0.27	0.06	-0.27
(*N* = 48)	(0.20)		(0.21)		(0.17)		(0.15)	
No Coaching	-0.28	0.09	-0.22	-0.05	-0.19	0.00	0.15	0.33
(*N* = 46)	(0.19)		(0.21)		(0.13)		(0.18)	

Note. Standard deviations are in parentheses. Fisher z values are equivalent to the correlational counterpart that is listed (within .01). REF refers to the reference group upon which d values are calculated; this is the group receiving no motivation, no coaching, and with a warning.

The correlation between item objectivity and item responses varied significantly as a result of coaching ($F_{1, 354} = 7.08$, $p < .01$) and motivation ($F_{1, 354} = 5.52$, $p < .05$). It should be noted that all the z values are negative, indicating that the more objective the item was judged to be, the lower were the students' scores on the biodata items. As expected, the strongest relationships between biodata items and rated objectivity were in the two conditions without motivation or coaching, and the weakest were in the two groups with both motivation and coaching. The effect size when compared to the reference group was largest ($d = 0.64$) when both coaching and motivation were provided, without a warning. When a warning was provided to the coached and motivated group, the effect size was reduced slightly to $d = 0.55$.

The relationship between item verifiability and item performance did not vary significantly across conditions. The strongest effects for verifiability were in the coached and motivated conditions with a warning ($d = 0.30$) and without a warning ($d = 0.25$).

For controllability, the relationship between item performance varied significantly as a result of the manipulation of coaching ($F_{1, 354} = 7.27$, $p < .01$) and motivation ($F_{1, 354} = 5.61$, $p < .05$). For controllability the largest gap in mean correlations was between groups without coaching or motivation and the groups that have both coaching and motivation. The effect size when compared to the reference group was largest ($d = 0.60$) when both coaching and motivation were provided, without a warning. When a warning was provided to those who were coached and motivated, the effect size was reduced slightly to $d = 0.53$. Relationships between controllability and biodata scores were once again negative, but slightly lower than those between objectivity and test scores for the biodata measure.

Judgments of the relevance of the items to the college setting were positively correlated with responses, but the overall correlation was only .11. These correlations did not vary significantly as a function of the main effects of any of the manipulations, but there was a significant coaching by motivation interaction ($F_{1, 354} = 4.39$, $p < .05$). The nature of this interaction suggests that those who are both motivated and coached have a greater likelihood of scoring higher on biodata items when those items appear relevant to academic performance, versus those who only experience the motivated or coached conditions alone, or neither of these conditions. This also suggests that coaching on the nature of biodata items is effective in helping individuals who are motivated to identify dimensions that are relevant to college performance.

These results provide support for the fact that item objectivity, verifiability, and controllability do produce modestly lower scores on biodata items, whereas relevance produces a slight increase in scores. Situational factors, represented by our manipulations, had a moderate impact on the size of these relationships for controllability and objectivity, but had little impact on the correlations between relevance and item responses or verifiability and item responses.

Discussion

This study has addressed three major questions. First, how do different testing conditions, independently and in concert, influence how an individual performs on biodata and SJIs? Second, how does an elaboration requirement affect biodata scores across different

conditions? Third, do certain biodata item characteristics make some items more susceptible to score changes than others?

Situational factors were demonstrated to be an important contributor to both biodata and SJI performance. Brief coaching was shown to improve scores on both of these noncognitive measures, an important issue when results may have high-stakes outcomes, such as in a selection situation. Coaching that includes the dimensions being captured by the measures appears effective in helping examinees to generate hypotheses about what characteristics are desirable and then presenting themselves appropriately. If these tests were to be used in a college admissions selection process, those who have access to coaching would likely be better able to improve their scores on biodata and situational judgment inventories. It is also the case, though, that even brief informational coaching may provide applicants with equal opportunity to present themselves favorably – and in the future such information may easily be found on the Internet or in popular-press magazines. Motivation was shown to have a significant effect on performance for biodata and SJI, and the significant interaction of coaching and motivation for SJIs (and the marginally significant interaction for biodata) suggests that maximum performance on these tests will occur when both extrinsic motivation and coaching are present or provided. To avoid inequity in selection due to unequal access to coaching, it may be advisable to offer coaching to all test takers. Given that the coaching was so brief and straightforward, it should be relatively easily provided to all test takers.

Warning statements, although not found in this study to be a significant factor in suppressing biodata inflation, did interact and operate in the direction expected for biodata when individuals are motivated to perform well. Warning statements had no effect on SJI performance. Warning statements may be relevant for biodata inventory administration as a way to limit dishonest responding, but they appear unnecessary with SJIs, and may actually make it salient to the respondent that responses to SJIs can be readily faked.

As found in previous research (e.g., Schmitt & Kunce, 2002, Schmitt et al., 2003), the elaboration technique does appear to lower biodata scores, perhaps for a variety of reasons: its check on the veracity of responses, its challenge to recall the details associated with responses, the unpleasant or fatiguing effect of responding, or some combination of these. Responses to elaborated biodata items were nearly one standard deviation lower than responses to non-elaborated items.

Understanding biodata item characteristics can serve to create a test that is less vulnerable to score inflation. We found that biodata items more susceptible to inflation were judged as being less objective, controllable, and verifiable, and more college relevant. For objectivity and controllability, this relationship also varied as a function of our manipulations, particularly motivation and coaching. To limit inflation, it may be important for biodata test items to be objective when possible. On the other hand, our results suggest that reducing the college relevance of items would make them less fakeable. However, we would not recommend reducing relevance. As suggested by Graham et al. (2002), verifiability of items was negatively related to item responses (-.18), but the correlation was small and it did not vary as a function of any of the manipulations. That less controllable items, counter to what we expected, were more susceptible to inflation, suggests that further research may be necessary to investigate the different types of controllability as identified by Graham et al. (2002). Judged relevance of the items to college student performance was positively related to item responses. These correlations were significantly higher under coached and motivated conditions, but none of the mean correlations in the various conditions exceeded .20.

One possible limitation of the elaboration results is that required elaboration may have been on items that were more verifiable. However, an examination of the judges' ratings of the verifiability of the elaborated items (Mean = 2.84, *SD* = 0.95) versus those of the non-elaborated items (Mean = 2.58, *SD* = 1.27) showed only a small difference (*d* = 0.24). Nevertheless, this verifiability, perhaps in combination with the elaboration requirement itself, may have limited the respondents' inclination to inflate responses. Subsequent research should be conducted that does not confound elaboration and item verifiability.

Another limitation of the study is the slightly different implications of the 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 discovered providing dishonest answers had a greater potential negative consequence for members of the motivated group, who would lose a cash payout for high performers, although statistically such an interaction was not found. Also, 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 and the generalizability of our results to high-stakes testing situations. To that extent our results are conservative; we suspect that the impact of motivation, at least, would be greater in real-world contexts.

Conclusions and implications

A respondent can improve his or her score on biodata and SJI when motivated, or when coached very briefly on how to do so. Based on the body of research that documents the effectiveness of warnings (Dwight & Donovan, 2003), as well as our own results, we recommend that test users consider implementing a warning statement in biodata administration.

In compiling an inventory of biodata items, test developers should consider items with a format and content that limit inflation. This may be achieved by focusing on items that are written in a way that maximizes their objectivity and verifiability and obscures their direct relevance. Controllability of item content should also be considered, but further research is necessary regarding the dimension of controllability. Adding an elaboration requirement for biodata items also seems to be an effective method of reducing item inflation.

Perhaps the most significant research needs are to evaluate the feasibility of using these measures in an actual admissions context or similar items in an employment context. Our best set of practical recommendations for reducing response inflation in biodata are the following: (a) Use objective biodata statements; (b) Require elaboration on biodata items; (c) Use warning statements that indicate that biodata responses will be checked for accuracy and that providing inflated responses may invalidate their test scores; (d) Try to balance the situational relevance of the item content with the ability to fake on those items. SJI scores, like biodata, are subject to coaching and motivational effects. Unequal access to coaching could precipitate subgroup differences in test performance, one of the issues that these tests are designed to avoid. We therefore suggest that companies or academic institutions that develop and administer SJI and biodata measures provide brief coaching that includes a description of what is being measured to all examinees. Although this may facilitate faking to some extent, the measures would likely still have the variability to allow for criterion-related validity. Importantly, brief coaching before the tests were administered would mini-

mize any gaps in scores precipitated by unequal access to coaching, gaps that may be related to racial subgroups and thus to adverse impact.

Appendix A

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 (cf. Stokes, 1994). A sample biodata item measuring leadership is:

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

Although studies have demonstrated the utility of biodata items in predicting job performance, other studies have shown that they are not impervious to inflation. As noted by Kluger, Reilly and Russell (1991) this may be a consequence of using continuous response scales rather than empirically keyed item response options. We assume that with a continuous response scale, the direction of the desirability of responses on the continuum may be more easily determined by the respondent, and fakers can operate effectively (e.g., in the sample item just provided, the last response option is best).

Elaborated biodata

A sample biodata item with an elaboration requirement is:

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)?

a) 0
b) 1-3
c) 4-7
d) 8-12
e) 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.

_____
_____
_____
_____

Situational judgment inventory

Sometimes regarded as job knowledge tests or low fidelity simulations, situational judgment inventories (SJIs) typically comprise a sample of hypothetical scenarios for which the respondent is asked to endorse the response representing the most appropriate course of action (and sometimes the least appropriate course of action as well). A sample SJI item 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?

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.

In many cases SJI content bears some similarity to job knowledge tests, but SJIs may still permit some level of score inflation, even when empirical keying is used. Respondents may be able to raise their scores, depending on the phrasing of the test questions (Nguyen, 2002). When the question phrasing requires speculation about how one might behave in a hypothetical scenario, that may encourage an individual to respond in a socially desirable way (also see Ployhart & Ehrhart, 2003, who also found instructional effects for SJIs).

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, 59-72.

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.

Cohen, J. Cohen, P., West, S. G., & Aiken, L. S. (2003). Applied multiple regression/correlation analysis for the behavioral sciences (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Cullen, M. J., & Sackett, P. R. (2004, April). Threats to the operational use of situational judgment tests in a student admission context. In P. R. Sackett (Chair), New developments in SJTs: Scoring, coaching and incremental validity. Symposium conducted at the 19th Annual Conference of the Society for Industrial and Organizational Psychology, Chicago, Illinois.

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.

Dwight, S. A., & Alliger, G. M. (1997). Reactions to overt integrity test items. Educational and Psychological Measurement, 57, 937-948.

Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? Human Performance, 16, 1-23.

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, 573-592.

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, 581-595.

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

Kluger, A. N., Reilly, R. R., & Russell, C. J. (1991). Faking biodata tests: Are option-keyed instruments more resistant? Journal of Applied Psychology, 76, 889-896.

Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. Personnel Psychology, 44, 763-792.

Maurer, T., Solamon, J., & Troxtel, D. (1998). Relationship of coaching with performance in situational employment interviews. Journal of Applied Psychology, 83, 128-136.

McDaniel, M. A., & Nguyen, N. T. (2001). Situational judgment tests: A review of practice and constructs assessed. International Journal of Selection and Assessment, 9, 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, 812-821.

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.

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, 660-679.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A.(2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. Journal of Applied Psychology, 89, 187-207.

Ployhart, R. E., & Ehrhart, M. G. (2003). Be careful what you ask for: Effects of response instructions on the construct validity and reliability of situational judgment tests. International Journal of Selection and Assessment, 11, 1-16.

Sackett, P. R., Burris, L. R., & Ryan, A. M. (1989). Coaching and practice effects in personnel selection. In C. L. Cooper & I. Robertson (Eds.), International review of industrial and organizational psychology (Vol. IX, pp. 145-183). Oxford, England: Wiley.

Schmitt, N., & Kunce, C. (2002). The effect of required elaboration of answers to biodata questions. Personnel Psychology, 55, 569-587.

Schmitt, N., Oswald, F. L., Kim, B. H, Gillespie, M. A., Ramsay, L. J., & Yoo, T. Y. (2003). Impact of elaboration on socially desirable responding and the validity of biodata measures. Journal of Applied Psychology, 88, 979-988.

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.

Vasilopoulos, N. L., Reilly, R. R., & Leaman, J. A. (2000). The influence of job familiarity and impression management on self-report measure scale scores and response latencies. Journal of Applied Psychology, 85, 50-64.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. Educational and Psychological Measurement, 59, 197-210.

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: Erlbaum.