2008

Item and Person Characteristics as Predictors of Faking

Nicholas Tyler Day
Wright State University

ITEM AND PERSON CHARACTERISTICS AS PREDICTORS OF FAKING

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

By

NICHOLAS TYLER DAY

B.S., Bowling Green State University, 2003

2008

Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

12/01/08

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY Nicholas Tyler Day ENTITLED Item and Person Characteristics as Predictors of Faking BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science .

David LaHuis, Ph.D.
Director, Thesis Committee

John M. Flach, Ph.D.
Chair, Department of Psychology

Daniel L. Weber, Ph.D.
Graduate Program Director

Committee on
Final Examination

David LaHuis, Ph.D.

Corey E. Miller, Ph.D.

John M. Flach, Ph.D.

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies

ABSTRACT

Day, Nicholas Tyler, M.S., Department of Psychology, Wright State University, 2008. Item and Person Characteristics as Predictors of Faking.

Applicants may be more motivated to fake than incumbents and may fake more on some items than others. The present study investigated both item and person characteristics as predictors of faking. At the item level, both item transparency and job-relevance were hypothesized to be associated with higher levels of faking. In contrast, item verifiability was hypothesized to be associated with lower levels of faking. At the person level, applicants were expected to have a higher prevalence of faking than incumbents. Data was taken from an existing pool of applicants (n = 507) and incumbents (n = 302) at a customer calling center. The study was performed using a multilevel-logistic regression (MLR) approach to estimating person response curve (PRC) for results for Conscientiousness, Extraversion, and Openness items. None of the item-level results were significant for Conscientiousness, but the analyses found significant item-level effects for Extraversion and Openness. First, item transparency was related to higher levels of faking. Also, individuals were more likely to fake for items of low verifiability than items of high verifiability. Unexpectedly, individuals were more likely to fake for items of low job-relevance than items of high job-relevance. The results for person-level effects showed that applicants exhibited substantial model fit over incumbents, although incumbents appeared to have higher levels of faking than incumbents. The results and implications are discussed.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

INTRODUCTION

During the application process, many applicants are likely to exaggerate or distort their personal attributes (e.g., Heron, 1956; McFarland & Ryan, 2000; Rosse et al., 1998). For instance, applicants may make themselves appear more dependable or energetic than they are in reality. Perhaps the easiest way for applicants to "fake good" is on their responses to personality scales. Applicant faking has become a major concern for personnel psychologists. This type of strategic faking in applicants may yield inaccurate test scores, which can compromise the use of test results in a selection context (e.g., Dunnette et al., 1962; Holden & Jackson, 1981; Pannone, 1984). Hough et al.'s (1990) review found no overall difference in criterion-related validities for faking and nonfaking applicants; however, other studies have shown that faking may affect the rank ordering of applicants (McFarland & Ryan, 2000; Christiansen et al., 1994; Rosse et al., 1998). Although there are numerous studies on the outcomes of faking, not many researchers have examined factors that may contribute to faking (McFarland & Ryan, 2000; Robie, 2006). I am not aware of research that simultaneously models the effects of both test item features and contextual factors on applicants faking on personality tests. It may be that certain test items lead to higher levels of faking than others, and that some applicants may fake to a larger extent than others.

In the present study, I investigated three item features: item verifiability, item transparency, and item job-relevance. I hypothesize that items that are less verifiable more transparent, and more job relevant will be associated with higher levels of faking. In

addition, I suggest that applicants will fake more and score higher than incumbents. Because these variables are on two levels of analysis, I used the multilevel logistic regression (MLR) person response curve (PRC) framework outlined by LaHuis and Copeland (in press). MLR uses estimates from the two-parameter logistic (2PL) item-response theory (IRT) or graded response model. I examined faking on each factor of the Big Five: Conscientiousness, Extraversion, Emotional Stability, Agreeableness, and Openness.

In the following sections, I will briefly review the 2PL model. This model is based on a dichotomous type of response pattern, where individuals answer either positively ("endorsed") or negatively ("nonendorsed"). Next I will describe the MLR approach to estimating PRC's (Reise, 2000) and how it can be used to test hypotheses related to applicant faking. For tests with polytomous items, I will explain using an extension of the 2PL model called the graded response model (GRM), which has been developed by Samejima (1969; 1996). Finally, I will describe the theoretical rationale for my hypotheses regarding faking.

Two-Parameter Logistic Model

IRT has many applications for studying test behavior (Baker, 2001; Embretson & Reise, 2000). In particular, the standard 2PL model has proven useful for analyzing personality/dichotomous data (Reise & Waller, 1990; Waller & Reise, 1989).

The equation for the 2PL model is:

$$P_{ij}\,(Y=1|\theta_j) = \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1+\exp[\alpha_i(\theta_j - \beta_i)]} \qquad (1)$$

The 2PL model specifies the probability of endorsing item *i* for person *j* as a function of a person's trait level (θ_j), an item's discrimination (α_i), and an item's difficulty (β_i). This

equation can be used to plot item response curves (IRC's), which demonstrate the functioning of an individual item (see Figure 1). Trait levels (θ) are assumed to have a mean of 0 and a standard deviation of 1, and item difficulty (β) is placed onto the same metric as θ.

The item discrimination parameter (α_i) are the slope of the IRC's, which represents the ability of the item to discriminate accurately between high and low trait levels. As seen by Figure 1, the dotted item has higher discrimination than items 1 or 3. Items with high discriminations have steeper slopes, and there is a clearer division between high and low trait levels. Item discriminations typically range usually from 0.75 to 1.75.

The IRCs' locations are determined by item difficulty (i.e., item threshold). In Figure 1, the first and second items both have a difficulty of zero, because average trait level on the X-axis corresponds to a 0.5 probability of endorsement on the Y-axis. In other words, respondents are 50% likely to endorse an item where their trait level matches item difficulty. Probability changes from 50% as trait level is estimated to be greater or less than the item's difficulty. This change in probability levels out, as individuals' trait levels get farther away from the item's difficulty. As can be seen by Figure 1, the dashed IRC has an item difficulty of 1.0 and appears further the right, because a 0.5 probability corresponds to a higher theta value. An Item Response Curve with lesser item difficulty would appear shifted to the left of the center theta value.

In general, items are modeled to represent different degrees of the latent trait being measured. The most difficult items can be useful for isolating individuals at top trait levels, whereas other items may be useful for partitioning individuals at a low-to-

moderate standing on the trait. In this way, item difficulty can be thought as the item's location threshold for estimating trait level.

Item and Person Fit

Item fit is assessed by comparing the model-implied responses with observed responses, using available statistical indices (e.g., a chi-square test). Chernyshenko et al. (2001) describe a procedure for computing 2PL model fit using chi-squares adjusted for sample sizes.

If item fit is adequate, person fit may also be examined. That is, item parameters and trait levels can be used to verify the probability of individual response patterns. Person fit can be assessed using the slopes of person response curves (PRC's). A PRC describes the relationship of how the probability of item endorsement decreases as item difficulty increases. The equation for graphing a PRC is:

$$P_{ij}\ (Y=1|\beta_i) = \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1+\exp[\alpha_i(\theta_j - \beta_i)]} \qquad (2)$$

The 2PL model predicts that individuals are unlikely to endorse items with much higher difficulty levels than their estimated trait level. Thus, person fit is evaluated by examining the negative slope of the PRC's for responses to items of increasing difficulty. A strong negative slope would indicate excellent person fit – because the probability of endorsement decreases markedly as item difficulty increases (see Figure 2; Honest curve). A less negative slope would indicate a lack of person fit – because examinees are answering questions of higher quality than should be expected by their trait levels (see Figure 2; Faking curve). Poor fit indicates that examinees have a relative lack of correspondence between their trait levels and response patterns. In selection settings, lack of person fit can most likely be attributed to faking on certain items.

MLR and Person Fit

The MLR approach is one way to assess person fit by comparing the slopes of PRC's. Reise (2000) developed the MLR approach for indices of person fit under a dichotomous 2PL IRT model, but the same approach can also be used for polytomous items (LaHuis & Copeland, in press). The basic approach treats item difficulty as a Level 1 predictor and person trait levels as a Level 2 predictor, which is represented by the following set of equations:

$$P_{ij}\ (Y=1|\ X_{ij}) = \frac{\exp(b_{0j} + b_{1j}\beta_{ij})}{1+\exp(b_{0j} + b_{1j}\beta_{ij})} \tag{3}$$

$$b_{0j} = \gamma_{00} + \gamma_{01}\ (\theta_j) + u_{0j} \tag{4}$$

$$b_{1j} = \gamma_{10} + u_{1j} \tag{5}$$

In Equations 3-5: b_{0j} represents the intercept for person j, and b_{1j} represents person j's slope coefficient. These equations produce PRC's like those in Figure 2. More negative b_{1j}'s indicate better fit. The γ's represent Level 2 coefficients, and the u's are the unique deviations for person j for the intercept and slope. The b_{0j} intercept coefficient represents the expected probability when all predictors are zero. The b_{1j} slope coefficient represents the shared prediction of scores. β_{ij} is the level of difficulty for item i, and θ_j is the trait level of person j.

Using this framework, MLR may be used to identify systematic variance in PRC's by testing for significant variance in the slopes for item difficulty. LaHuis and Copeland (in press) suggest that individual differences in faking may be a cause of systematic differences in PRC slopes. For example, a faker might endorse many of the toughest items, but overlook other items that are not as difficult. This would result in poor person fit to the 2PL model.

One of the benefits of MLR is the ability to test *a priori* hypotheses concerning how item and person characteristics relate to faking, such as variance related to item features and job level. MLR treats item-level data as nested within the individuals. In this way, item features are specified as Level 1 variables, while individual level variables are specified as Level 2 variables.

Polytomous Data

The graded-response model (GRM; Samejima, 1969; 1996) is an extension of the 2PL model for data that is designed for polytomous responses. Most personality tests are scaled on more than two possible answers. The advantage for designing tests with polytomous items is that responses have greater range and flexibility of scoring, thus helping to diminish error and increase precision (Spector, 1992).

Samejima's GRM assumes an item has *m* ordered categories, in order to allow the estimation of item parameters at each category. IRT estimation is done for ($k = m - 1$) categories – which define the boundary response functions (BRF's). BRF's are the probability of endorsing upper response options versus the probability of endorsing lower response options. For example, the BRF (P^*_{i3}) is the probability of choosing response option 3 or above. The BRF (P^*_{i4}) is the chance of choosing response option 4 or above. On a 6-choice response item, the BRF (P^*_{i6}) would involve the probability of choosing the highest option. The equation for a BRF is:

$$P^*_{ik}\ (Y=1|\theta_j) = \frac{\exp[\alpha_i(\theta_j - \beta_{ik})]}{1+\exp[\alpha_i(\theta_j - \beta_{ik})]} \qquad (6)$$

This equation substitutes the probability of choosing between categories in the BRF, instead of the probability of choosing between responses in the dichotomous model (see Equation 1). The BRF equation has parameters equivalent to the standard 2PL model

– except for β_{ik}, the threshold parameter, which represents item difficulty for response option *k*. For a 6-choice response item, there are 5 possible β_{ik} parameters. The item discrimination value (α_i) is held constant. Figure 3 shows an example of consecutive BRF's for a 6-choice response item.

BRF's can be used to calculate the probability of endorsing a single response option. See the following equations for calculating the probability of selecting each individual response option:

$$P_{i1}(\theta) = 1 - P^{*}_{i2}(\theta) \quad (7)$$

$$P_{i2}(\theta) = P^{*}_{i2}(\theta) - P^{*}_{i3}(\theta) \quad (8)$$

$$P_{i3}(\theta) = P^{*}_{i3}(\theta) - P^{*}_{i4}(\theta) \quad (9)$$

$$P_{i4}(\theta) = P^{*}_{i4}(\theta) - P^{*}_{i5}(\theta) \quad (10)$$

$$P_{i5}(\theta) = P^{*}_{i5}(\theta) - P^{*}_{i6}(\theta) \quad (11)$$

$$P_{i6}(\theta) = P^{*}_{i6}(\theta) - 0 \quad (12)$$

These equations show that the probabilities of selecting a particular option (P_{ik}), given theta, are calculated cumulatively by subtracting one BRF from the previous BRF. It is possible to conduct a MLR analysis on only the BRF's that are theoretically useful. As shown in Equation 3, the GRM can be thought of as a series of dichotomous 2PL models. For this study of faking, it is likely that applicant faking is most prevalent for the probability of choosing the highest response option versus lower options (BRF = P^{*}_{i6}).

In the following sections, I develop several hypotheses about how item and person characteristics relate to faking. Specifically, at the item level, I propose that item

transparency, job-relevance, and verifiability may influence faking. At the person level, I suggest that applicants exhibit higher levels of faking than incumbents.

Item-Level Characteristics

Item Verifiability

Verifiable items are likely to discourage faking on questionnaires, by asking for responses where the examinee could be held accountable for lying. Mael (1991) states that, "A verifiable item is an item that can be corroborated from an independent source" (p. 777). For example, applicants may be less likely to fake on biographical questionnaire items if they are verifiable from other records of employment (Asher, 1972). A verifiable test item might ask for something that is commensurable with employment records, such as punctuality.

Compared to most forms of selection, personality inventories are often less verifiable in nature. This is because personality items tap into subjective internal states, behavioral intents, and hypothetical responses (Asher, 1972; Hough et al., 1990; Mael, 1991). However, people do tend to respond in ways that are consistent with the impression they think that others have of them, albeit positive or negative (Schlenker, 1980, Schlenker et al., 2008). Fakers may have a high degree of discomfort with items that are even somewhat verifiable, so they would be motivated to give an honest response. Cognitive dissonance theory asserts that people in uncertain circumstances are motivated to give responses that are aligned with the perceived impressions of others (Festinger, 1957). Thus, personality items should be less fakable if they are related to outward behaviors that could be observed by others, rather than internal states of disposition (Mael, 1991).

Some research has focused on item verifiability and the prevalence of faking. Donovan et al. (2003) found that the applicants' perceptions of verifiability was correlated negatively (r = -.67) with their prevalence of faking on noncognitive measures, though this effect was overshadowed by perceived severity of the deception. In another study of item features, verifiable items exhibited less susceptibility to distortion than nonverifiable items (Mael, 1991). Also, Becker & Colquitt (1992) found verifiability was related to less distortion as compared to other item features on a biodata form. More research is needed in this area. Based on this, I predict the verifiability of the item will be negatively related to faking.

Hypothesis 1: Easily verifiable items will permit less faking than those that are more difficult to verify, after controlling for item difficulty and trait level. Thus, verifiability will be negatively related to the probability of endorsement.

Item Transparency

With transparent items, applicants can easily guess the response that would produce a higher score on the personality construct (Alliger, Lilienfeld, & Mitchell, 1995). This characteristic makes transparent items amenable to faking. Non-transparent items tend to be abstruse or idiosyncratic, while still tapping into the target construct (Jackson, 1971). However, non-transparent items also tend to be less valid or theoretically based (Duff, 1965; Wiener, 1948). This helps explain why transparent items are more prevalent in personality inventories (e.g., Abrahams et al., 1971).

Boyle & Start (1989) lamented that self-report tests typically include a large number of transparent items, which are clearly relevant to the construct, and therefore easy to fake. Non-transparent items might decrease the ability to fake, because there may be no obvious response that would provide a "correct" score on the construct being

measured (McFarland & Ryan, 2000). Much of the research on transparency has focused on the utility of "subtle scales" that have been developed in order to curb faking (e.g., Hough & Paulin, 1994; Barge & Hough, 1986; Owens, 1976). This approach to test construction has had mixed success, as many of these scales have been found vulnerable to faking (e.g., Meehl & Hathaway, 1946; Schrader & Osburn, 1977; Thornton & Gierasch, 1980). While it may be difficult to ascertain the construct underlying non-transparent items, it is often still possible to guess the correct direction to make a desired response (Snell et al., 1999). Unfortunately, very few of these studies have focused at the item level.

Although Zickar & Drasgow (1996) proposed that an item transparency would be an important feature of fakable tests, this has not been examined directly. In the present study, I tested relationships between item transparency and faking using the MLR approach. I expect that item transparency will adversely affect the honesty of respondents.

Hypothesis 2: Item transparency will be positively related to faking, after controlling for item difficulty and trait level. Thus, transparency will be positively related to the probability of endorsement.

Item Job-Relevance

Another item-level predictor of faking may be item job relatedness (Leary & Kowalski, 1990). Kroger & Turnbull (1975) found that individuals faked more successfully on personality tests when made aware of the accurate job description. Similarly, Kluger et al. (1991) discovered that graduate students scored marginally higher on biodata if provided with a specific job title, than if asked more generally to simulate

"applying for a job." Personality inventories may be most successful for those applicants who are able to target which items are specifically job related.

Without adequate job knowledge, applicants may actually fake in the wrong direction on some jobs, because their conceptions might include negative or inaccurate stereotypes (Mahar et al., 1995). In one meta-analysis, applicants for a sales job appeared to actually fake in the wrong direction on a scale of agreeableness (Birkeland et al., 2006). Other jobs, such as junior manager, may be particularly easy for applicants to guess the ideal job-relevant characteristics (Martin et al., 2002).

The choice of which items to fake is primarily a matter of personal judgment (Furnham, 1990). Kreitler & Kreitler (1981) found that a sample of Israeli military had extreme responses most related to judgments of relevance for items of a personality scale. Also, this study found that test takers were likely to respond neutrally for items of questionable relevance. I suggest job applicants may tend to give neutral responses and fake less on items that are not job-relevant, while they tend to give more extreme responses to questions judged highly relevant.

Hypothesis 3: Item job-relevance will be positively related to faking controlling for item difficulty and trait level. Thus, job-relevance will be positively related to the probability of endorsement.

Individual-Level Characteristics

At the individual level, applicants tend to score higher than incumbents do on personality scales, with no observed differences between groups other than the possibility of faking (Birkeland et al., 2006). For instance, Schmit & Ryan (1993) compared the factor structure of the Big Five for applicant and incumbent job groups. The applicant group had a different response pattern than the incumbent group for four of the five

NEO-FFI subscales. That is, applicants uniformly endorsed a category of items regardless of their underlying trait level. Schmit & Ryan suggested that this category of responses was best described by an "ideal employee factor" – which resembles faking.

Robie et al. (2001) conducted a large study comparing applicant sales managers and incumbent sales managers. In support of the faking hypothesis, they found that applicants scored approximately 0.5 standard deviations higher than incumbents on personality scales measuring the Big Five. These results were compared with the laboratory study by Zickar and Robie (1999) who found similar differences in test scores between faking and honest groups and explained that, "…some items were more easily faked than others" (p. 559). Robie et al. (2001) were unable to pinpoint the same item-related causes of faking.

The use of a between-subjects design has been criticized somewhat in the faking literature (e.g., Ellingson et al., 2006; Hogan et al., 2007). Zickar et al. (2004) found substantial variation in distortion for both applicants and incumbents, which compromised the ability to compare groups directly. Incumbents may engage in self-presentation bias in much the same way as applicants. Therefore, a broad comparison in test scores between applicants and incumbents may overlook actual patterns of distortion that occurs in both groups. This is why it is important to treat faking as a continuous variable and to investigate applicant-incumbent differences in conjunction with interactions at the test level (Ellingson et al., 2006).

Differences in mean test scores between applicants and incumbents have been most recently demonstrated by Birkland et al.'s meta-analysis (2006). The results of this meta-analysis were similar to results of another meta-analysis (Viswesvaran & Ones,

1999) which compared groups of “fake good” or “honest” respondents. According to Birkeland et al., “The degree to which [applicants] distort their scores, however, is (a) less than the degree that they distort scores when instructed to fake and (b) depends on the personality dimension being measured, the type of job, and the type of test” (p. 325). Several studies have found either measurement invariance or a similar factor structure between applicants and incumbents on tests of personality (e.g., Griffin et al., 2004; Robie et al., 2001; Tsaousis & Nikolaou, 2001). However, research agrees that applicants tend to score higher than incumbents on impression management scales (e.g., Dunnette et al., 1962; Rosse et al., 1998). In general, Birkeland et al. (2006) found that job applicants scored higher than incumbents on the Big Five traits of Conscientiousness (d = 0.45) and Emotional Stability (d = 0.43).

Thus, there is some indication that applicants fake more than incumbents do. However, this effect has not been examined using the MLR approach. That is, it not clear if there are applicant and incumbent differences in PRC’s. Based on previous research, I would expect that applicants will be more likely to have PRC’s that are consistent with faking. That is, I suggest that the slopes of the PRC’s will be less negative than those for incumbents.

Hypothesis 4: The slopes of the PRC’s will be less negative for applicants than those for incumbents.

Method

Participants

Archival data was collected by a personality testing center. Participants were 507 applicants for and 302 incumbents in customer call service positions. Demographic data were unavailable.

Twelve graduate students volunteered to rate items of the Work Style 5 in terms of verifiability, transparency, and job-relevance. Students rated each item on a series of scales, using a written key for each scale (Appendix A). Their answers were averaged to provide ratings for verifiability, transparency, and job-relevance of each item. A similar method was used by Robie (2006) to identify item subtlety. See Table 1 for the reliabilities, means, and standard deviations of item ratings from this study.

Measures

Personality. The assessment used in archival data was the Work Style 5. The Work Style 5 is adapted from the NEO-IPIP, and it has 17 items per scale. Data was collected for each factor: Conscientiousness , Extraversion, Emotional Stability, Agreeableness, and Openness. Like all of the Work Style 5 scales, Conscientiousness was highly reliable with a alpha coefficient of 0.87. Next, for Extraversion there was an alpha coefficient of 0.83. The Emotional Stability alpha was 0.88. For Agreeableness, the alpha reliability was 0.80. Finally, Openness had a reliability coefficient of 0.85. Examples of the items and item ratings are presented in Appendix B.

Lie Scale. The Work Style 5 included a 7-item embedded subscale with Unlikely Virtues (UV) which was used for some exploratory analyses. The alpha reliability of the lie scale was 0.61.

Item Ratings. Graduate students had 255 ratings across 85 items of the Work Style 5, because there were 3 ratings for 17 items each construct. This includes 85 ratings each for item reliability, verifiability, and job-relevance. The item reliabilities for ratings across the Big Five were the following: $\alpha = 0.88$ for verifiability, $\alpha = 0.72$ for transparency, and $\alpha = 0.87$ for job-relevance. Reliabilities were also computed across raters. Inter-rater reliability is reported in Table 2 for each rating and construct.

Analyses

IRT Estimation. The GRM item and trait level parameters were obtained using Multilog 7.03 (Thissen, 2003). The fit of the model was evaluated using adjusted chi square to degrees of freedom ratio (Chernyshenko et al., 2001). I report these statistics for item singles, pairs, and triples. Ratios below three (adjusted $\chi^2/df < 3.00$) indicate acceptable fit for the 2PL model to the data.

MLR Analyses. The first step of the MLR analyses is to specify an MLR equation with the IRT estimates of item difficulty (Level 1) and person trait level (Level 2), while allowing intercepts to vary. I tested for variance in the intercepts using the recommended chi-squared statistics (see Raudenbush & Bryk, 2002). The intercepts should be explained completely by trait level estimates so residual variation should be close to zero and nonsignificant. With nonsignificant intercept variance after controlling for trait level estimates, the intercept can be specified as fixed. However, if τ_{00} is significantly different from zero, this would indicate differential test functioning (DTF). This means that the

entire test discriminates on some factor(s) other than trait level. In the unlikely case of DTF, the intercepts should be allowed to vary for the third and final steps.

The second step evaluates whether or not the slopes of the PRC's vary across individuals. In this step, the γ_{10} term is the grand mean of person slopes, which is allowed to vary for individuals' deviations (u_{1j}) in response patterns. If the chi-square test for slope variance is significant, there are systematic differences in individuals' PRC slopes.

The third and fourth steps add predictors to explain variance in the intercepts and/or slopes. These steps tested my hypotheses about faking. Item transparency, job-relevance, and verifiability will be added as Level 1 predictors. For example, it may be that transparent items would encourage more faking than subtle items. A positive correlation for item transparency would indicate less negative slopes for transparent items than subtle items. Controlling for item difficulty and trait-level estimates, individuals would be more likely to endorse items that are transparent. Finally, applicant/incumbent status will be added at Level 2 as a predictor of negative slopes.

Results

IRT Estimation – Model Fit

I investigated fit of the IRT model and eliminated some items where I could identify a pattern of misfit. This was necessary where the chi-square to degrees of freedom statistics were above three ($\chi^2/df > 3.00$) for the scales of Agreeableness, Emotional Stability, and Conscientiousness. At least two scales, Openness and Extraversion, were allowed to include all of the items with acceptable fit ($\chi^2/df < 3.00$).

For the Agreeableness scale, the mean adjusted chi-square to degrees of freedom was below three for item singles (adjusted $\chi^2/df = 0.00$). It was above four for doubles (adjusted $\chi^2/df = 4.22$) and triples (adjusted $\chi^2/df = 4.31$). For one item removed, the mean adjusted chi-square to degrees of freedom ratio was below three for item singles (adjusted $\chi^2/df = 0.00$). However, it was above three for doubles (adjusted $\chi^2/df = 4.08$) and triples ($\chi^2/df = 3.66$). This suggested that the GRM could not reach the accepted standard of fit for Agreeableness. Thus, Agreeableness was excluded from MLR analysis.

For the Emotional Stability scale, the mean adjusted chi-square to degrees of freedom was below three for item singles (adjusted $\chi^2/df = 0.00$). It was above four for doubles (adjusted $\chi^2/df = 4.28$) and triples (adjusted $\chi^2/df = 4.66$). For two items removed, the mean adjusted chi-square to degrees of freedom ratio was below three for item singles (adjusted $\chi^2/df = 0.00$). It was above three for doubles (adjusted $\chi^2/df = 3.30$) and triples ($\chi^2/df = 4.09$). Again, this suggested that the GRM exhibited misfit for the items. Thus, the Emotional Stability scale was excluded from MLR analysis.

For the Conscientiousness scale, the mean adjusted chi-square to degrees of freedom was below three for item singles (adjusted χ^2/df = 0.00). It was above three for doubles (adjusted χ^2/df = 3.56) and triples (adjusted χ^2/df = 4.32). For one item removed, the mean adjusted chi-square to degrees of freedom ratio was below three for item singles (adjusted χ^2/df = 0.00), item doublets (adjusted χ^2/df = 1.98) and triples (χ^2/df = 2.04). This suggested acceptable fit after removing one item for the Conscientiousness scale.

For the Extraversion scale, the mean adjusted chi-square to degrees of freedom was below three for item singles (adjusted χ^2/df = 0.00), doubles (adjusted χ^2/df = 2.54) and triples (adjusted χ^2/df = 2.45). This suggested the GRM fit acceptably well for the scale of Extraversion.

For the Openness scale, the mean adjusted chi-square to degrees of freedom was below three for item singles (adjusted χ^2/df = 0.00) and doubles (adjusted χ^2/df = 2.42). It was near three for item triples (adjusted χ^2/df = 3.42). This suggested reasonable fit for the scale of Extraversion.

Conscientiousness

Table 3 presents the results for the Conscientiousness scale.

Step 1. As expected for Conscientiousness, trait levels were positively related (γ = 2.70, t (807) = 52.89, $p < .01$) and thresholds were negatively related (γ = -1.68, t (12,941) = -27.42, $p < .01$) to the probability of endorsing the topmost response option (see Table 3). This indicated that the probability of endorsement decreased along with trait level, and there was a lower probability of endorsement with increased item thresholds (i.e., difficulty). The chi-square test showed nonsignificant intercept variance (χ^2 (807) = 780.85, $p > .50$) for Conscientiousness, which suggested there was no

systematic variance to be found across persons (i.e., differential test functioning), after controlling for trait levels and item thresholds.

Step 2. The next step tested for slope variance using chi-square values. The intercept was fixed but the slope for item difficulty varied. Results indicated significant slope variance (χ^2 (807) = 1,084.62, $p < .01$), indicating systematic variance in PRC's.

Step 3. Step 3 tested hypotheses for item-level predictors. I added item verifiability, transparency, and job relatedness as Level 1 predictors. There was a nonsignificant but positive relationship between item verifiability and the probability of item endorsement ($\gamma = 0.08$, t (808) = 1.13, $p = 0.26$), after controlling for the item thresholds and trait levels. Similarly, a nonsignificant positive relationship between the probability of item endorsement and item transparency was found ($\gamma = 0.15$, t (808) = 0.96, $p = 0.34$), after controlling for item threshold and trait level. Finally, A nonsignificant relationship was found between job-relevance and the probability of item endorsement ($\gamma = 0.02$, t (808) = 0.21, $p = 0.84$). Although two of the relationships were in the expected direction for item transparency and job-relevance with endorsement, the positive relationship between item verifiability and endorsement was somewhat unexpected.

Step 4. In step 4, after controlling for item thresholds, trait levels, and item-level predictors, applicant/incumbent group was entered as a Level 2 predictor. Incumbents were coded as 0 and applicants were coded as 1. Contrary to expectations, results showed that applicant status was negatively related to slopes for Conscientiousness ($\gamma = -0.57$, t (808) = -4.20, $p < .01$). See Figure 4 for the graph of PRC's between applicants and incumbents on Conscientiousness. This graph shows that there is a more negative curve

for applicants than incumbents, which can contrasted with Figure 2. Applicants were less likely than incumbents to endorse items that exceeded their estimated trait levels.

Extraversion

Results for the Extraversion scale are presented in Table 4.

Step 1. Trait levels were positively related ($\gamma = 2.33$, t (807) = 33.941, $p < .01$) and thresholds were negatively related ($\gamma = -1.06$, t (12,941) = -27.81, $p < .01$) to the probability of endorsing the highest response option. However, the chi-square test for the intercept variance component had systematic variance (χ^2 (807) = 1,172.47, $p < .01$) that was not accounted for by either trait level or item thresholds. This indicated that there was differential test functioning for Extraversion, so the intercepts were allowed to vary in subsequent steps.

Step 2. In this step, both the intercept and the slope for item difficulty varied. Results indicated significant intercept variance (χ^2 (807) = 1364.43, $p < .01$) and slope variance (χ^2 (808) = 1229.82, $p > .01$).

Step 3. In step 3, as expected, item verifiability was related significantly to the probability of endorsement for Extraversion ($\gamma = -0.24$, t (808) = -3.06, $p < .01$). Individuals were less likely to endorse items that were more verifiable. In addition, item transparency was positively related to the probability of endorsement ($\gamma = 0.21$, t (808) = 2.91, $p < .01$). For Extraversion, individuals were more likely to endorse items that were transparent. Finally, it was surprising there was a significant negative relationship between job-relevance and the probability of endorsement for Extraversion ($\gamma = -0.25$, t (808) = -3.68, $p < .01$). The expected relationship would have been positive, because individuals should have been more likely to endorse the items that are most job-relevant.

Step 4. There was a nonsignificant difference in item difficulty slopes between applicants and incumbents for Extraversion (γ = -0.08, t (808) = -1.05, p = .30). See Figure 5 for the corresponding graph of PRC's. Because of the variance found in intercepts, incumbents have a significantly different intercept than applicants. This means that incumbents had a scoring advantage over applicants regardless of estimated trait levels.

Openness

Table 5 presents the results for the Openness scale.

Step 1. As expected for Openness, trait levels were positively related (γ =2.61, t (807) = 51.513, p < .01) and thresholds were negatively related (γ = -0.84, t (12,941) = -21.57, p < .01) to the probability of endorsing the highest response option (see Table 6). The chi-square test revealed nonsignificant variance in the intercepts (χ^2 (807) = 772.78, p > .50). This indicated the absence of differential test functioning.

Step 2. In Step 2, I fixed the intercept, but allowed the item difficulty slope to vary. Results suggested significant variance in the slope (χ^2 (807) = 1,190.71, p < .01).

Step 3. As expected, item verifiability was negatively related (γ = -0.27, t (808) = -3.15, p < .01), and item transparency was positively related (γ = 0.33, t (808) = 4.70, p < .01) to the probability of item endorsement. For Openness, individuals were less likely to endorse items that had higher verifiability, and individuals were more likely to endorse items of high transparency. Again, unexpectedly there was a significant negative relationship between job-relevance and the probability of item endorsement (γ = -0.19, t (808) = -3.58, p < .01). Thus, individuals were less likely to endorse items that were more job-relevant for the scale of Openness.

Step 4. Unexpectedly, applicants had significantly more negative item difficulty slopes that incumbents (γ = -0.50, t (808) = -4.89, p < .01). See Figure 6 for the corresponding graph of PRC's. This graph shows that incumbents have a more positive slopes than applicants for Openness. The graph is most divergent for items of low difficulty.

Hypotheses Summary

Hypotheses 1 stated that respondents should be less likely to fake on items that are more verifiable, because there is the threat to be caught in a lie. Thus, there ought to be a negative relationship between verifiability and item endorsement. For Conscientiousness, I found a nonsignificant relationship (γ = 0.08, t (808) = 1.13, p = 0.26) after controlling for the item thresholds and trait levels. As expected, however, item verifiability was related significantly to the probability of endorsement for Extraversion (γ = -0.23, t (808) = -3.06, p < .01) and Openness (γ = -0.27, t (808) = -3.15, p < .01). Therefore, Hypothesis 1 was partially supported.

Hypothesis 2 stated that more transparent items should lead to more faking. This hypothesis would be supported by a positive relationship between item transparency and the probability of endorsing an item. A nonsignificant positive relationship between slope and item transparency was found for Conscientiousness (γ = 0.15, t (808) = 0.96, p = 0.34) after controlling for item threshold and trait level. As was expected, item transparency was positively related to slope for Extraversion (γ = 0.21, t (808) = 2.91, p < .01) and Openness (γ = 0.33, t (808) = 4.70, p < .01), after controlling for item threshold and trait level. This indicated that individuals were more likely overall to endorse items

of high transparency than low transparency, after controlling for item thresholds (difficulty) and trait levels. Thus, Hypothesis 2 was partially supported.

Hypothesis 3 stated that participants should fake more on behaviors that are seen as desirable to potential employers. This would be supported if job-relevance is positively related to the probability of item endorsement. A nonsignificant relationship was found between job-relevance and slope for Conscientiousness ($\gamma = 0.02$, t (808) = 0.21, p = 0.84). Unexpectedly, there was a significant negative relationship between job-relevant items and endorsement probability for Extraversion ($\gamma = -0.25$, t (808) = -3.68, $p < .01$) and Openness ($\gamma = -0.19$, t (808) = -3.58, $p < .01$). This showed that individuals were less likely to endorse items of high job-relevance than low job-relevance, after controlling for item thresholds (difficulty) and trait levels. Thus, I found no support for Hypothesis 3.

Finally, I believed that applicants should fake more than incumbents as reflected by Hypothesis 4. Support for Hypothesis 4 would be found if applicants had less negative slopes than incumbents. Contrary to expectations, results showed that applicants had more negative related slopes for Conscientiousness ($\gamma = -0.57$, t (808) = -4.20, $p < .01$) and Openness ($\gamma = -0.50$, t (808) = -4.89, $p < .01$), after controlling for item threshold, trait levels, item verifiability, item transparency, and item job-relevance. Thus, I found the opposite of expected results for Hypothesis 4.

Additional Analyses

I conducted an Independent Samples *t*-test between applicant and incumbent groups for theta levels. Table 6 shows a *t*-test comparison of means between incumbents and applicants. Applicants scored significantly higher than incumbents on Conscientiousness (t (807) = -9.93, $p < .001$), Extraversion (t (541.84) = -4.78, $p < .001$),

Openness (t (807) = -7.18, $p < .001$), Emotional Stability (t (807) = -10.78, $p < .001$), and Agreeableness (t (807) = -4.72, $p < .001$). Additionally, the applicant group scored significantly higher on a lie scale score than incumbents (t (807) = -8.92, $p < .001$). This comparison was in support of the overall expectations of applicant-inflated scores.

I also tested for a relationship between lie scale scores and the slopes for item difficulty. I believed that lie scale scores would be positively related to item-difficulty slopes, so that increased lie scale scores would lead to increasingly poorer fit. The results are summarized in Tables 7-9. Lie scale scores were unrelated to slopes for Conscientiousness ($\gamma = 0.003$, t (807) = 0.54, $p = .58$) or Extraversion ($\gamma = -0.002$, t (807) = -0.75, $p = .46$), after controlling for item threshold, trait level, item verifiability, item transparency, and item job-relevance (see Table 7 and 8). Lie scale scores were negatively related to slopes for Openness ($\gamma = -0.02$, t (807) = -3.38, $p < .01$), after controlling for item threshold, trait level, and item-level predictors (see Table 9). Contrary to expectations, lie scale scores were related to increased person fit for one of three constructs.

Finally, I checked for cross-level interactions between group-level predictors and those item-level predictors that had significant variance components in Step 3. For Conscientiousness, there was significant variance in the slopes for item verifiability: ($\tau^2 = 0.53$, $p < .01$), item transparency ($\tau^2 = 0.86$, $p < .01$) and job relatedness ($\tau^2 = 0.69$, $p < .01$). For Extraversion, I did not find significant variance in slopes for item verifiability ($\tau^2 = 0.25$, $p > .05$), item transparency ($\tau^2 = 0.11$, $p > .05$), or item job-relevance ($\tau^2 = 0.27$, $p > .05$). Openness had no significant variance components for item verifiability (τ^2

= 0.74, $p > .05$), item transparency ($\tau^2 = 0.29$, $p = .42$), but there was significant variance in slopes for item job-relevance ($\tau^2 = 0.30$, $p = 0.02$).

As mentioned earlier in unexpected results, Conscientiousness had a small positive relationship between item verifiability and the probability of item endorsement. In the cross-level analysis, applicants had significantly more positive relationships between item verifiability and the probability of item endorsement for Conscientiousness ($\gamma = 0.27$, t (807) = 2.83, $p < .01$), after controlling for item thresholds and trait levels (see Figure 7). This graph shows that applicants were actually more likely to endorse items of high verifiability than low verifiability, whereas incumbents had almost equal endorsement across items of low and high verifiability. Lie scale scores did not significantly affect the relationship between item verifiability and the probability of item endorsement ($\gamma = 0.00$, t (807) = 0.07, $p = .943$) (see Figure 8), after controlling for item thresholds and trait levels (see Table 7). Figure 8 shows an almost equal slope between the upper 75th percentile and the lower 25th percentile of lie scores. Only group membership (see Figure 7) helped to explain the cross-level relationship between item verifiability and slopes for Conscientiousness, albeit in an unexpected direction that showed increased likelihood of endorsement for items high on verifiability.

For item transparency there was a small positive relationship for Conscientiousness with the probability of item endorsement. Results of the cross-level interaction suggested that this was smaller for applicants ($\gamma = -0.43$, t (807) = -3.08, $p < .01$) (see Figure 9). Figure 9 shows that incumbents were significantly more likely to endorse items of high transparency than low transparency, whereas the endorsement of applicants remained stable across items of different transparency. Similarly, the upper

range of lie scale scores was associated with decreases in the size of the relationship between item transparency and slopes ($\gamma = -0.02$, t (807) = -2.52, $p < 0.05$) (see Figure 10). This graph shows that people who passed the lie test were more likely to endorse items of high transparency, whereas there was not as much of a relationship between job-relevance and endorsement for those who failed the lie test.

Incumbents had more of a negative relationships between item job-relevance and the probability of endorsement ($\gamma = 0.28$, t (807) = 3.45, $p < 0.01$) (see Figure 11), after controlling for item threshold and trait level. Figure 11 shows that applicants were more likely to endorse items of high job-relevance than low job-relevance, whereas incumbents were slightly in the opposite direction. Similarly, higher Lie scale scores were associated with more positive relationships between item job-relevance and slopes for Conscientiousness ($\gamma = 0.02$, t (807) = 4.33, $p < .001$) (see Figure 12).

For Openness, incumbents also had more of a negative relationship with item job-relevance and the probability of endorsement ($\gamma = 0.06$, t (807) = 5.98, $p < 0.01$) (see Figure 13), after controlling for item threshold and trait level. Figure 13 shows a negative relationship for incumbents, but almost no effect of job-relevance on endorsement for applicants. Similarly, low Lie scale scores influenced the relationship between item job-relevance and endorsement ($\gamma = 0.004$, t (807) = 9.17, $p < 0.001$) (see Figure 14), such that people who passed the lie test were less likely to endorse job-relevant questions and more likely to endorse irrelevant items, whereas people who failed the lie test had slightly opposite relationship with job-relevance and item endorsement.

Discussion

MLR IRT analysis has many applications for studying the responses of examinees. In this study, I was able to examine simultaneously the effects of both item and person level characteristics on faking, which presented a unique advantage for using MLR analysis by studying both levels related to faking. Rarely have both levels been used for a study of test behavior. I found some support for hypotheses and also some unexpected results that could lead to new directions in research.

Hypotheses

I found some support for Hypothesis 1. Both Extraversion and Openness had evidence for faking with items that were less verifiable. Items rated with high verifiability ("I am the first to act," or, "I can handle a lot of information") seemed able to curb much of the faking with these scales. However, these results were not significant for the Conscientiousness scale.

For Hypothesis 2, I believed that examinees would be more likely to fake on items that were transparent. The results for Conscientiousness were again nonsignificant. However, Extraversion and Openness showed that examinees were more likely to fake on items that were transparent. Transparent items (e.g., "I let myself be pushed around," or, "I excel in what I do") led to the most faking.

Hypothesis 3 predicted that item job-relevance would motivate examinees to fake. The opposite results were found – that examinees were more likely to fake on items that were less job-relevant for Extraversion and Openness. Though the results were not

significant for Conscientiousness, they were in the same unexpected direction. Overall, this might be because job-irrelevant items carry less of a moral quandary, given that the perceived importance of the question is low. That is, it would be easier to rationalize faking on questions that perhaps "unfairly" or "unimportant" test for personal characteristics outside of the job realm. This may help explain the findings that individuals were more likely to inflate scores for job-irrelevant items than for job-relevant items.

Hypothesis 4 predicted that incumbents would have better person fit than incumbents. I found unexpectedly that applicants had significantly better person fit than incumbents for both Conscientiousness and Openness scales. A potential explanation for this is there may differences in response processes used. For example, the GRM used in the present study assumed a dominance response process where the probability of item endorsement relates monotonically to individuals' trait levels. With the dominance response model, individuals' probability of endorsing the item, "I try to follow rules," increases as their conscientiousness increases.

However, other response processes than the dominance model have been theorized. Chernyshenko, Stark, Drasgow, and Roberts (2007) and Stark, Chernyshenko, Drasgow, and Williams (2006) have recently suggested that ideal point models should be considered for personality measures. These models suggest that individuals judge how well an item describes them in terms of the underlying trait and tend to endorse items that they feel match their level of the trait. They will tend to endorse items that they feel match their trait levels. The mismatch may be because they believe their trait level is less than or exceeds that indicated by the item. The former is termed disagreeing from below

and the latter is labeled disagreeing from above. For example, individuals may not endorse the "I try to follow rules" item because they hardly ever try to follow rules or because they always follow rules. This type of disagreement causes a bell-shaped IRF. The decrease in the probability of endorsement associated with disagreeing from above is referred to as folding.

It may be that applicants use a dominance response process, and incumbents use an ideal point response process because of the differences in testing situations. If this were the case, the 2PLM would fit better for applicants than incumbents because applicants' response would be consistent with the model, while incumbents would be inconsistent with the monotonic assumptions of the model. That is, PRC's indicate how well the IRT model fits for individuals. My results indicate that the 2PLM fits significantly better for applicants than incumbents.

Additional Analyses

This study conducted additional analyses to check the hypothesis of faking against the finding that applicant PRC's fit better than incumbents to the 2PLM. It was verified that applicants had a significantly higher lie scale score and higher estimated trait levels than incumbents on all Big Five measures (see Table 6). That is, applicants scored approximately 0.4 to 0.6 standard deviations higher than incumbents on estimated trait levels. The magnitude of this difference is supported by prior research on the prevalence of faking for applicants and incumbents (Robie et al., 2001; Zickar & Robie, 1991).

For both Conscientiousness and Openness, the relationship between item job-relevance and endorsement was more negative for incumbents, and it appeared to be somewhat of a positive relationship for applicants (see Figures 11 and 13).The incumbent

curves show lower rates of endorsement as item job-relevance increased. Perhaps this was part of the reason for incumbents to have more positive PRC's than applicants, because incumbents appeared more cautious at endorsing some items than others if questions were high on job-relevance. The slightly positive rates of endorsement for applicants would have been expected due to faking.

For Conscientiousness, the relationship between item verifiability and endorsement was stronger for applicants (see Figure 7). Although incumbents had equal endorsement for items of low and high verifiability, applicants had more positive rates of endorsement for items that were increasingly verifiable. It may be that applicants thought that verifiable items were more appealing to a prospective employer, and thus, they may have weighted the importance of these items in their responses. Item verifiability seemed important to applicant responses, but it is apparent less so for incumbents.

In contrast, incumbents placed greater weight on item transparency. That is, transparent items tended to have positive rates of endorsement for incumbents, which strengthened the relationship between item transparency and endorsement (see Figure 9). Although the applicant curve shows no relationship between transparency and endorsement, the characteristic of item transparency seemed important to the responses of incumbents whose response process might have been influenced by relatedness of items to the construct.

It is interesting to speculate on the overall pattern of the cross-level interactions. It appears that applicants were placing greater weight on item characteristics that could be viewed as favorable by the organization. That is, the more job-relevant and verifiable the items were, then the more likely applicants would endorse them. It is possible that

applicants believed items with these characteristics were the ones on which organizations placed the most weight. In contrast, these item characteristics did not appear to matter for incumbents.

The reverse was true for item transparency which mattered more for incumbents, but not as much for applicants. It may be that applicants did not place importance on item transparency because they believed that the organization did not place a lot of weight on transparent items.

Limitations

One of the possible limitations is that the study used only a single profession from which to draw personality scores. Multiple vocations would add much towards the generalization of results. There are also prior findings that applicants are better able to fake for some professions than others (e.g., Mahar et al., 1995, Martin et al., 2002, Birkeland et al., 2006). Different jobs will have different requirements, so item job-relevance is likely to vary. Data from other jobs would have been useful for additional analyses of item- and group-level relationships with faking.

Furthermore, the present study was limited to a between subjects design for applicants and incumbents. The tests for variance may have been constricted by possible between-group differences that could have been unrelated to the present model of faking. For instance, the group of incumbents could have had personality standings that matched for the job from attraction, selection, and attrition (see Schneider et al., 1995). If incumbents had a limited range for Conscientiousness – due to its relationship to overall job performance and attrition – this would help to explain the lack of significance in the present study using incumbents as a comparison group. A future study can have more

power statistically, for example, by testing and retesting a group of candidates who are hired to the position.

Also, graduate students' ratings of job-relevance may have differed from the examinees' actual impressions of job-relevance. Applicants and incumbents may have different ratings job-relevance, depending on their knowledge of job characteristics. The present study was restricted to graduate student ratings of item job-relevance.

Finally, there was no way to check for demographic variables in the current study, because it was based on data that was already collected from a customer call center. Demographic characteristics may have explained the differential test functioning for Extraversion, or some person level characteristics could have been added to the model for prediction of faking. Characteristics such as age, test experience, and gender would be useful for a study of faking.

Future Research

MLR and the 2PL model may provide insight into test answers and item characteristics, as well as person characteristics. The factors that contribute to faking are relatively unexplored in the domain of personality testing. Future research should investigate different professions and uses the same ratings of item characteristics. The same results should be duplicated for other personality inventories than the Work Style 5. I believe that further research using MLR IRT methods would help provide a better model of faking, and it also would provide support in understanding the vagaries of this complicated technique.

Future research should include additional self-ratings (e.g., job experience, attitudes, or testing experience), which might explore the many possible contributors to

faking. A confidential survey could provide answers to many questions regarding the prevalence and attitudes toward response distortion. For example, a set of questions could ask for conditions where people have lied in their past, conditions where lying is socially acceptable, and reactions toward dishonest behaviors. I believe that questions related to lying are rarely asked in an experimental situation where people are encouraged to be honest about their responses. Furthermore, demographics and attitudes can help to identify the roots of differential test functioning.

An experimental study could go toward confirming the effect of the ideal point response model or dominance model on fit. Controlled conditions can be instructed to endorse questions for the ideal point model instructions: “If the question fits you exactly in your everyday behavior,” and as for the dominance model instructions: “If the question would be scored equal or below your trait level.” Another type of study that could be useful is a statistical simulation that answers some questions in a monotonic fashion and another set of questions in a non-monotonic fashion, and then the results of MLR IRT analyses could be compared. The results of these experiments could be referenced back to the present study of faking.

References

Abrahams, N., Neumann, I., & Githens, W. H. (1971). Faking vocational interests. *Personnel Psychology, 24,* 5-12.

Alliger, G. M., Lilienfeld, S. O., & Mitchell, K. E. (1996). The susceptibility of overt and covert integrity tests to coaching and faking. *Psychological Science, 7*(1), 32-39.

Asher, J. J. (1972). The biographical item: Can it be improved? *Personnel Psychology, 25,* 251-269.

Baker, F. B. (2001). *The basics of item response theory.* U.S.: ERIC Clearinghouse on Assessment and Evaluation.

Barge, B. N., & Hough, L. M. (1986). Utility of biographical assessment: A review and integration of the literature. In L. M. Hough (Ed.), *Utility of temperament, biodata, and interest assessment for predicting job performance: A review of the literature* (ARI Research Note No. 88-02, pp. 91-130). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Barrick, M. R., & Mount, M. K. (1996). Effects of impression management and self-deception on the predictive validity of personality constructs. *Journal of Applied Psychology, 81(3),* 261-272.

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.

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317+335.s

Bliese, P. D. (2002). Multilevel random coefficient modeling in organizational research: Examples using SAS and S-PLUS. In F. Drasgow & N. Schmitt (Eds.), *Modeling in Organizational Research: Measuring and Analyzing Behavior in Organizations* (pp. 401-445). San Francisco, CA: Jossey-Bass, Inc.

Boyle, G. J., & Start, K. B. (1989). Comparison of higher-stratum motivational factors across sexes using the Children's Motivation Analysis Test. *Personality and Individual Differences, 10(4),* 483-487.

Chernyshenko, O. S., Stark, S., Chan, K. Y., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. *Multivariate Behavioral Research, 36,* 523-562.

Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under assumptions of an ideal point response process: Toward increasing the flexibility of personality measures. *Psychological Assessment, 19,* 88-106.

Christiansen, N. D., Goffin, R. D., Johnston, N. G., & Rothstein, M. G. (1994). Correcting the 16PF for faking: Effects on criterion-related validity and individual hiring decisions. *Personnel Psychology, 47,* 847-860.

Donovan, J. J., Dwight, S. A., & Hurtz, G. M. (2003). An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance, 16(1),* 81-106.

Duff. F. (1965). Item subtlety in personality inventory scales. *Journal of Consulting Psychology, 29,* 565-570.

Dunnette, M. D., McCartney, J., Carlson, H. C., & Kirchner, W. K. (1962). A study of faking behavior on a forced-choice self-description checklist. *Personnel Psychology, 15,* 13-24.

Ellingson, J. E., Sackett, P. R., & Connelly, B. S. (2007). Personality assessment across selection and development contexts: Insights into response distortion. *Journal of Applied Psychology, 92(2),* 386-395.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Festinger, L. A. (1957). *A theory of cognitive dissonance.* New York, NY: Stanford University Press.

Furnham, A. (1986). Response bias, social desirability, and dissimulation. *Personality and Individual Differences, 17(3),* 385-400.

Furnham, A. (1990). Faking personality questionnaires: Fabricating different profiles for different purposes. *Current Psychology: Research and Reviews, 9,* 46-55.

Goffin, R. D., & Christiansen, N. D. (2003). Correcting personality tests for faking: A review of popular personality tests and an initial survey of researchers. *International Journal of Selection and Assessment, 11,* 340-344.

Griffin, B., Hesketh, B., & Grayson, D. (2004). Applicants faking good: Evidence of item bias in the NEO PI-R. *Personality and Individual Differences, 36,* 1545-1558.

Griffith, R. L., Yoshita, Y., Peterson, M. H., & Malm, T. (2006, April). Addressing elusive questions: Investigating the faking-performance relationship. In Griffith, R. G., & Yoshita, Y. (Chairs). Deceptively simple: Applicant faking behavior and prediction of job performance. Symposium conducted at the 21st Annual

conference for the Society for Industrial and Organizational Psychology, Dallas, TX.

Heron, A. (1956). The effects of real-life motivation on questionnaire response. *Journal of Applied Psychology, 40(2),* 65-68.

Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92(5),* 1270-1285.

Holden, R. R., & Jackson, D. N. (1981). Subtlety, information, and faking effects in personality assessment. *Journal of Clinical Psychology, 37,* 379-386.

Hough, L. M. (1998). Effects of intentional distortion in personality measurement and evaluation of suggested palliatives. *Human Performance, 11,* 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 disortion on those validities [Monograph]. *Journal of Applied Psychology, 75(5),* 581-595.

Hough, L. M., & Paulin, C. (1994). Construct-oriented scale construction: The rational approach. In G. S. Stokes, M. D. Mumford, & W. A. Owens (Eds.), *The biodata handbook: Theory, research, and biographical information in selection and performance prediction.* Palo Alto, CA: Consulting Psychologists Press.

Jackson, D. N. (1971). The dynamics of structured personality tests. *Psychological Review, 78,* 229-248.

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.

Kreitler, S., & Kreitler, H. (1981). Test item content: Does it matter? *Educational and Psychological Measurement, 41(3),* 635-642.

Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology, 43(1),* 48-55.

LaHuis, D. M., & Copeland, D. (In press). Investigating faking using a multilevel logistic regression approach to measuring person fit.

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin, 107(1),* 34-47.

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

Mahar, D., Cologon, J., & Duck, J. (1995). Response strategies when faking personality questionnaires in a vocational selection setting. *Personality and Individual Differences, 18,* 605-609.

Martin, B. A., Bowen, C. C., & Hunt, S. T. (2002). How effective are people at faking on personality questionnaires? *Personality and Individual Differences, 32,* 247-256.

McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85(5),* 812-821.

Meehl, P. E., & Hathaway, R. (1946). The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, *30*(5), 525-564.

Mumford, M. D., & Owens, W. A. (1987). Methodology review: Principles, procedures, and findings in the application of background data measures. *Applied Psychological Measurement, 11(1),* 1-31.

Owens, W. A. (1976). Background data. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology*, (pp. 609-644). Chicago, IL: Rand McNally.

Pannone, R. D. (1984). Predicting test performance: A content valid approach to screening applicants. *Personnel Psychology, 37*, 507-514.

Peterson, M. H., & Griffith, R. L. (2006). Faking and job performance: A multi-faceted issue. In R. Griffith & M. Peterson (Eds.), *A closer examination of applicant faking behavior*. Greenwich, CT: Information Age Publishing.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications, Inc.

Rees, C. J., & Metcalf, B. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology, 18*, 156-165.

Reise, S. P. (2000). Using multilevel logistic regression to evaluate person-fit in IRT models. *Multivariate Behavioral Research, 35(4)*, 543-568.

Reise, S. P., & Waller, N. G. (1993). Traitedness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology, 65(1)*, 143-151.

Robie, C. (2006). Effects of perceived selection ratio on personality testing. *Social Behavior and Personality, 34(10),* 1233-1244.

Robie, C.; Zickar, M. J.; & Schmit, M. J. (2001). Measurement equivalence between applicant and incumbent groups: An IRT analysis of personality scales. *Human Performance, 14(2)*, 187-207.

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.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph,* No. 17.

Samejima, F. (1996). The graded response model. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory.* New York, NY: Springer.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations*. Monterey, CA: Brooks/Cole Publishing.

Schlenker, B. R., Wowra, S. A., Johnson, R. M., & Miller, M. L. (2008). The impact of imagined audiences on self-appraisals. *Personal Relationships, 15*(2), 247-260.

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.

Schneider, D., Smith, D. B., & Goldstein, H. W. (1995). Attraction-selection-attrition: Toward a person-environment psychology of organizations. In W. B. Walsh, K. H. Craik, & R. H. Price (Eds.), *Person-environment psychology: New directions and perspectives (2nd edition),* pp. 61-85. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Schrader, A. D., & Osburn, G. (1977). Biodata faking: Effects of induced subtlety and position specificity. *Personnel Psychology, 30*(3), 395-404.

Snell, A. F., Sydell, E. J., & Lueke, S. B. (1999). Towards a theory of applicant faking: Integrating studies of deception. *Human Resource Management Review, 9(2),* 219-242.

Spector, P. E. (1992). *Summated rating scale construction* (Sage University Paper series on Quantitative Applications in the Social Sciences, No. 82). Newbury Park, CA: Sage.

Stark, S., Chernyshenko, O. S., Drasgow, F., & Williams, B. A. (2006). Examining assumptions about item responding in personality assessment: Should ideal point methods be considered for scale development and scoring? *Journal of Applied Psychology, 91*(1), 25-39.

Thissen, D. (2003). *Multilog 7.0.* [Computer Software]. Chicago, IL: Scientific Software International.

Thornton, G. C., & Gierasch, P. F. (1980). Fakability of an empirically derived selection instrument. *Journal of Personality Assessment, 44,* 48-51.

Tsaousis, I., & Nikolaou, I. E. (2001). The stability of the five-factor model of personality in personnel selection and assessment in Greece. *International Journal of Selection and Assessment, 9,* 290-301.

Viswesvaran, C., & Ones, S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197-210.

Waller, N. G., & Reise, S. P. (1989). Computerized adaptive personality assessment: An illustration with the absorption scale. *Journal of Personality and Social Psychology, 57(6),* 1051-1058.

Wiener, D. N. (1948). Subtle and obvious keys for the Minnesota Multiphastic Personality Inventory. *Journal of Consulting Psychology, 12,* 164-170.

Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement, 20(1),* 71-87.

Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering Faking Samples in Applicant, Incumbent, and Experimental Data Sets: An Application of Mixed-Model Item Response Theory. *Organizational Research Methods*, *7*(2), 168-190.

Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84(4),* 551-563.

Appendix A

Rated each question of the NEO-IPIP on the following scales:

Verifiability –

1 = unverifiable … the question is subjective and relates to innermost attitudes.

2 = slightly verifiable … the question is borderline subjective and relates to internal attitudes.

3 = somewhat verifiable … the question is borderline objective and relates to external behaviors.

4 = highly verifiable … the question is objective and can be easily corroborated.

Transparency –

1 = very obvious … the question clearly fits in with the other questions of the construct.

2 = somewhat obvious … the question somewhat fits in with other questions of the construct.

3 = somewhat subtle … the question is somewhat different from other questions of the construct.

4 = very subtle … the question does not fit in clearly with other questions of the construct.

Job-Relevance –

1 = not job-related … the question is not related to performance of job duties.

2 = slightly job-related … the question could conceivably be related to slight aspects of job performance.

3 = somewhat job-related … the question is related to performance of supplemental behaviors on the job.

4 = very job-related … the question is highly related to performance of actual job duties.

Appendix B

Sample Item from the Work Style 5:	Mean Item Ratings (min 1, max 4): Verifiability	Transparency	Job-Relevance
"Finish what I start." (Conscientiousness)	3.72	1.36	3.91
"Do a lot in my spare time." (Extraversion)	3.09	2.45	1.64
"Keep my cool." (Emotional Stability)	2.63	1.45	3.09
"Am concerned about others." (Agreeableness)	2.27	1.73	2.27
"Adapt well to new situations." (Openness)	2.82	1.54	3.00

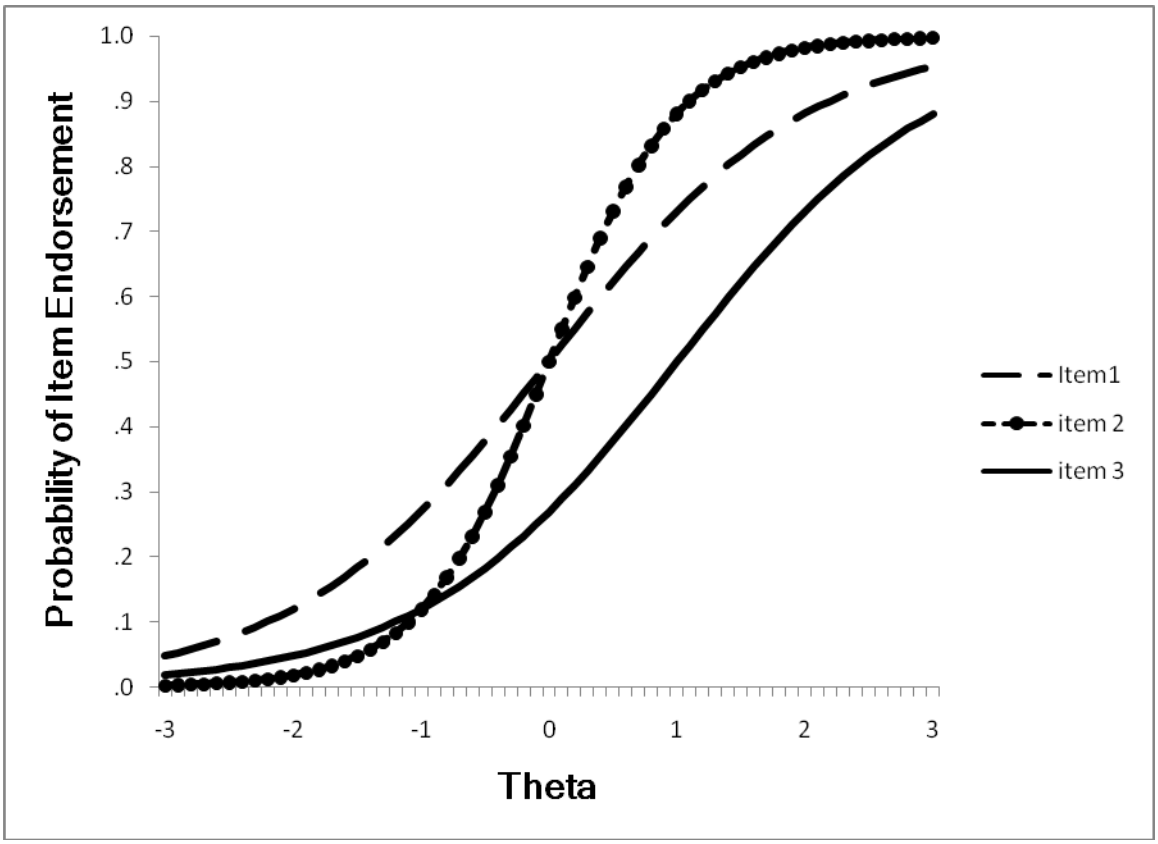

Figure 1. Example of Item Response Curves

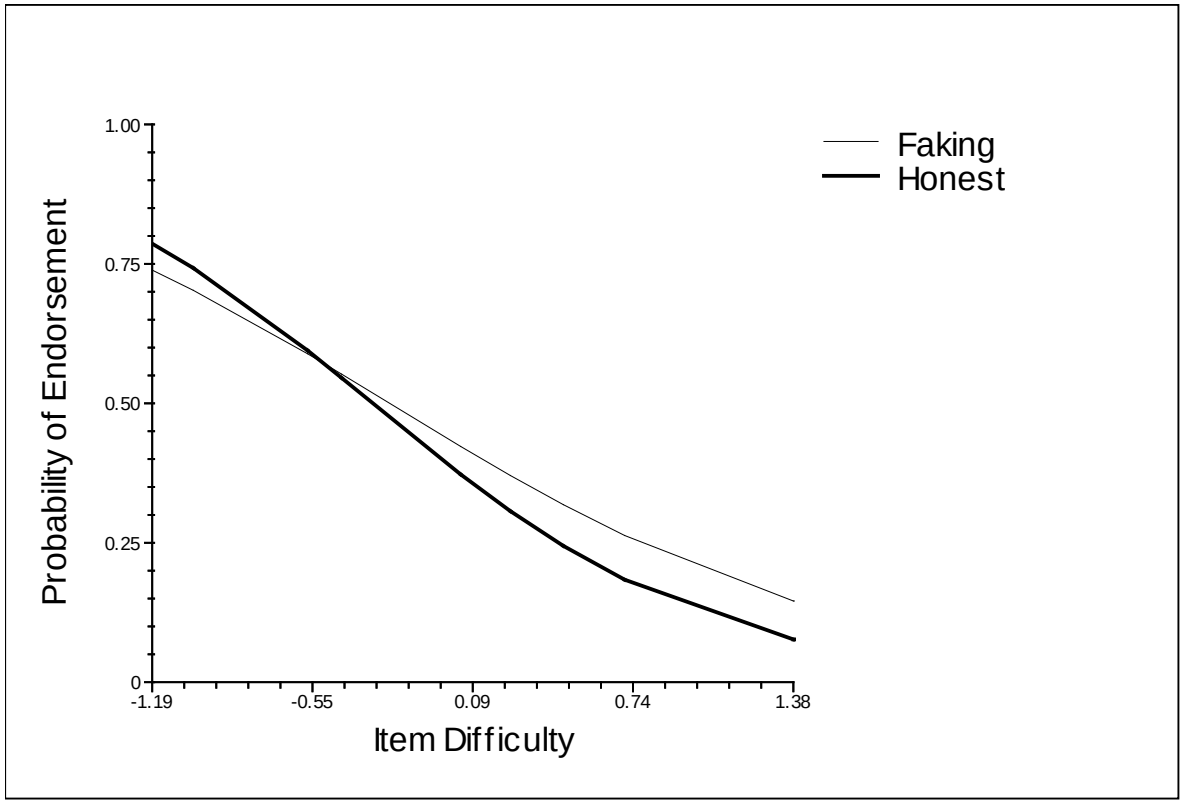

Figure 2. Example of Person Response Curves

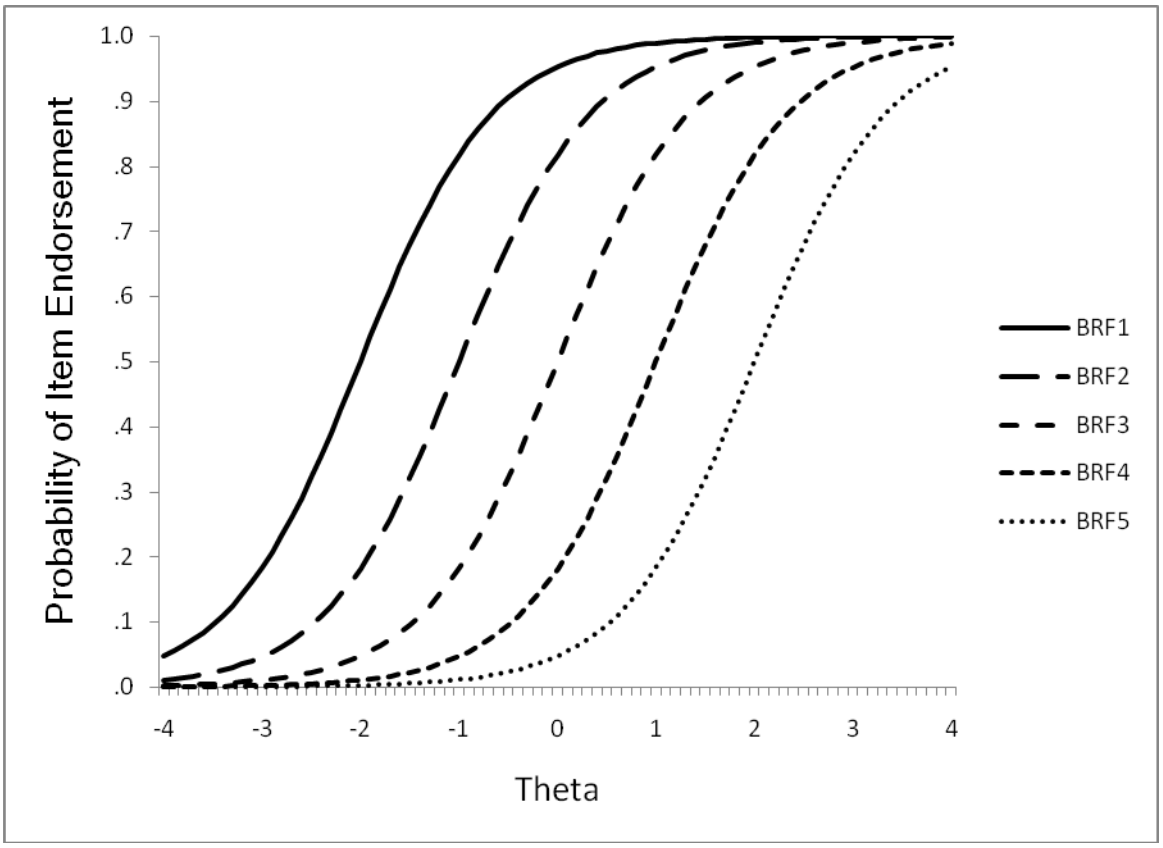

Figure 3. Example of BRF's

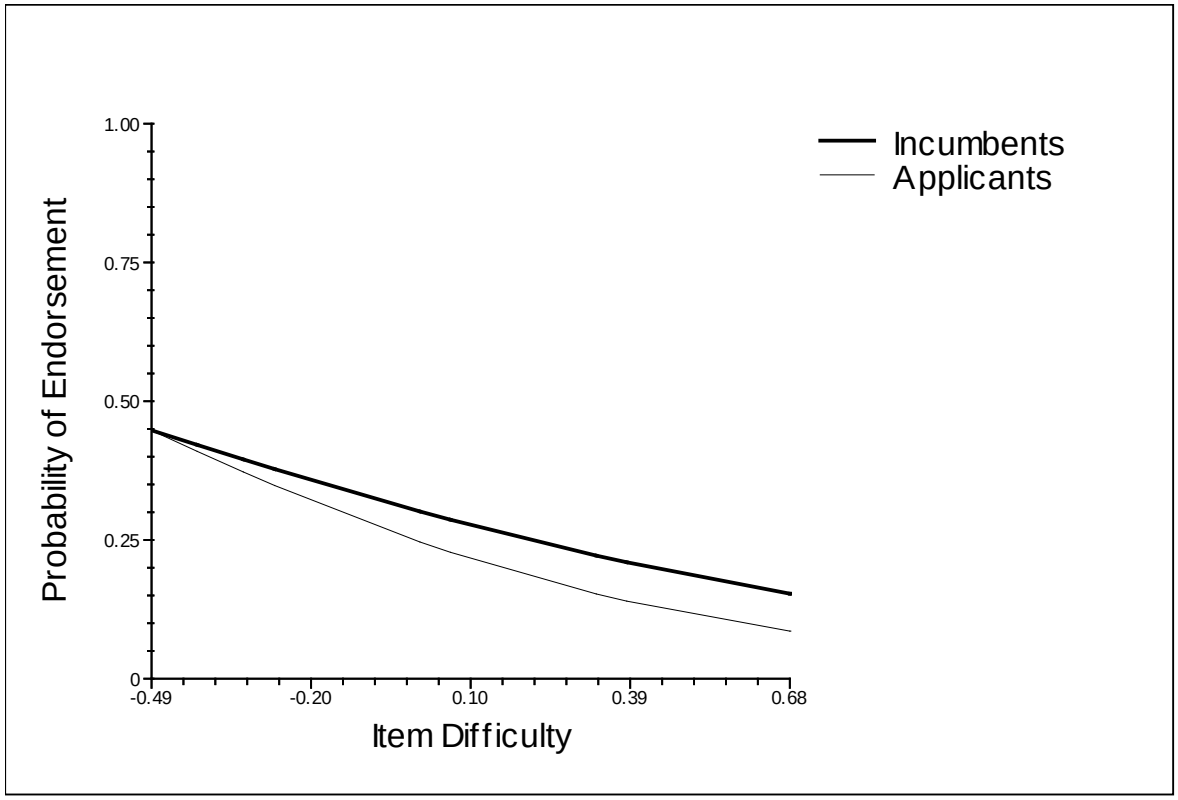

Figure 4. Graph of PRC's for Conscientiousness

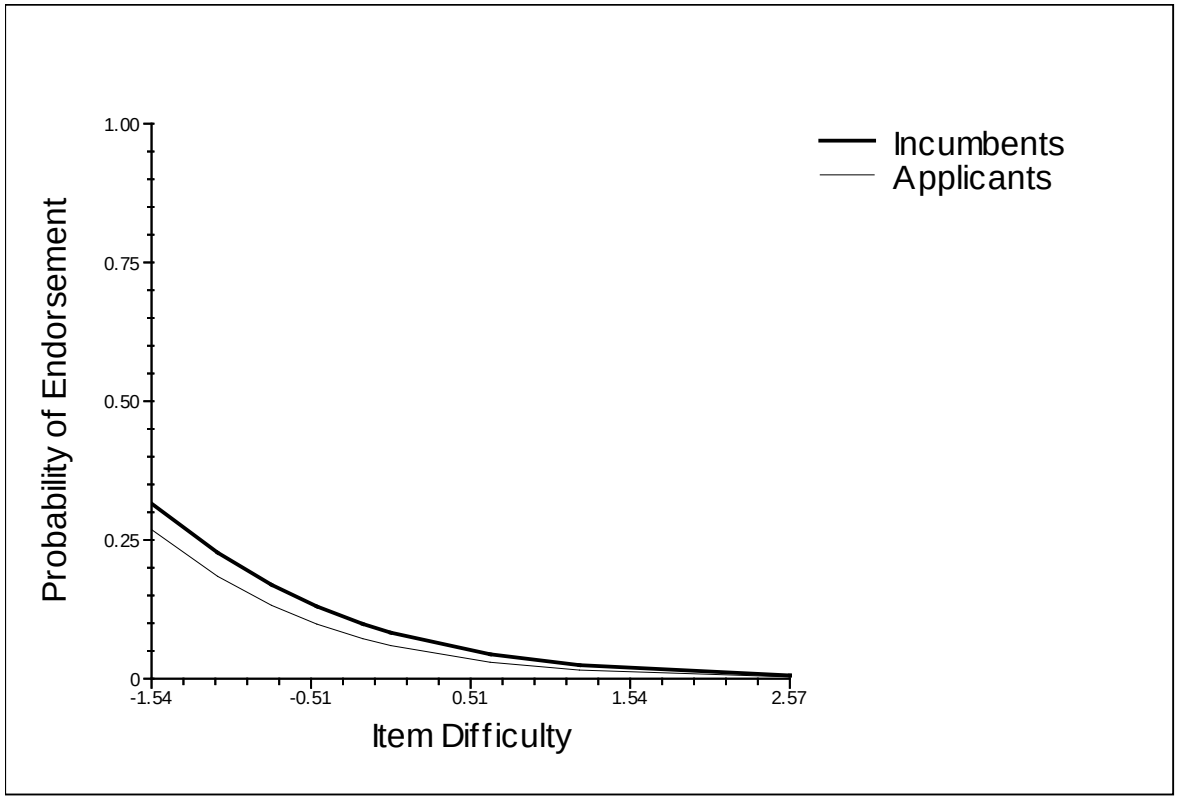

Figure 5. Graph of PRC's for Extraversion

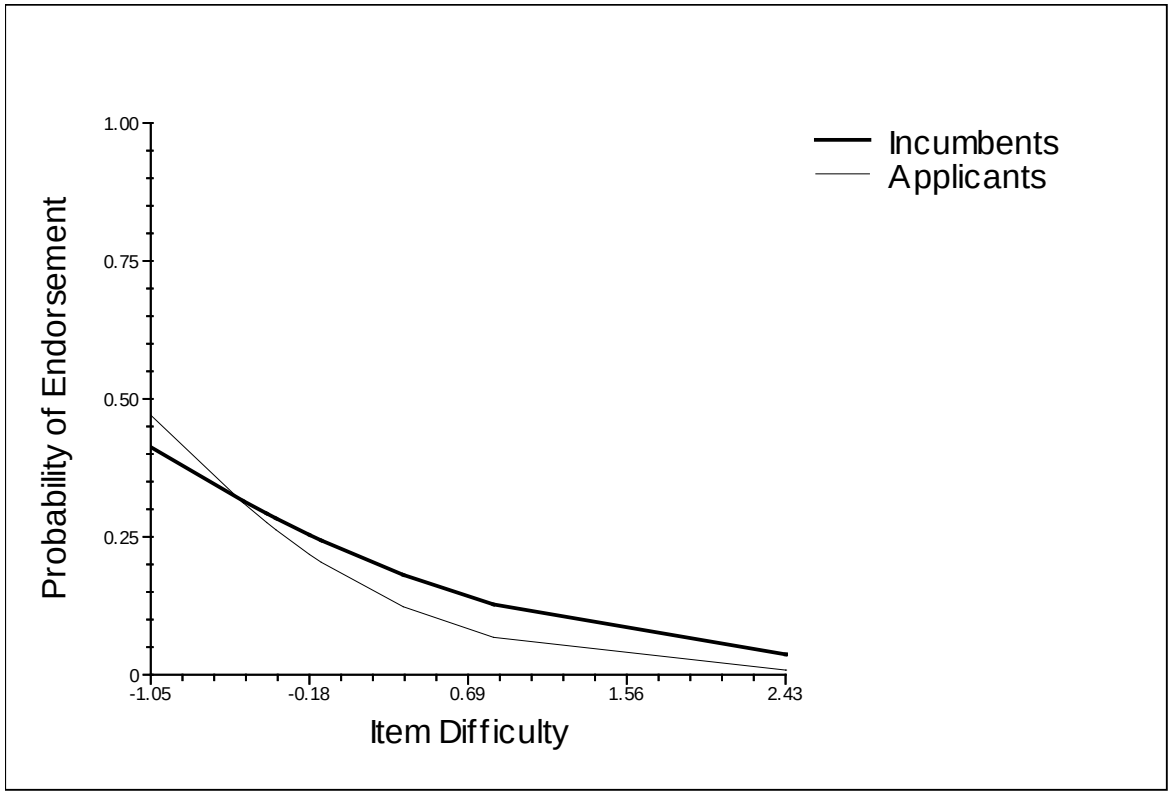

Figure 6. Graph of PRC's for Openness

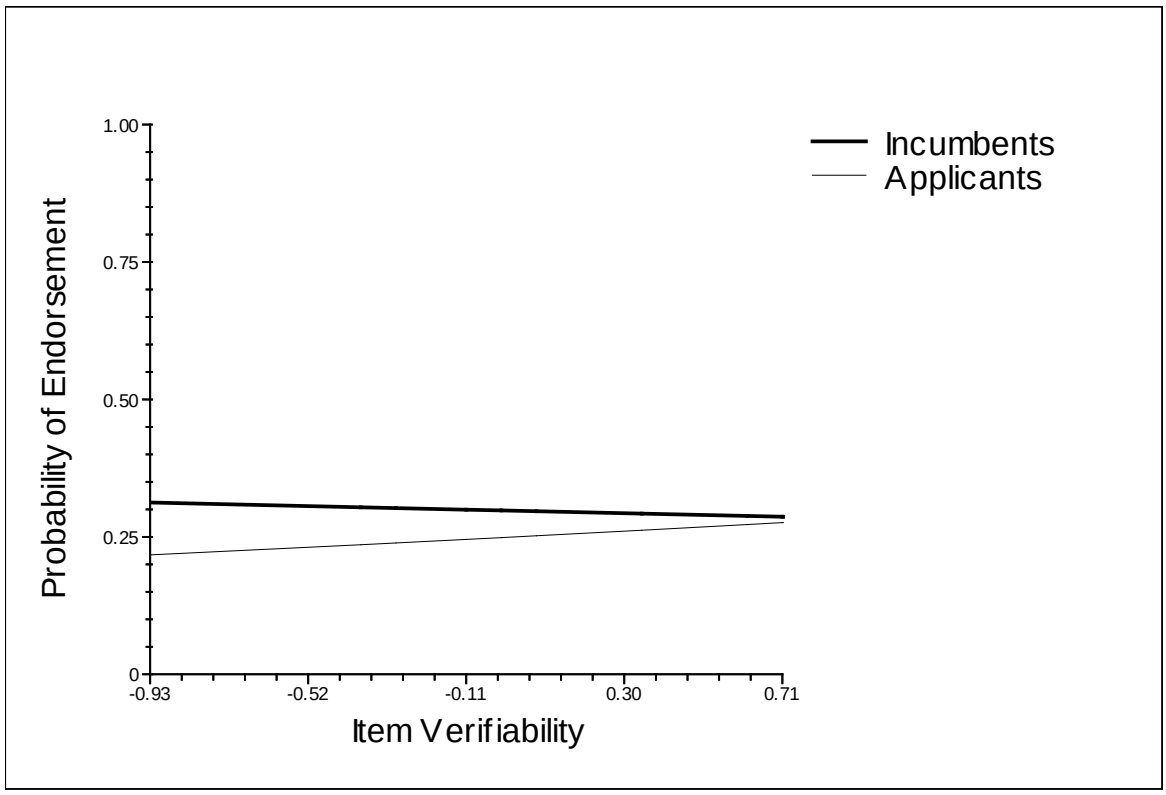

Figure 7. Interaction between group membership and item verifiability for Conscientiousness

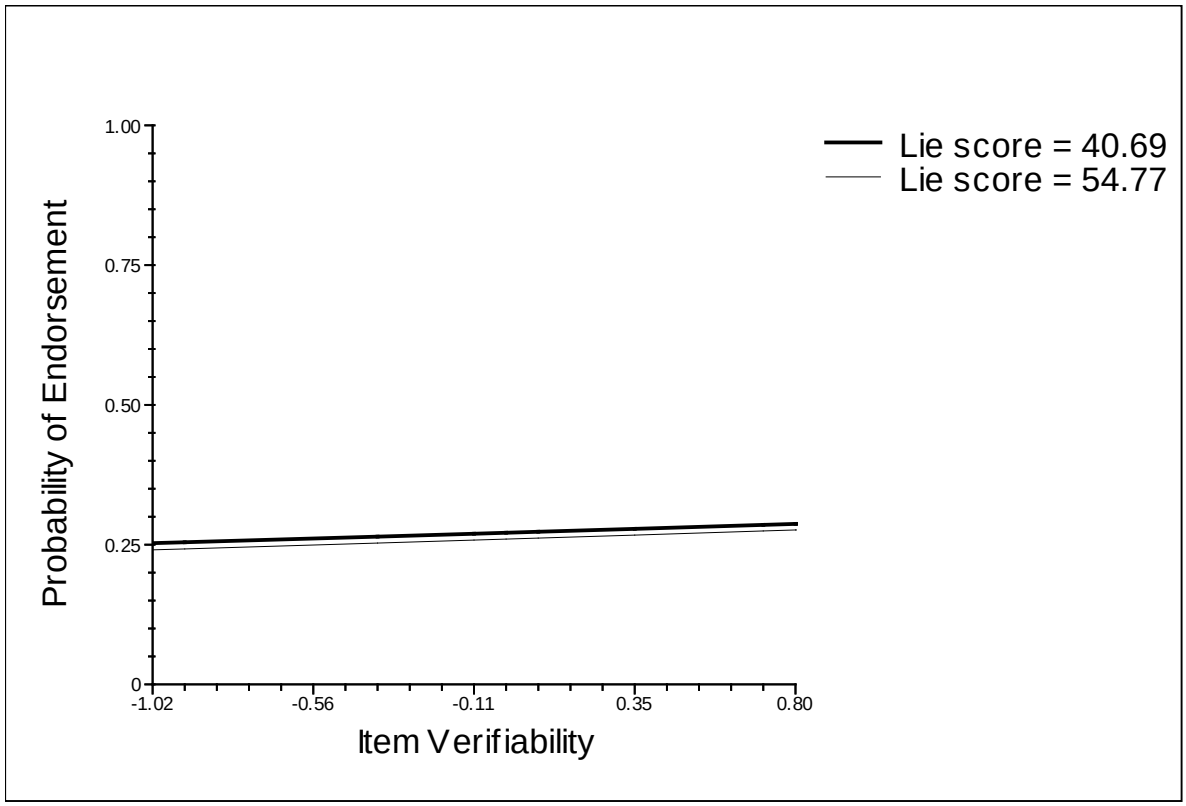

Figure 8. Interaction between lie scale score and item verifiability for Conscientiousness

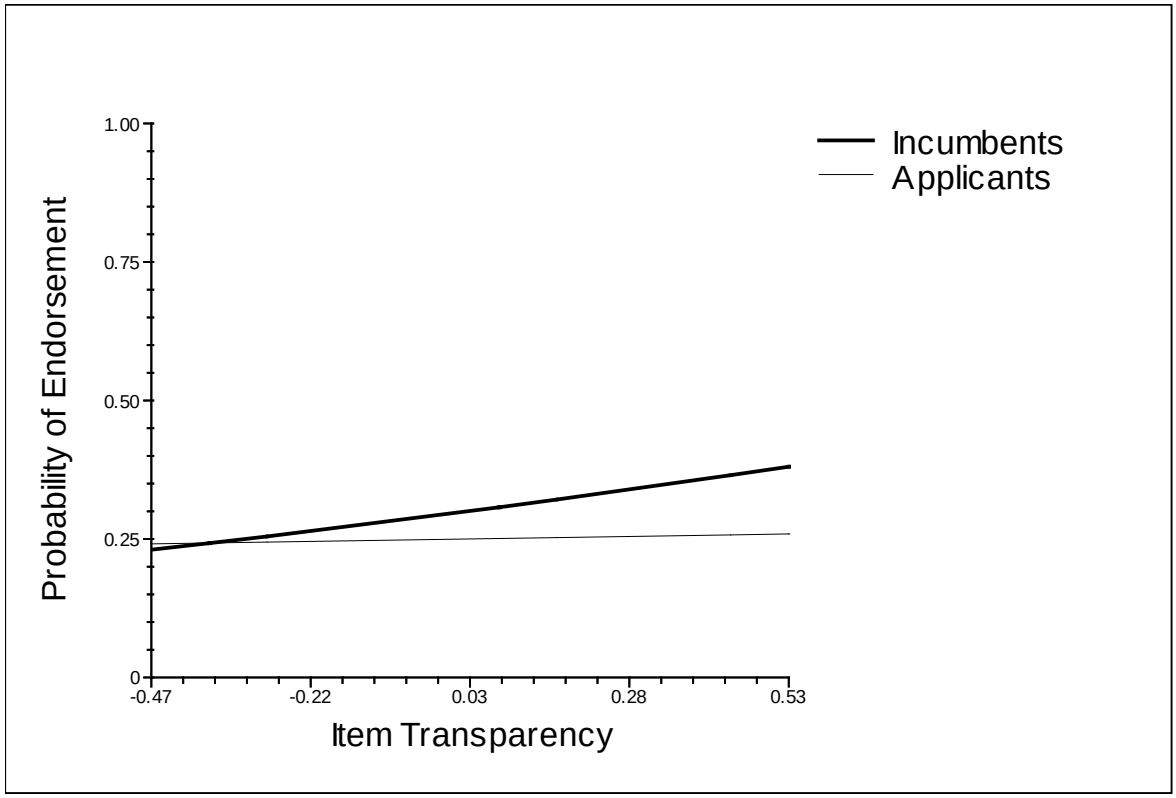

Figure 9. Interaction between group membership and item transparency for Conscientiousness

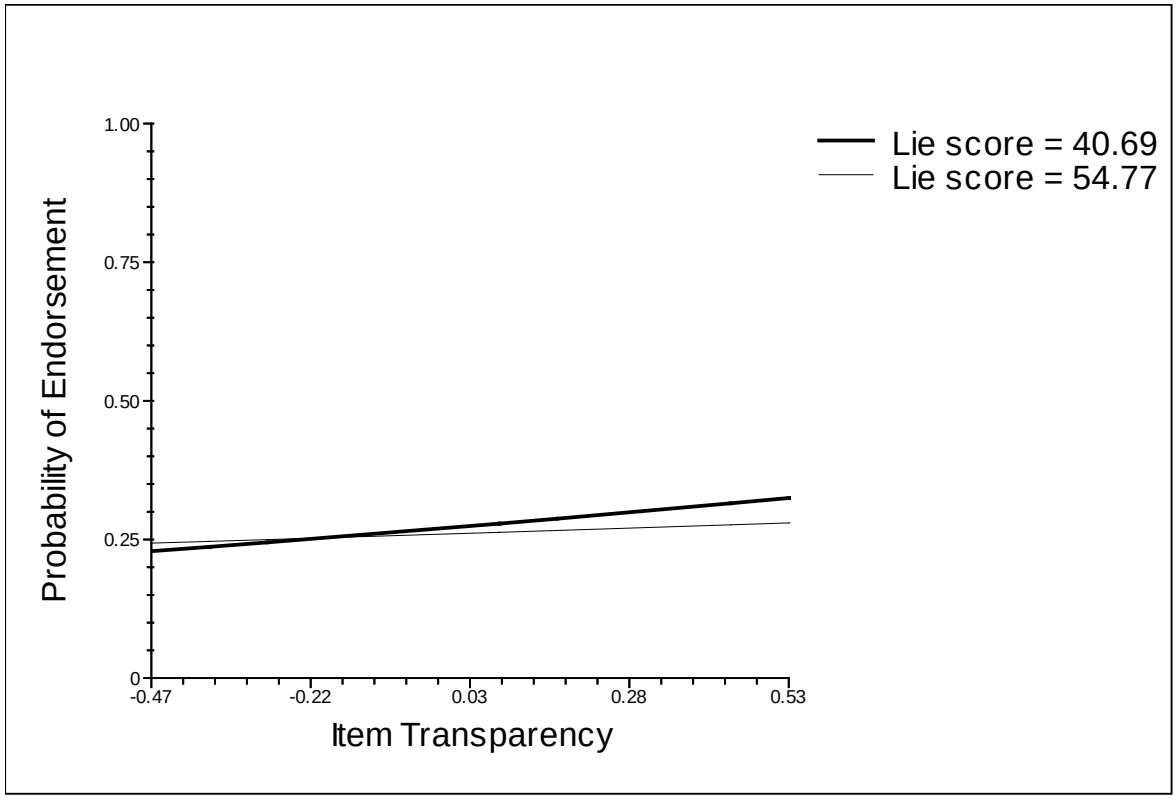

Figure 10. Interaction between lie scale score and item transparency for Conscientiousness

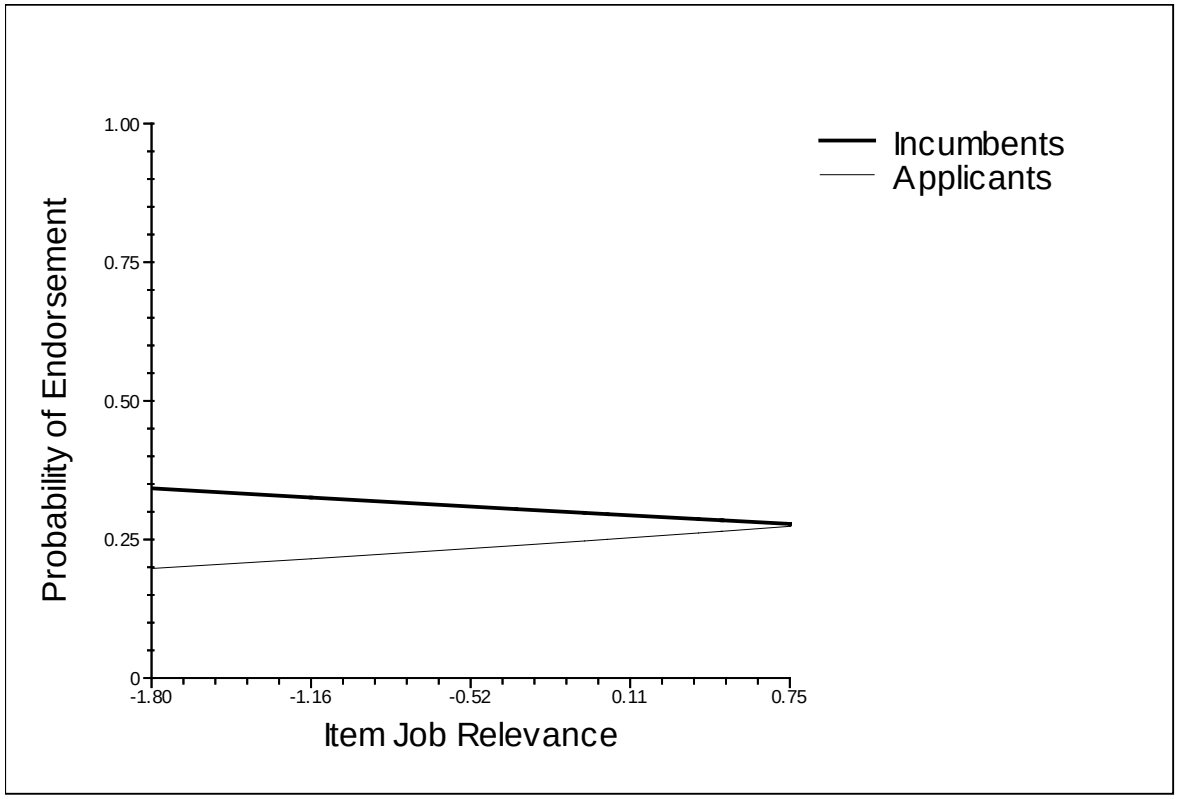

Figure 11. Interaction between group membership and item job-relevance for Conscientiousness

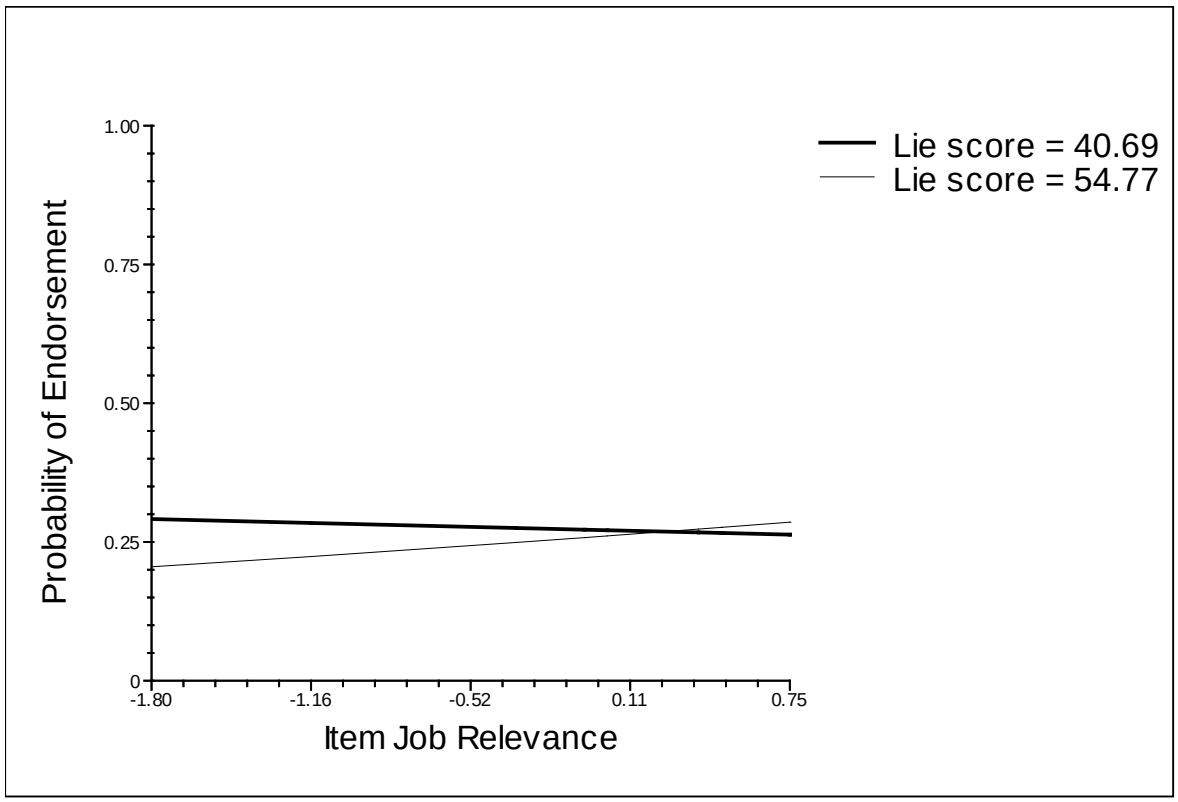

Figure 12. Interaction between lie scale score and item job-relevance for Conscientiousness

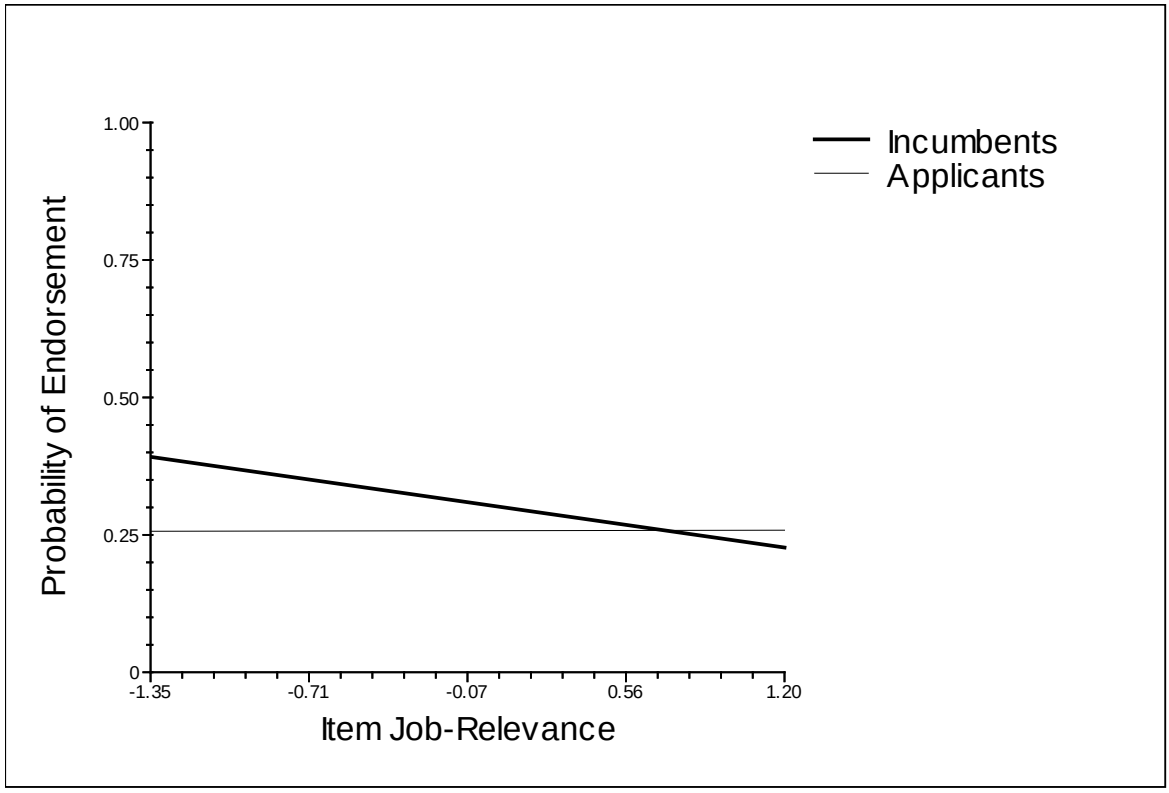

Figure 13. Interaction between group membership and item job-relevance for Openness

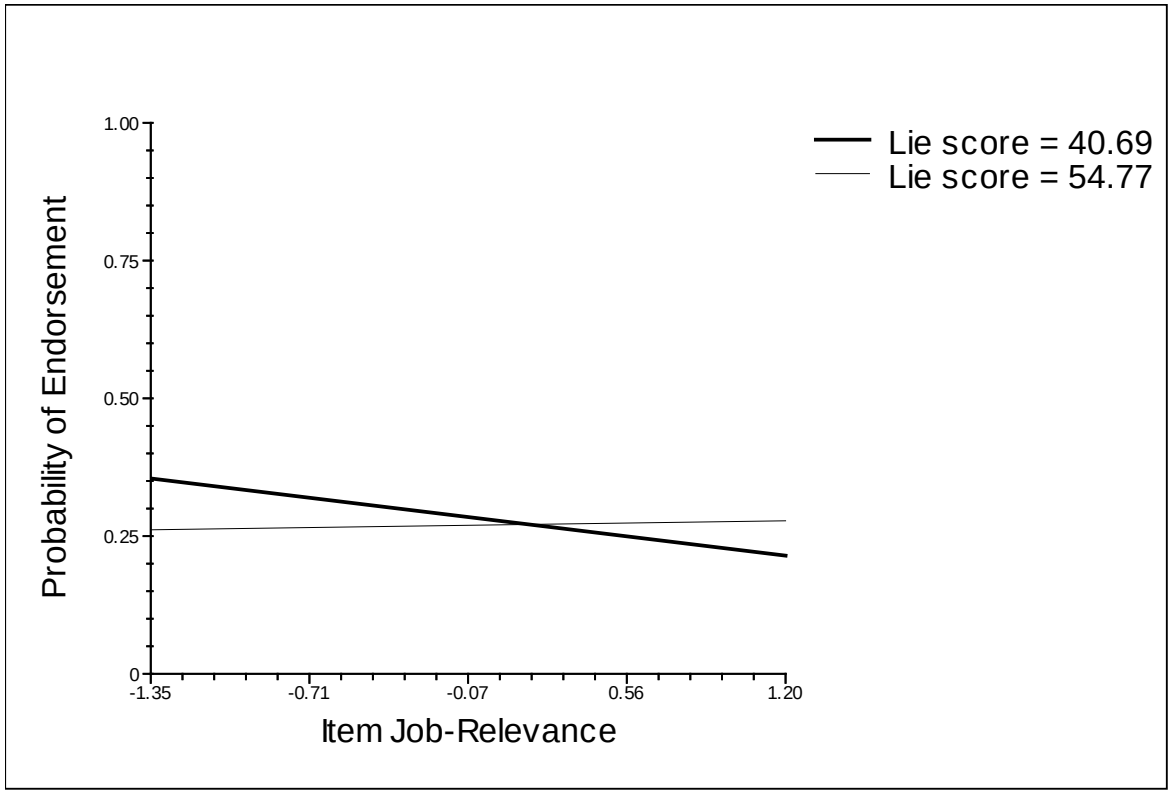

Figure 14. Interaction between lie scale score and item job-relevance for Openness

Table 1.

Descriptive Ratings for Items

Measure	Alpha	M	S.D.	
Item Verifiability:				
Work Style 5	0.88	2.50	0.65	85
Conscientiousness	0.73	2.98	0.51	17
Extraversion	0.82	2.78	0.76	17
Agreeableness	0.87	2.18	0.72	17
Emotional Stability	0.93	2.10	0.67	17
Openness	0.82	2.44	0.60	17
Item Transparency:				
Work Style 5	0.72	1.99	0.75	85
Conscientiousness	0.78	1.78	0.62	17
Extraversion	0.77	1.92	0.72	17
Agreeableness	0.56	2.08	0.76	17
Emotional Stability	0.75	1.98	0.80	17
Openness	0.50	2.16	0.85	17
Item Job-Relevance:				
Work Style 5	0.87	2.63	0.72	85
Conscientiousness	0.59	3.16	0.45	17
Extraversion	0.84	2.35	0.83	17

Agreeableness	0.90	2.51	0.90	17
Emotional Stability	0.83	2.53	0.88	17
Openness	0.81	2.62	0.81	17

Table 2.

Reliabilities, Means, and Standard Deviations for Inter-Rater Agreement (n = 11)

Measure	Alpha	Mean	S.D.	Items
Inter-Rater Agreement:				
Work Style 5	0.86	2.37	0.64	255
Item Verifiability	0.88	2.50	0.63	85
Item Transparency	0.72	1.98	0.47	85
Item Job-Relevance	0.87	2.63	0.66	85
Conscientiousness	0.92	2.64	0.81	51
Extraversion	0.83	2.35	0.62	51
Agreeableness	0.77	2.25	0.52	51
Emotional Stability	0.78	2.20	0.54	51
Openness	0.87	2.41	0.66	51

Table 3.

Results from the MLR analyses for Conscientiousness Response Option 5

	Probability of endorsement							
	Step 1		Step 2		Step 3		Step 4	
	γ	SE	γ	SE	γ	SE	Γ	SE
Item-level								
Threshold	-1.68 *	.06	-1.67*	.07	-1.64*	.08	-1.26*	.12
Verifiability					0.08	.07	0.08	.07
Transparency					0.15	.16	0.15	.16
Job-Relevance					0.02	.08	0.02	.08
Person-level								
Intercept	-4.51*	.08	-4.50*	.08	-4.54*	.08	-4.44*	.08
Trait Level	2.70*	.05	2.69*	.05	2.71*	.05	2.77*	.05
Group							-0.28*	.05
Cross-level interactions								
Difficulty by Group							-0.57*	.14
Verifiability by Group							0.27*	.09
Transparency by Group							-0.43*	.14
Job-Relevance by Group							0.28*	.08
Variance components								
Intercept	.00							
Item difficulty slope			.46*		1.00*		.99*	
Verifiability slope					.53*		.54*	

Transparency slope	.26*	.86*	.92*
Job-Relevance slope		.69*	.72*

Note. * $p < .05$.

Table 4.

Results from the MLR analyses for Extraversion Response Option 5

	Probability of endorsement							
	Step 1		Step 2		Step 3		Step 4	
	γ	SE	γ	SE	Γ	SE	Γ	SE
Item-level								
Threshold	-1.06*	.04	-1.06*	.04	-1.12*	.04	-1.08*	.06
Verifiability					-0.24*	.08	-0.23*	.08
Transparency					0.21*	.07	0.21*	.07
Job-Relevance					-0.25*	.07	-0.25*	.07
Person-level								
Intercept	-3.67*	.07	-3.67*	.07	-3.74*	.06	-3.52*	.08
Trait Level	2.33*	.07	2.34*	.07	2.42*	.07	2.44*	.07
Group							-0.36*	.08
Cross-level interactions								
Difficulty by Group							-0.08	.08
Verifiability by Group							N/A	
Transparency by Group							N/A	
Job-Relevance by Group							N/A	
Variance components								
Intercept	.19*		.12*		.10*		.11*	
Item difficulty slope			.09*		.08*		.07*	
Verifiability slope					.24		.25	

Transparency slope	.26*	.12	.52*
Job-Relevance slope		.27	.29

Note. * $p < .05$.

Table 5.

Results from the MLR analyses for Openness Response Option 5

	Probability of endorsement							
	Step 1		Step 2		Step 3		Step 4	
	γ	SE	γ	SE	Γ	SE	Γ	SE
Item-level								
Threshold	-0.84*	.04	-0.87*	.05	-1.15*	.06	-0.85*	.09
Verifiability					-0.27*	.09	-0.28*	.09
Transparency					0.33*	.07	0.32*	.07
Job-Relevance					-0.19*	.05	-0.20*	.05
Person-level								
Intercept	-4.24*	.07	-4.28*	.07	-4.38*	.07	-4.25*	.08
Trait Level	2.61*	.05	2.63*	.05	2.69*	.05	2.74*	.05
Group							-0.28*	.05
Cross-level interactions								
Difficulty by Group							-0.50*	.10
Verifiability by Group							N/A	
Transparency by Group							N/A	
Job-Relevance by Group							0.06*	.01
Variance components								
Intercept	.00							
Item difficulty slope			.55*		.21		.23	
Verifiability slope					.74		.73	

Transparency slope	.26*	.321	.22*
Job-Relevance slope		.30*	.30

Note. * $p < .05$.

Table 6.

Differences between Incumbent and Applicant Groups on Theta Scores and Lie scale scores

	Incumbents (n=302)		Applicants (n=507)			
Measure	M	SD	M	SD	t	df
Conscientiousness	1.01	.61	1.47	.63	-9.93*	807
Extraversion	0.34	.58	0.53	.48	-4.78*	541.84**
Openness	0.91	.61	1.21	.57	-7.18*	807
Emotional Stability	0.62	.59	1.08	.56	-10.78*	807
Agreeableness	1.38	.58	1.60	.60	-4.72*	807
Lie scale score	43.66	10.89	50.39	10.08	-8.92*	807

*$p < .001$.

**Equal variances *not* assumed for *t*-test.

Table 7.

Results from the lie score MLR analyses for Conscientiousness Response Option 5

	Probability of endorsement							
	Step 1		Step 2		Step 3		Step 4.5	
	γ	SE	γ	SE	Γ	SE	Γ	SE
Item-level								
Threshold	-1.68 *	.06	-1.67*	.07	-1.64*	.08	-1.81*	.32
Verifiability					0.08	.07	0.08	.07
Transparency					0.15	.16	0.15	.16
Job-Relevance					0.02	.08	0.02	.08
Person-level								
Intercept	-4.51*	.08	-4.50*	.08	-4.54*	.08	-4.40*	.12
Trait Level	2.70*	.05	2.69*	.05	2.71*	.05	2.75*	.06
Lie scale score							-0.00	.00
Cross-level interactions								
Difficulty by Lie scale score							0.00	.01
Verifiability by Lie scale score							0.00	.00
Transparency by Lie scale score							-0.02*	.01
Job-Relevance by Lie scale score							0.02*	.00
Variance components								

Intercept	.00			
Item difficulty slope		.46*	1.00*	.98*
Verifiability slope			.53*	.52*
Transparency slope		.26*	.62*	.53*
Job-Relevance slope			.69*	.68*

Note. * $p < .05$.

Table 8.

Results from the lie scoreMLR analyses for Extraversion Response Option 5

	Probability of endorsement							
	Step 1		Step 2		Step 3		Step 4.5	
	γ	SE	γ	SE	γ	SE	Γ	SE
Item-level								
Threshold	-1.06*	.04	-1.06*	.04	-1.12*	.04	-0.98*	.19
Verifiability					-0.24*	.08	-0.25*	.08
Transparency					0.21*	.07	0.21*	.07
Job-Relevance					-0.25*	.07	-0.25*	.07
Person-level								
Intercept	-3.67*	.07	-3.67*	.07	-3.74*	.06	-4.34*	.18
Trait Level	2.33*	.07	2.34*	.07	2.42*	.07	2.31*	.07
Lie scale score							0.01*	.00
Cross-level interactions								
Difficulty by Lie scale score							-0.00	.00
Verifiability by Lie scale score							N/A	
Transparency by Lie scale score							N/A	
Job-Relevance by Lie scale score							N/A	
Variance components								

Intercept	.19*	.12*	.10*	.08*
Item difficulty slope		.09*	.08*	.09*
Verifiability slope			.24	.24
Transparency slope		.26*	.12	.15*
Job-Relevance slope			.27	.28

Note. * $p < .05$.

Table 9.

Results from the lie score MLR analyses for Openness Response Option 5

	Probability of endorsement							
	Step 1		Step 2		Step 3		Step 4.5	
	γ	SE	γ	SE	γ	SE	Γ	SE
Item-level								
Threshold	-0.84*	.04	-0.87*	.05	-1.15*	.06	-0.38	.24
Verifiability					-0.27*	.09	-0.26*	.09
Transparency					0.33*	.07	0.33*	.07
Job-Relevance					-0.19*	.05	-0.19*	.05
Person-level								
Intercept	-4.24*	.07	-4.28*	.07	-4.38*	.07	-4.18*	.12
Trait Level	2.61*	.05	2.63*	.05	2.69*	.05	2.73*	.06
Lie scale score							-0.00	.00
Cross-level interactions								
Difficulty by Lie scale score							-0.02*	.00
Verifiability by Lie scale score							N/A	
Transparency by Lie scale score							N/A	
Job-Relevance by Lie scale score							0.00*	.00
Variance components								

Intercept	.00			
Item difficulty slope		.55*	.21	.25
Verifiability slope			.74	.71
Transparency slope		.26*	.39	.32*
Job-Relevance slope			.30*	.30*

Note. * $p < .05$.