

2010

# The effect of a reasoning warning on faking in personality testing for selection and the perception of procedural justice

T Ryan Dullaghan  
*University of South Florida*

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

---

## Scholar Commons Citation

Dullaghan, T Ryan, "The effect of a reasoning warning on faking in personality testing for selection and the perception of procedural justice" (2010). *Graduate Theses and Dissertations*.  
<http://scholarcommons.usf.edu/etd/1619>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

The Effect of a Reasoning Warning on Faking in  
Personality Testing for Selection and the Perception of Procedural Justice

by

T. Ryan Dullaghan

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Arts  
Department of Psychology  
College of Arts and Sciences  
University of South Florida

Major Professor: Walter Borman, Ph.D.  
Stephen Stark, Ph.D.  
Marcie Finkelstein, Ph.D.

Date of Approval:  
December 12, 2008

Keywords: neuroticism, agreeableness, conscientiousness,  
openness to experience, extraversion, impression management

© Copyright 2010, T. Ryan Dullaghan

## Table of Contents

List of Tables	iii
Abstract	iv
Chapter One – Introduction	1
Faking Behavior	2
The Utility of Warnings	6
Procedural Justice	10
Identifying Faking	13
Self-Presentation/Direct Evidence	13
Mean Comparisons	16
Variance	17
Summary	17
Chapter Two – Method	20
Participants	20
Measures	21
Warning Prompts	21
Personality Test	22
Procedural Justice	23
Impression Management	24
Procedure	24
Chapter Three – Results	26
Comparison of Sources	26
Personality	28
Procedural Justice	36
Impression Management	38
Chapter Four – Discussion	39
Personality	39
Procedural Justice	41
Impression Management	42
Limitations	43
Future Research	44
Conclusion	46
References	47

Appendices	54
Appendix A: International Personality Item Pool Items	55
Appendix B: Procedural Justice Questionnaire Items	57
Appendix C: Selection Procedural Justice Scale Items	58
Appendix D: Impression Management Scale	59

## List of Tables

Table 1	Follow-up ANOVAs to Survey MANOVA: Full sample	27
Table 2	Follow-up ANOVAs to Source MANOVA: Full sample	27
Table 3	Means and standard deviations for Nurse and Student samples for the IPIP	29
Table 4	Means and standard deviations for Nurse and Student samples for procedural justice	30
Table 5	Intercorrelations among study variables: Nurse sample	31
Table 6	Intercorrelations among study variables: Student sample	32
Table 7	Descriptive and ANOVA statistics with Tukey's post hoc tests for the personality variables: Nurse sample	35
Table 8	Descriptive and t-test statistics for the procedural justice variables: Nurse sample	37
Table 9	Descriptive and t-test statistics for the procedural justice variables: Student sample	38

The Effect of a Reasoning Warning on Faking in  
Personality Testing for Selection and the Perception of Procedural Justice

T. Ryan Dullaghan

ABSTRACT

A major concern with using personality tests in the selection process is the prevalence of applicant faking behavior which can influence the rank order of applicants such that fakers are hired at an elevated frequency. This study examined the effects of the detection/consequence warning and a more applicant-friendly warning on faking and perceived procedural justice. I hypothesized that a positive warning (reasoning warning) and a detection/consequence warning would show similar mean personality trait levels compared to honest responses, with all means showing less socially desirable responding than no warning prompt. Results suggested that the detection/consequence warning is more effective at reducing faking behavior in the selection context, and the content of the warning has no impact on perceived procedural justice.

## Chapter One

### Introduction

Utilizing personality tests in the selection process has become increasingly popular over the past decade and a half since the development of the five-factor model of personality and research showing the validity of the five-factor model in predicting job performance, organizational status, income, and a number of other variables (Barrick & Mount, 1991; Mount & Barrick, 1995; Judge, Higgins, Thoreson, & Barrick, 1999; Tett, Jackson, & Rothstein, 1991). In the early 1990s, meta-analyses demonstrated the robustness of the Big Five personality factors of neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness (Barrick & Mount, 1991; Tett et al., 1991). Not only have some of these personality traits been shown to predict job performance, but personality can show high incremental validity above general mental ability (GMA) in the prediction of job performance (Mount, Witt, & Barrick, 2000; Schmidt & Hunter, 1998) compared to other selection procedures. Furthermore, unlike GMAs, personality tests used for selection show little to no bias based on age, race, or gender (Hough, Oswald, & Ployhart, 2001). Unfortunately, personality data can be faked much more easily than other components of a selection battery (Donovan, Dwight, & Hurtz, 2003).

Faking is intentionally distorting responses on a personality test so as to appear higher or lower on a trait than the respondent's true trait score (McFarland & Ryan, 2000). In most cases where respondents are motivated to distort their responses

(selection, assessment, classification), faking so as to appear less desirable may not be a serious issue, so most research on the topic has focused on faking in the positive direction (Jones & Abraham, 2008).

Warning prompts have been shown to be effective methods for reducing applicant faking on personality tests. There is, however, evidence that applicants have a negative perception of personality tests being used in the selection process (Rosse, Miller, & Stecher, 1994), and adding a warning prompt appears to also elicit a negative response from applicants (McFarland, 2003). McFarland did not, however, compare unwarned conditions to the warned conditions.

Little research has examined the target of the applicant's negative perceptions, but the selection system (Rosse et al., 1994) and the organization (McFarland, 2003) have been hypothesized to be potential targets. If currently used warnings do, in fact, harm applicant perceptions of an organizational application process, these applicants could potentially withdraw from the selection process, tell other prospective employees not to apply to the organization, and/or enter the organization with a negative view of the organization's justice. New types of warnings that seek to improve applicant perceptions of procedural justice could potentially mitigate the effects of including a personality test in the selection battery. The purpose of this study is to examine the influence of a more applicant-friendly warning on faking, as well as see if such a warning will improve applicant perceptions of the use of a personality test in selection.

### *Faking Behavior*

Presumably, applicants for a job are motivated to show themselves to be a relatively ideal candidate for the organization. As such, applicants could be tempted to

distort their responses toward some conception of the type of person who best fits a given job. Many studies have found higher (more socially desirable) personality trait group means for faking groups than control/non-faking groups (Barrick & Mount, 1996; Hough, 1998; Rosse, Stecher, Miller, & Levin, 1998). Typically, this distortion is referred to as faking, but various researchers have called it impression management, response distortion, intentional distortion, social desirability, and dissimulation (McFarland & Ryan, 2000).

In a sample of recent job applicants, Donovan et al. (2003) assessed the base rate of applicant faking. The researchers created a survey asking participants whether or not they had engaged in 29 faking activities during a recent job application process. Responses showed that participants tended to fake the personality components of the selection process more often than other components. Participants reported faking on components relating to personality 27.8% to 53.8% of the time versus only 4.7% to 9.5% for the biographical components. The aspects of personality addressed included hardworking, prompt, and thorough (45.2%), dependability and reliability (29%), agreeableness (27.8%), and downplaying negative attributes (53.8%). Biographical data showed a much lower prevalence of faking – outright made up information (9.5%), listed unearned rewards (4.7%), and exaggerated experience (7.7%). Thus it appears that faking in personality testing may occur in a meaningful proportion of the job applicant population.

There is still debate as to the effects of faking on the validity of personality measures in employee selection (Barrick & Mount, 1996; McFarland & Ryan, 2000; Mueller-Hanson, Heggstad, & Thornton, 2003; Ones, Viswesvaran & Reiss, 1996).

Topping & O’Gorman (1997) found that for all Big Five traits but Agreeableness, intentionally faked responses significantly decreased the validity of responses to the NEO-FFI when compared to personality ratings from friends or family. On the other hand, in a study using two applicant samples, Barrick and Mount (1996) found that when correcting for response distortion, criterion-related validities against voluntary turnover and job performance decreased, but not a significant amount. The validities were similar despite the finding that the structural models correcting for either self-deception or impression management fit better than the unadjusted model. In a meta-analysis of the social desirability literature, Ones et al. (1996) found similar results regarding the robustness of criterion-related validities, as have others (Hough, 1998; Hough, Eaton, Dunnette, Kamp, & McCloy, 1990).

There is much evidence, however, that faking responses can change the rank order of candidates. In selection scenarios with low selection ratios, this rank order is what will determine who gets hired and who does not. Mueller-Hanson et al. (2003) simulated the applicant process to examine the criterion-related validity and rank-order effects of faking. For a control condition, there were no differences in the predictor-criterion relationships among three achievement motivation subgroups (low, middle, and high achievement motivation). For the faking condition, however, the personality-performance relationship for the lower ( $r = .45$ ) and upper thirds ( $r = .07$ ) were significantly different. Further, the applicants selected based on their rank-order on the non-cognitive measure were consistently composed of more from the faking group than the control group. As the selection ratio decreased, the proportion of faking group members selected increased, and thus the error in selection. An increase in the error in selection due to faking has been

found consistently in other research looking at the effect of faking on rank-order (Ellingson, Sackett, & Hough, 1999; Hough, 1998; Komar, Brown, Komar, & Robie, 2008; Rosse et al., 1998).

There is evidence that faking affects the psychometric properties of personality tests such that the Five-Factor structure does not always fit faked responses. Schmit and Ryan (1993) hypothesized that personality factor structure would depend on the purpose of the test administration. Two such conditions include a simple for-research-only examination of personality, or as used in a selection battery. Results showed that the Five Factor structure fit better for student (voluntary) samples than applicant (involuntary) samples. In the applicant sample, an “ideal-employee factor” (Schmit and Ryan, 1993, pp. 971) appeared in the factor analysis, reflecting several work-related traits. The authors believe the NEO-FFI’s five broad personality traits may not be accurately applied in employee selection. Furthermore, the subscales of the NEO-FFI showed complex loading patterns, so these subscales should also be used cautiously when assessing applicant responses.

Smith, Hanges, and Dickson (2001), on the other hand, found no changes in factor structure using the HPI between student, applicant, and incumbent samples. Marshall, de Fruyt, Rolland, and Bagby (2005) similarly found that the factor structure on the NEO-PI-R remained stable across applicant and non-applicant groups. Unexpectedly, in Smith et al.’s (2001) study, the model fit the applicant sample better than the student sample, in disagreement with Schmit and Ryan’s (1993) findings. Unfortunately, I have found no further studies seeking to resolve the differences between Schmit and Ryan and Smith et al.’s findings.

Potential inconsistency of factor structures has serious implications for the utility of personality tests in the selection setting. Consider conscientiousness, the trait most often found to be related to job performance across occupations. If items used in creating the composite conscientiousness score do not fit in the conscientiousness latent factor, the inclusion of such items may introduce noise into the predictor-criterion relationship, thus reducing this relationship.

### *The Utility of Warnings*

As alluded to above, items on most personality tests are somewhat transparent. On most measures, a general knowledge of the Big-Five model will enable respondents to identify which trait each item reflects, and therefore determine how to fake their responses to the item, should they choose to do so. Rees and Metcalfe (2003) found that 36% of participants thought it was easy to fake on a personality questionnaire, and another 36% did not find faking unethical. Wolford and Christiansen (2008) approached the fakeability issue from a slightly different perspective. With the increasing popularity of personality measures being used in the workplace, books have appeared that coach readers on how to fake personality tests. In the Wolford and Christiansen study, the authors provided some participants with sections of a book that coached readers how to improve their scores on a personality test; another group did not receive this material. During the testing session, all participants were told to respond to the personality test as if they were applying for a car salesman position. Results showed that those who read the coaching materials did show higher mean scores for extraversion and conscientiousness than those who did not receive the coaching materials. The results suggest that with knowledge of what a personality test is asking for, and the motivation to fake,

participants can distort their responses in such a way as to increase their overall scores on the test.

There has been a great deal of research the past decade or so looking into how to reduce such distortion. Although faking has been found to affect applicant rank-order and the scale's psychometric properties (as discussed above), researchers have developed methods for dealing with applicant faking behavior. Warnings have been shown to effectively reduce the prevalence of faking in the application process (Dwight & Donovan, 2003). For many personality traits, mean scores for the warned conditions were significantly lower than for the unwarned condition. However, a recent study comparing honest, faked, and warned groups suggests that warning the applicant not to fake may not reduce test scores down to the same level as the honest condition (Donovan, Dwight, & Schneider, 2008). Research needs to be conducted looking more precisely at how warnings impact personality test scores.

McFarland and Ryan (2000) provided a model of the applicant faking process to describe the way variables interact to create variance in faking on non-cognitive measures. First in the model are influences on beliefs toward faking, including values, morals, religion, personality traits, etc. These influences then affect an individual's beliefs toward faking, which in turn determine an individual's intention to fake. The relationship between beliefs toward faking and intention to fake, however, is moderated by situational influences such as desire for the job and warnings. Intention to fake's relationship with faking behavior is moderated by both the ability to fake (self-monitoring, knowledge of the construct being measured, and item transparency) and the opportunity to fake. Opportunity to fake addresses the limitation for fakers that those

already high on the trait may not be able to positively distort their responses. Finally, the model asserts faking behavior will influence a number of outcomes including validities, test scores, scale reliabilities, and the factor structure. Although the model has not yet been tested in its entirety, it does provide a framework from which researchers can examine the expected effects of warnings on faking behavior.

Pace and Borman (2006) summarized five types of warnings that could be used to reduce applicant faking. First is the detection warning, in which test takers are informed that faked responses can be detected. Second is the consequence warning, which refers to the test administrator explaining the penalty a respondent will receive for faking his or her responses. Independently, these two warnings appear to have little effect on the prevalence of faking. Combined, however, they have been shown to be effective at reducing trait mean scores (Dwight & Donovan, 2003). In the Dwight and Donovan study, the condition in which the detection warning was paired with the consequence warning elicited lower mean scores (reduced faking) compared to the unwarned condition, detection warning-only condition, and consequence-only condition. The detection/consequence warning has had the most wide-spread use in the warning literature.

The moral conviction warning appeals to the applicant's sense of right and wrong. Upon reading, it presumably brings to mind associations of honesty with appropriate behavior and positive self-image, which may reduce faking levels as well.

Another approach to reducing faking, although not necessarily considered a warning, is the for-research-only prompt. Essentially, the for-research-only prompt explains why personality tests are used in selection, and how accurate responses aide in

improving the selection process. The administrator explains that a person's responses will not be tied to the person's name or any other identifying information, nor will responses affect the respondent in any way. Responses from this approach, since applicants will have a decreased motivation to fake, are presumably closer to their true personality scores.

Lastly, there is the reasoning warning, the "let's reason together" approach (Pace, Borman, Penney, Xu, & Bearden, 2005). This warning appeals to test-taker interests rather than the consequences of faking. In a friendly tone, the warning points out that it may not be in the test-taker's best interest to fake his or her responses because dishonest responses may result in the respondent being classified into jobs he or she is not best suited for. It continues by pointing out that incorrect classification may result in performance problems and a lack of job fit. The researchers developed this instruction prompt for use in a military setting, where test takers will be classified into jobs based on the test results. In the typical organizational selection context, the reasoning warning would need to be rephrased to reflect conditions of being offered the job (or not) while still maintaining the friendly, helpful tone.

The reasoning warning appeals to many levels of McFarland and Ryan's (2000) model of faking. The warning could affect beliefs toward faking in that it should increase the belief that faking is not useful and faking will not help the applicant reach his or her goals. Further, the warning may make the applicant realize that he or she does not fit the job, based on the personality traits needed to be successful and satisfied in the job. Lastly, with a decreased belief in the effectiveness of faking, there will be a decreased intention

to fake. It is also possible, however, that such a warning could inform an applicant that it is possible to fake responses, thereby leading to an increase in faking behavior.

Pace et al. (2005) conducted the only study I am aware of utilizing the reasoning warning. Although the study was terminated before all data were collected, preliminary results showed promise for the reasoning warning. Based on the available data, there were no significant mean differences between personality scores on a number of personality traits for the detection/consequence warning condition, reasoning warning condition, and for-research-only condition. Due to the early termination of the Pace et al. (2005) study, however, the researchers were unable to compare the means from the three warning conditions with those of an unwarned group.

### *Procedural Justice*

Procedural justice has been defined as the fairness of the procedures that are used to determine organizational outcomes (Lind & Tyler, 1988). There is much evidence that applicants tend to have negative perceptions of the use of personality measures in the selection process because personality tests are not seen as directly relevant to the job itself (Holtz, Ployhart, & Dominguez, 2005; Gilliland, 1993; Gilliland, 1994). The content of a personality test is not closely related to the content of the job, and it is not clear to most people what the relationship is between personality traits and job outcomes.

Gilliland (1993) attempted to integrate the literature on applicant perceived procedural justice in the selection context by proposing a model of applicant reactions to employment selections systems. While the model attempts to explain applicant reactions to the entire selection process, I will focus only on the parts of the model relating to the testing itself, where personality testing can best fit. Gilliland identifies four procedural

rules relating to the formal characteristics of the selection process – job relatedness, opportunity to perform, reconsideration opportunity, and consistency. Job relatedness is the extent to which the test itself appears to be related to the job content and could be a valid predictor of job performance. The opportunity to perform refers to the applicant having a say in the information used for the selection decision. The reconsideration opportunity rule refers to the applicant’s opportunity to challenge the decision process. Specifically, applicants need to be able to view their scores and understand the scoring process. Finally, the consistency of administration rule refers to the need for applicants to believe that all applicants were treated fairly and equally in the selection process. Other rules relevant to the selection battery itself include honesty and the propriety of questions. Honesty is the truthfulness the organization shows when communicating with applicants. The propriety of questions refers to items being appropriate and not showing prejudice. Combined, the above rules interact to influence the applicant’s overall opinion of the organization’s selection process.

Gilliland tested certain components of his proposed model in a later study (Gilliland, 1994). He found that perceived procedural justice was higher for applicants whose selection procedures were viewed as most relevant to their job. He manipulated the relevance of the testing procedure to the job by having groups of participants take different tests (a work sample test, a cognitive ability test, or an overt integrity test). He found that the work sample test was seen as most relevant to the future job, whereas the integrity test was seen as the least relevant. Further, participants completing the work sample test held the greatest perception of procedural justice, whereas those completing the overt integrity test held the lowest.

Using a threat (in the form of a detection/consequence warning) in the application process can negatively influence the applicant's perception of the organization (Gilliland, 1993; McFarland, 2003). McFarland (2003) examined the direct effect of the detection/consequence warning on applicant's perception of procedural justice. She outlined a number of ways warnings can negatively affect applicant perceptions of the selection process. First, she pointed out that the detection warning could make applicants feel that the organization is not allowing them to perform to as high a degree as possible. Second, a detection warning informs applicants that the test can, indeed, be faked. As such, applicants could believe that the measure would not accurately predict anything about a future employee's performance. Third, including items that assess dishonesty in the personality test could be considered an unwarranted invasion of privacy.

McFarland hypothesized that a warning which explains that a measure of social desirability is within the scale, and explains what the scale will be used for, will improve applicant perceptions of the selection process. Although means on multiple components of procedural justice for the warned group were lower, the differences were non-significant, suggesting that warnings had little effect on perceived procedural justice. However, McFarland points out that the use of personality tests in selection is perceived negatively overall, as seen in low mean scores on measures of procedural justice. Other research has supported the notion that including a personality test in a selection battery produces negative applicant reactions (Rosse et al., 1994; Steiner & Gilliland, 1996), and Rees and Metcalfe (2003) found a quarter of participants thought personality questionnaires could not effectively predict job performance, suggesting these personality tests violate a number of procedural justice rules.

As addressed above, Pace and Borman (2006) proposed the reasoning warning as a more applicant-friendly approach to reducing faking. The use of the reasoning warning could improve perceptions of procedural justice in the use of personality tests for selection by explaining to respondents that honest answers do, in fact, make personality measurement effective and honest answers are in their own best interests. The detection/consequence warning could have shown no effect on perceived procedural justice in McFarland's (2003) study because it did not directly address any of Gilliland's (1993) procedural justice rules. Logically, explaining the reasons behind including the personality test in the selection battery, and honesty from the organization can improve applicant perceptions of procedural justice through Gilliland's job relatedness and honesty rules (as discussed above). Related to the McFarland and Ryan (2000) model, the reasoning warning could also reduce the intention to fake due to a desire to reciprocate with honesty and fairness to the organization. In sum, the reasoning warning could potentially improve applicant perceptions of procedural justice through addressing several procedural justice rules.

### *Identifying Faking*

Several methods for detecting faking and dealing with its effects have been proposed in the research. Below, I will briefly summarize three approaches – self-presentation or direct evidence, mean comparisons, and variance comparisons.

### *Self-Presentation/Direct Evidence*

Some personality measures have items built into them meant to identify fakers. Using one or more of these methods, statistical correction can be made to obtain an adjusted score for applicants (Ellingson et al., 1999). However, these statistical

corrections have not shown improved validity (Barrick & Mount, 1996; Mueller-Hanson et al., 2003). Alternatively, researchers have attempted to identify faking and honest groups based on a median split on impression management scales. Stark, Chernyshenko, Chan, Lee and Drasgow (2001) examined the psychometric properties of a number of personality scales using IRT differential item functioning (DIF) analysis, with honest and faking groups determined through such a median split. They found that fewer items showed DIF using the median split on an impression management measure compared to data from applicant and non-applicant samples. They concluded that due to DIF and differential test functioning, the scale might measure different underlying constructs for applicant and non-applicant samples. As such, this median split should not be applied to the analysis and interpretation of personality data.

Rosse et al. (1998) controlled for response distortion using an impression management scale. They found that the rank order of applicants was very different between controlled and uncontrolled applicant personality scores, especially at the top of the distribution, where employment decisions at smaller selection ratios will be made. While controlling for social desirability has been used in an attempt to decrease the effects of faking on selection, researchers are not entirely clear if controlling for social desirability or impression management actually reflects the individual's true personality.

Ellingson et al. (1999) compared honest and corrected applicant scores in a within-subjects study. They had participants take a personality test twice, once under an honest condition, and once under a "fake good" condition, then obtained corrected scores by correcting the "fake good" scores for intentional distortion using a social desirability scale. Although correcting for distortion did bring group mean scores for the faking group

down to the level of honest condition means, the correlations between honest and corrected scores were low ( $r = .09-.26$ ). To assess the effect of correcting scores on selection, Ellingson et al. created a series of selection scenarios based on mock-applicant personality test scores then examined which participants would be selected using a top-down selection procedure using each person's true and corrected scores. The authors considered a correct selection decision to have occurred when applicants who were selected based on their honest scores were also selected with their corrected scores. In some scenarios, social desirability corrections improved the proportion of applicants correctly selected. In other scenarios, corrections had no effect or reduced the proportion correctly selected. Consequently, it is possible that applicant rankings based on corrected personality scores may not result in true-score rankings. Thus, a social desirability measure may not be the most effective method for identifying or controlling for applicant faking behaviors.

Some researchers maintain that social desirability may reflect valid personality-related variances (Ellingson et al., 1999; Ones et al., 1996) and should not be used to correct personality scores. Furthermore, there is a body of research that has looked at the effectiveness of adding a frame of reference to the personality measure. For example, Schmit, Ryan, Stierwalt, and Powell (1995) added a school context to each item on a Conscientiousness scale. They found increased validity for all Conscientiousness subscales using a school frame of reference in predicting performance relative to college GPA, as compared to the scale with no frame of reference. Hunthausen, Truxillo, Bauer, and Hammer (2003) found similar results using a work-place sample and job performance criterion. Both studies' results suggest that socially desirable responding can

reflect true work performance, assuming the applicant does exhibit the reported behaviors at work, even though his or her personality-based behaviors outside the work context may be notably different.

### *Mean Comparisons*

The most common method for identifying faking is the mean comparison method using a between-subjects design (see Birkeland, Manson, Kisamore, Brannick, & Smith, 2006 meta-analysis) (Barrick & Mount, 1996; Dwight & Donovan, 2003; McFarland, 2003; McFarland & Ryan, 2000; Rosse et al., 1998). As addressed above, considerable research has found a statistical increase in average scores for faking groups (Barrick & Mount, 1996; Hough, 1998). Researchers have used a number of methods to identify or create the faking group used in these mean comparisons. In some studies using non-applicant samples, participants have been given a reward for “getting” the job (financial incentives). Others have instructed participants to intentionally distort their responses so they get the job. In one study, McFarland (2003) simply asked if applicants lied on the test to increase their scores.

For applicant samples, some researchers have presumed that applicants are motivated to fake their responses to make the best impression they can, whereas an incumbent sample should show much less faking behavior (Rosse et al., 1998). Hogan, Barrett, and Hogan (2007) took a different approach. Applicants completed a personality test during an initial employment process. The majority of applicants were not offered jobs. Six months later, many of those who did not previously receive a job offer reapplied to the same organization. Hogan et al. suggested that these re-applicants would be more

highly motivated to get the job the second time around, although they did not find support for this supposition.

Regardless of the method used to induce or discourage faking, the analyses are simple – compare group means on individual personality traits between the faking and non-faking groups and examine the effect sizes to evaluate differences.

### *Variance*

For some of the most popular personality tests, such as the NEO-PI-R and NEO-FFI (Costa & McCrae, 1992) participants respond to the personality measure using a five-point Likert scale according to how much they agree the items describe them. Assuming applicants want the job, responses may be positively distorted, and moving responses toward the “five” mark will reduce the response variance, so faking groups should be expected to show less variance than an honest group.

Also, some researchers have found that applicants distort their responses to a different extent on different personality variables (McFarland & Ryan, 2000; Rosse et al., 1998). These findings suggest that applicants view certain personality traits as more relevant to the job than others, demonstrating another source of variance in faking conditions. Identifying a faking group based on different variances, however, would be difficult to argue. Furthermore, comparing variance would do little to help identify individual fakers. The value in comparing variance between conditions would be to suggest there are different treatment effects between groups.

### *Summary*

Presumably, when applicants are within the selection context, they will want to present the most favorable impression of themselves. While the composition of the Big

Five personality traits vary depending on the measure, all seek to assess five similar, recognizable traits. When distorting responses, the applicant who is trying to appear more favorable than other applicants would likely respond more highly on items related to Extraversion, Openness, Agreeableness, and Conscientiousness, and less high on items related to Neuroticism.

In sum, although evidence suggests that faking on a personality test does not seriously affect criterion-related validities for a number of important organizational outcomes, faking does affect the rank-order of job candidates, especially near the top of the score distribution. Furthermore, the inclusion of a personality measure in the selection process can negatively influence perceptions of the selection process, and perhaps the organization. As such, I will test the effect of the reasoning warning in reducing the prevalence of faking and the perception of procedural justice. If the reasoning warning can be shown to be as effective as the detection/consequence warning at reducing applicant faking on a personality measure, and if the reasoning warning shows improved perceptions of procedural justice, then the reasoning warning could be argued to be the better warning prompt to use. I hypothesize that:

*H1: A personality test with no warning will show higher mean scores for Extraversion, Openness, Agreeableness, and Conscientiousness than the for-research-only (control) condition, the detection/consequence warning, and the reasoning warning, and lower mean scores for Neuroticism.*

*H2: A personality test with the detection/consequence warning or the reasoning warning will show no difference in mean scores from the for-research-only condition for any personality variables.*

H3: *The perception of procedural justice in the reasoning warning condition will be higher than with the detection/consequence warning condition.*

## Chapter Two

### Method

#### *Participants*

Participants were recruited from two sources. The first sample (Nursing sample) came from nurses working in the medical field in Louisville, Ky. The other source was undergraduate students in psychology courses at a southeastern United States public university (Student sample). I obtained responses from 267 nursing participants (89.5% female, 9% male, 1.5% missing). Ages ranged from 18 to 58 ( $M = 25.63$ ;  $SD = 6.51$ ). The majority of the nursing participants were White/Caucasian (82.8%), with 13.5% Black/African American, .7% Asian/Pacific Islander, .4% Native American, and 2.6% Other/Mixed Race.

I obtained 155 responses from the Student sample. Two responses showed uniform responding throughout, and two more responses were completed in around one minute (96 items), which was two minutes shorter than the next response time, so 151 responses were included in the analyses (79.5% female, 20.5% male). Ages ranged from 18 to 42 ( $M = 21.75$ ,  $SD = 3.64$ ). The majority of Student participants were White/Caucasian (49.7%), with 19.9% Black/African American, 17.2% Hispanic/Latino, 7.3% Asian/Pacific Islander, and 6% Other/Mixed Race. This sample was employed in a number of industries including the medical industry (30.5%), retail/service industry (17.9%), and professional industries (15.2%), among others, with 25.2% unemployed.

As I have no expectations for differences between groups based on gender, race, tenure, employment status, etc., there were no exclusionary criteria for participants in either sample. I collected no identifying information above basic demographics from participants. No compensation was given to participants, and participation was completely voluntary for both samples.

### *Measures*

*Warning Prompts:* Participants received one of four instruction prompts. In order to create a control condition with honest responses, the for-research-only condition (Honest condition) differed from the other conditions. Participants in the for-research-only condition were presented with only the following paragraph preceding the personality test:

For-research-only condition (Honest condition):

*You are about to take a personality test. As you answer the following questions, please be as honest as you can. Your responses will be used for research purposes only. There will be no identifying information kept with your responses, and all responses will be kept strictly confidential. Honest answers will help us to get an idea of the typical person's true personality.*

Participants in the other three conditions (no warning, detection/consequence warning, and reasoning warning) were told to pretend they are a job applicant trying to get a job in the field of nursing. They were also told the personality test they were about to take would be a key part of the job selection process. This general instruction prompt was included in the remaining three forms:

General instruction prompt:

*Pretend you are a job applicant trying to get your ideal job in the field of nursing. The personality test you are about to take is a very important part of the job selection process, so it is important that you do well. Please respond to the test as you would if you were applying for this job. The test will be used in the decision to hire all job candidates.*

The Detection/Consequence and Reasoning warnings then followed the general instruction prompt with their respective warnings:

Detection/Consequence warning:

*Please be aware of the following two points:*

- 1. This test contains questions designed to identify those who slant their responses to make themselves look like a better candidate than they are. Research has shown that these questions are an effective way of identifying individuals who provide inaccurate information about themselves.*
- 2. Dishonest or distorted self-descriptions will invalidate your results. In other words, faking will result in your not being considered for the job.*

Reasoning warning:

*This test has been designed to find job candidates who will be most successful and happy in this job. Please respond honestly to the following questions. Those who respond dishonestly may find themselves in jobs they are not well suited for, which may in turn result in poor performance and dissatisfaction with the job, so it is in your own best interest to answer the following questions honestly.*

*Personality Test:* I used the International Personality Item Pool (IPIP) (Goldberg et al., 2006; Appendix A) as my personality assessment tool. The IPIP has 10 items for

each Big Five trait (50 items total); Neuroticism ( $\alpha = .84$ ) (“Often feel blue”), Extraversion ( $\alpha = .84$ ) (“Feel comfortable around people”), Openness to Experience ( $\alpha = .62$ ) (“Have a vivid imagination”), Agreeableness ( $\alpha = .75$ ) (“Have a good word for everyone”), and Conscientiousness ( $\alpha = .84$ ) (“Am always prepared”). The IPIP items included in assessment of the Big Five traits were selected based on rank-ordered correlations with scores on the NEO PI-R. The items showing the highest correlations with each trait were selected to be included in the measure. The developers performed a visual content analysis to identify items that addressed the same issue. If they found two items that were too similar in content, they removed the item with the lower correlation, and the next item in the rank-order was included. Responses are on a five-point Likert-type scale, with anchors of 1 = Very Inaccurate, 3 = Neither Inaccurate nor Accurate, and 5 = Very Accurate.

*Procedural Justice:* I used a modified version of Colquitt’s (2001) Procedural Justice Questionnaire (PJQ; Appendix B). Colquitt created the items based on Leventhal’s (1980) procedural justice rules. The phrasing of some of the original items referred to a present outcome rather than the current selection process, so I made tense and small phrasing changes to make the items fit the current study’s testing situation. A sample item includes “Have you had an influence over the outcome arrived at by these procedures?” The measure is a seven-item scale with a five-point Likert-type response with anchors 1 = to a small extent, and 5 = to a large extent. Internal consistency reliability is high,  $\alpha = .83$ .

I also used certain subscales of Bauer et al.’s (2001) Selection Procedural Justice Scale (SPJS; Appendix C). Since the focus of this study is on the applicant perception of

procedural justice regarding the use of a personality test in the selection process and the use of various warnings, I am only including the subscales relating to the test itself. The included subscales are job-relatedness – predictive (“A person who scored well on this test will be a good nurse,”  $\alpha = .83$ ), chance to perform (“I was able to show what I can do on this test,”  $\alpha = .91$ ), propriety of questions (“The test itself did not seem too personal or private,”  $\alpha = .78$ ), and job-relatedness content (“The content of the test was clearly related to the nursing job,”  $\alpha = .93$ ). The included subscales have 11 items total, with a high reliability of  $\alpha = .84$ .

*Impression Management:* I have also included the Impression Management (IM) subscale of Paulhus’s Balanced Inventory of Desirable Responding (BIDR) (Paulhus, 1991; Appendix D). Paulhus defines impression management as intentionally changing behavior for an audience. While much research has shown that accounting for variance in responding to personality tests in the selection context based on social desirability does not bring individual scores do the “honest” level, social desirability scales are still often used in personality research. The measure is a 20 item scale with a seven-point Likert-type response with anchors 1 = Not True, 4 = Somewhat True, and 7 = Very True. Reliability for the IM scale was found to be moderate,  $\alpha = .77$ .

### *Procedure*

I collected data from the Nursing sample in a series of testing sessions during two typical work days. Testing sessions ranged from early morning to late evening to cover multiple work shifts, with 13 to 56 participants in each session, for a total of eight sessions. The testing session involved a short introduction by the experimenter explaining that participation was entirely voluntary and no adverse outcome would result from them

deciding not to participate. Participants then read and signed the informed consent form, which was collected before distributing the surveys. After all consent forms had been collected, participants were given the experiment's survey face down, instructed to read the survey instructions, and completed the measure. Once all participants had finished, the experimenter debriefed participants as to the purpose of the study, including a brief summary of previous research on warnings and revelation of the different warning prompts presented. Any remaining questions were also answered. Before ending the debriefing session, the experimenter asked participants not to discuss the study with any of their coworkers, so as not to compromise the manipulation.

The Student sample completed the survey through the online data-collection SONA system. This system required participant to sign into the website, then select the study to participate in. They were asked to select "ok" to indicate their consent to participate. They were then directed to a digital version of the survey to complete. Due to restraints on the SONA system, data had to be collected from one condition at a time. The order of conditions collected was Reasoning Warning, Detection/Consequence Warning, Faking, then Honest.

## Chapter Three

### Results

Many of the analyses below required supplementary tests to an omnibus statistic (ANOVA). In these cases, I used the Bonferroni procedure to control for Type I error. In the Bonferroni procedure, the alpha level (often .05) is divided by the number of post hoc analyses performed. For example, if five post hoc analyses are performed to an ANOVA, an alpha level of .01 would be used to test the post hoc tests for significance. Also, interpreting effect sizes, I used Cohen's (1988) conventions of  $d = .2$  as a small effect,  $d = .5$  as a medium effect, and  $d = .8$  as a large effect (Cohen, 1988).

#### *Comparison of Sources*

The first step in my data analyses was to examine participant responses for differences between the Nursing and Student samples using the multivariate MANOVA test with Source (Student or Nurse) and Survey (Reasoning, Detection/Consequence, Faking, Honest) as the independent variables, and all personality, justice, and impression management variables as the dependent variables. The MANOVA for Survey was significant, Wilk's  $\Lambda = .86$ ,  $F(36, 1158.94) = 1.74$ ,  $p = .005$ . Follow-up ANOVAs suggested that differences could be found only for Neuroticism and Impression Management (Table 1). All other variables had low power (below .55), suggesting significant results may not have been detected.

Table 1

*Follow-up ANOVAs to Survey MANOVA: Full sample*

	MS	F	p	power
Neuroticism	1.8	3.86	.010	.82
Extraversion	.98	2.12	.097	.54
Openness	.30	1.19	.314	.32
Agreeableness	.55	2.02	.111	.52
Conscientiousness	.17	.42	.736	.14
PJQ	.06	.11	.957	.07
SPJS	.47	1.19	.314	.32
JR – Predict	1.88	2.15	.093	.55
Perform	.16	.20	.895	.09
Propriety	1.29	1.84	.139	.48
JR – Content	2.38	2.27	.080	.57
Impression Mgmt	2.34	3.74	.011	.81

\* Alpha level = .05; results are for F(3, X), where X varies depending on sample size due to missing data.

Table 2

*Follow-up ANOVAs to Source MANOVA: Full sample*

	MS	F	p	power
Neuroticism	1.72	3.68	.056	.48
Extraversion	.98	2.13	.145	.31
Openness	.77	3.05	.081	.41
Agreeableness	8.08	29.79	.000	.99
Conscientiousness	9.61	24.62	.000	.99
PJQ	.73	1.26	.263	.20
SPJS	.01	.01	.915	.05
JR – Predict	4.05	4.64	.032	.58
Perform	1.71	2.23	.136	.32
Propriety	3.83	5.45	.020	.64
JR – Content	4.13	3.94	.048	.51
Impression Mgmt	5.34	8.52	.004	.83

\* Alpha level = .05; results are for F(3, X), where X varies depending on sample size due to missing data.

The MANOVA for Source was also significant, Wilk's  $\Lambda = .84$ ,  $F(12, 392) = 6.25$ ,  $p = .000$ . Follow-up ANOVAs suggested that differences could be found for Agreeableness, Conscientiousness, Job Relatedness – Predictive, Propriety, Job Relatedness – Content, and Impression Management. As with the Survey follow-ups, all non-significant variables had low power of .48 or below (Table 2).

There was no significant multivariate interaction between Survey and Source, Wilk's  $\Lambda = .912$ ,  $F(36, 1158.94) = 1.019$ ,  $p = .439$ , power = .93.

Since the Source analyses showed that there were several differences between variables due to the sample but the interaction was not significant, I decided to analyze Students and Nurses separately. *M* and *SD* for all variables by condition can be found in Tables 3 and 4, and bivariate correlations among variables by Source can be found in Tables 5 and 6).

### *Personality*

I had two hypotheses regarding the personality measures. I first hypothesized that responses from the Faking condition would show higher mean scores for Extraversion, Openness, Agreeableness, and Conscientiousness than the Honest, Faking, and Reasoning conditions, and lower mean scores for Neuroticism. I also hypothesized that the Honest condition would show no difference in mean scores from the Reasoning warning condition nor the Detection/Consequence condition. While there is a fair amount of support suggesting that the Detection/Consequence warning elicits responses that are consistently less socially desirable than the Faking condition, the remaining hypotheses are less clearly summarized.

Table 3

*Means and standard deviations for Nurse and Student samples for the IPIP*

Variable by Condition	Nurse		Student	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Reasoning	N = 65		N = 38	
Neuroticism	2.32	.73	2.66	.72
Extraversion	3.61	.68	3.47	.80
Openness	3.48	.50	3.66	.40
Agreeableness	3.92	.54	3.60	.53
Conscientiousness	3.96	.64	3.58	.61
Detection/Consequence	N = 67		N = 37	
Neuroticism	2.49	.63	2.61	.61
Extraversion	3.47	.68	3.42	.68
Openness	3.35	.49	3.50	.48
Agreeableness	3.83	.54	3.61	.53
Conscientiousness	3.83	.54	3.61	.53
Faking	N = 69		N = 38	
Neuroticism	2.12	.64	2.50	.68
Extraversion	3.83	.62	3.49	.67
Openness	3.49	.52	3.56	.44
Agreeableness	4.07	.49	3.69	.49
Conscientiousness	4.13	.63	3.49	.69
Honest	N = 66		N = 38	
Neuroticism	2.77	.73	2.48	.69
Extraversion	3.57	.65	3.72	.68
Openness	3.53	.52	3.50	.60
Agreeableness	3.87	.54	3.61	.50
Conscientiousness	3.74	.66	3.71	.67

Table 4

*Means and standard deviations for Nurse and Student Samples for procedural justice*

Variable by Condition	Nurse		Student	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Reasoning	N = 64		N = 38	
PJQ	3.63	.84	3.63	.73
SPJS	2.66	.69	2.59	.65
Job Relatedness – Pred	2.27	1.00	2.51	.86
Perform	2.13	.92	2.22	.86
Propriety	3.88	.90	3.48	.79
Job Relatedness – Content	2.29	.99	2.09	1.06
Detection/Consequence	N = 66		N = 37	
PJQ	3.65	.75	3.51	.73
SPJS	2.60	.70	2.52	.63
Job Relatedness – Pred	2.25	.97	2.36	.95
Perform	2.18	.86	2.23	.84
Propriety	3.67	.95	3.51	.83
Job Relatedness – Content	2.16	.98	1.77	.82
Faking	N = 68		N = 38	
PJQ	3.66	.88	3.53	.55
SPJS	2.70	.56	2.77	.52
Job Relatedness – Pred	2.46	.87	2.72	.93
Perform	2.17	.80	2.36	.77
Propriety	3.85	.82	3.61	.76
Job Relatedness – Content	2.30	1.08	2.39	1.27
Honest	N = 65		N = 38	
PJQ	3.62	.80	3.53	.61
SPJS	2.65	.66	2.72	.59
Job Relatedness – Pred	2.19	.92	2.39	.98
Perform	2.16	.97	2.36	.95
Propriety	3.86	.87	3.92	.69
Job Relatedness - Content	2.28	.94	1.95	1.06

Table 5

*Intercorrelations among study variables: Nurse sample*

	1	2	3	4	5	6	7	8	9	10	11
1. Neuroticism	--										
2. Extraversion	-.42	--									
3. Openness	-.12	.15	--								
4. Agreeableness	-.43	.15	.31	--							
5. Conscientiousness	-.52	.17	.22	.47	--						
6. PJQ	-.13	.03	.07	.22	.15	--					
7. SPJS	-.09	-.01	.06	.14	.11	.39	--				
8. JR – Predict	-.11	.01	.16	.13	.09	.19	.67	--			
9. Perform	-.05	-.01	.02	.11	.06	.25	.85	.57	--		
10. Propriety	-.09	.07	.08	.10	.13	.45	.58	.19	.22	--	
11. JR – Content	-.01	-.12	-.08	.04	.05	.18	.64	.29	.44	.18	--
12. Impression Mgmt	-.22	-.05	.05	.38	.42	.10	.14	.16	.12	.04	.07

\* Correlations at magnitude .12 and above are significant at the .05 level.

Table 6

*Intercorrelations among study variables: Student sample*

	1	2	3	4	5	6	7	8	9	10	11
1. Neuroticism	--										
2. Extraversion	-.43	--									
3. Openness	-.17	.15	--								
4. Agreeableness	-.33	.17	.24	--							
5. Conscientiousness	-.46	.32	.21	.29	--						
6. PJQ	-.20	.17	-.01	.27	.31	--					
7. SPJS	-.15	.02	-.03	.06	.03	.07	--				
8. JR – Predict	-.08	-.01	.05	.03	-.03	.00	.69	--			
9. Perform	-.02	.02	-.02	-.01	.02	-.07	.87	.58	--		
10. Propriety	-.25	.14	.05	.19	.22	.37	.43	.07	.13	--	
11. JR – Content	-.08	-.11	-.15	-.02	-.14	-.09	.62	.27	.45	-.03	--
12. Impression Mgmt	-.24	-.02	.17	.48	.18	.06	.07	.09	.04	-.03	.09

\* Correlations at magnitude .17 and above are significant at the .05 level.

To test my personality hypotheses, I performed a series of one-way ANOVAs with the survey form as the independent variable and the mean score for each of the five personality variables as the dependent variables. For the Nursing sample, one-way ANOVAs showed significant differences between condition means for Neuroticism,  $F(3, 1) = 11.09, p = .000, \text{power} = .999$ , Extraversion,  $F(3, 1) = 3.81, p = .011, \text{power} = .814$ , Agreeableness,  $F(3, 1) = 2.69, p = .047, \text{power} = .650$  and Conscientiousness,  $F(3, 1) = 4.95, p = .002, \text{power} = .910$ , but not for Openness,  $F(3, 1) = 1.51, p = .212, \text{power} = .397$  (Table 5). I then followed each significant ANOVA with Tukey's post-hoc tests based on my five planned comparisons. In order to control for Type 1 error, I used the Bonferroni procedure to set .01 as my significance criterion for each Tukey's post-hoc test. A summary of all one-way ANOVAs and Tukey's post-hoc tests can be found in Table 7.

For Neuroticism, the mean for the Reasoning condition ( $M = 2.32, SD = .73$ ) was significantly lower than the Honest condition ( $M = 2.77, SD = .73$ ),  $p = .001$ , with a moderate effect size,  $d = -.62$ , and the Faking condition ( $M = 2.12, SD = .64$ ) was significantly lower than the Honest condition,  $p = .008, d = .58$ . The Faking condition was also lower for the Detection/Consequence condition ( $M = 2.49, SD = .63$ ),  $p = .000, d = .73$ .

For Extraversion, the Detection/Consequence condition ( $M = 3.47, SD = .68$ ) was significantly lower than the Faking condition ( $M = 3.83, SD = .62$ ),  $p = .007$ , with a moderate effect size,  $d = .55$ . For Conscientiousness, the Faking condition ( $M = 4.13, SD = .63$ ) was significantly greater than the Honest condition ( $M = 3.74, SD = .66$ ),  $p = .002, d = .60$ . The comparison between the Detection/Consequence ( $M = 3.83, SD = .54$ ) and

Faking conditions for Conscientiousness,  $p = .026$ ,  $d = -.51$ , was excluded from interpretation because significance did not reach the .01 alpha level.

The comparison between the Detection/Consequence ( $M = 3.83$ ,  $SD = .54$ ) and Faking ( $M = 4.07$ ,  $SD = .49$ ) conditions for Agreeableness,  $p = .042$ ,  $d = -.47$ , was likewise excluded from interpretation.

I tested my second hypothesis, which hypothesized no differences between groups, by examining the effect sizes between the warned conditions and the Honest condition. For the Reasoning-Honest difference, Extraversion had a small effect size,  $d = .09$ . For Conscientiousness, however, the effect size was slightly larger,  $d = .34$ , and for Neuroticism, there was a significant difference between means, with a large effect size,  $d = .62$ . For the Detection/Consequence-Honest difference, Neuroticism had a moderate effect size,  $d = .41$ , but Extraversion and Conscientiousness had quite small effect sizes,  $d = .15$  and  $d = .15$ . Overall, this hypothesis was supported for the Detection/Consequence-Honest comparison, but there was not much support for the Reasoning-Honest comparison. Although effect sizes of magnitude .41 and .34 are large enough to be considered meaningful, the means did not differ significantly, so the differences could have been due to error.

For the Student sample, a one-way ANOVA showed no significant difference between condition means for Neuroticism,  $F(3,1) = .61$ ,  $p = .612$ , power = .174, Extraversion,  $F(3,1) = 1.36$ ,  $p = .256$ , power = .357, Openness,  $F(3,1) = .88$ ,  $p = .452$ , power = .239, Agreeableness,  $F(3,1) = .23$ ,  $p = .879$ , power = .092, nor Conscientiousness,  $F(3,1) = .78$ ,  $p = .509$ , power = .214. As none of the one-way ANOVAs was significant, no post-hoc tests were performed.

Table 7

*Descriptive and ANOVA statistics with Tukey's post hoc tests for the personality variables: Nurse Sample*

Variable by Condition	$F(3, 1)^*$	$p$	N	M	SD	Tukey's PHST*
Neuroticism	11.09	.000				
Reasoning			65	2.32 <sup>a</sup>	.73	a = .001
Detection/Consequence			67	2.49 <sup>b</sup>	.63	b = .008
Faking			69	2.12 <sup>bc</sup>	.64	c = .000
Honest			66	2.77 <sup>ac</sup>	.73	
Extraversion	3.81	.011				
Reasoning			65	3.61	.68	a = .007
Detection/Consequence			67	3.47 <sup>a</sup>	.68	
Faking			69	3.83 <sup>a</sup>	.62	
Honest			66	3.57	.65	
Openness	1.51	.212				
Reasoning			65	3.48	.50	ns
Detection/Consequence			67	3.35	.49	
Faking			69	3.49	.52	
Honest			66	3.53	.52	
Agreeableness	2.69	.047				
Reasoning			65	3.91	.54	a = .042 (ns)
Detection/Consequence			67	3.83 <sup>a</sup>	.54	
Faking			69	4.07 <sup>a</sup>	.49	
Honest			66	3.87	.54	
Conscientiousness	4.95	.002				
Reasoning			65	3.96	.64	a = .026 (ns)
Detection/Consequence			67	3.83 <sup>a</sup>	.54	b = .002
Faking			69	4.13 <sup>ab</sup>	.63	
Honest			66	3.74 <sup>b</sup>	.66	

\* Significance for the one-way ANOVAs are at the .05 alpha level. Significance for the Tukey's post-hoc test is at the .01 alpha level.

### *Procedural Justice*

I hypothesized that participant perceptions of procedural justice within this study would be greater for participants in the Reasoning condition than the Detection/Consequence condition. This hypothesis was not supported.

I conducted independent samples t-tests to examine responses for differences between participant perceptions of procedural justice based on mean scores on Colquitt's (2001) PJQ and Bauer et al.'s (2001) overall SPJS, as well as the individual subscales included in the SPJS (Job Relatedness – Predictive, Chance to Perform, Propriety of Questions, and Job Relatedness – Content). For Nurses, the mean perception of procedural justice did not differ between the Reasoning and Detection/Consequence conditions for any of the justice measures or subscales (*M*, *SD* and t-test statistics can be found in Table 8). Similarly, no differences between means were found for Students either (see Table 9). Contrary to my hypothesis, the mean perception of procedural justice did not differ between the Reasoning and Detection/Consequence conditions for any of the justice measures or subscales for neither Nurses nor Students.

Table 8

*Descriptive and t-test statistics for the procedural justice variables: Nurse sample*

Variable by Condition	<i>t</i> *	df	<i>p</i>	N	M	SD
Colquitt's PJQ	-.15	128	.879			
Reasoning				64	3.63	.84
Detection/Consequence				66	3.65	.75
Bauer et al.'s SPJP	.51	129	.614			
Reasoning				64	2.66	.69
Detection/Consequence				67	2.60	.70
Job-Relatedness Predictive	.11	129	.911			
Reasoning				64	2.27	1.00
Detection/Consequence				67	2.25	.97
Perform	-.37	129	.711			
Reasoning				64	2.13	.92
Detection/Consequence				67	2.18	.86
Propriety	1.29	129	.201			
Reasoning				64	3.88	.90
Detection/Consequence				67	3.67	.95
Job-Relatedness Content	.72	129	.470			
Reasoning				64	2.29	.99
Detection/Consequence				67	2.16	.98

\* Significance for the independent samples t-tests are at the .05 alpha level.

Table 9

*Descriptive and t-test statistics for the procedural justice variables: Student sample*

Variable by Condition	t*	df	p	N	M	SD
Colquitt's PJQ	.71	73	.483			
Reasoning				38	3.63	.73
Detection/Consequence				37	3.51	.73
Bauer et al's SPJP	.49	73	.626			
Reasoning				38	2.59	.65
Detection/Consequence				37	2.52	.63
Job-Relatedness Predictive	.71	73	.481			
Reasoning				38	2.51	.86
Detection/Consequence				37	2.36	.95
Perform	-.06	73	.949			
Reasoning				38	2.22	.86
Detection/Consequence				37	2.23	.84
Propriety	-.17	73	.869			
Reasoning				38	3.48	.79
Detection/Consequence				37	3.51	.83
Job-Relatedness Content	1.47	73	.146			
Reasoning				38	2.09	1.06
Detection/Consequence				37	1.77	.82

\* Significance for the independent samples t-tests are at the .05 alpha level.

*Impression Management*

Using an ANOVA, I also found no significant mean differences for the Impression Management scale for either the Nursing sample,  $F(3, 1) = 2.33, p = .075$ , nor the Student sample,  $F(3, 1) = 1.81, p = .148$ .

## Chapter Four

### Discussion

#### *Personality*

I hypothesized that responses from the Faking condition would show higher mean scores for Extraversion, Openness, Agreeableness, and Conscientiousness than the Honest, Detection/Consequence, and Reasoning conditions, and lower mean scores for Neuroticism. I also hypothesized that the Honest condition would show no difference in mean scores from the Reasoning warning condition nor the Detection/Consequence condition. Since the ANOVAs for the student sample showed no differences between any means, the following discussion will refer to the Nursing sample results, unless otherwise specified.

My first hypothesis was partially for the Faking and Detection/Consequence comparison. The Faking condition elicited higher mean scores than the Detection/Consequence warning for Extraversion, and lower mean scores for Neuroticism. Also, there is some support for the Faking and Honest comparison. The Faking condition elicited higher mean scores than the Honest condition for Conscientiousness, and lower for Neuroticism. The second part of my hypotheses (that the Honest condition would show no significant difference in mean scores from the warned conditions) was supported for all but the Reasoning warning condition in Neuroticism, which was significantly higher than the honest mean. Examination of effect sizes revealed mostly small or negligible effects for both warnings, although the

Detection/Consequence-Honest difference approached a moderate effect size even though this difference was not statistically significant.

Consistent with previous research, these results suggest that the Detection/Consequence warning is an effective method for reducing applicant mean scores on a personality test within the job selection context because means were frequently lower than the Faking condition and did not differ from the Honest condition. The Reasoning warning, however, appears to be less effective. In fact, for Neuroticism, there is a significant difference between the Reasoning warning condition and the Honest condition, but no difference between the Reasoning condition and the Faking condition. Evidence thus suggests that a warning that the test administrators can identify those who distort their responses and threatening them with removing the applicants from the selection process if they do fake their responses is needed to significantly reduce applicant scores.

Dwight and Donovan (2003) pointed out that there are other possible reasons mean scores are decreased in warned conditions compared to faking conditions. One reason they mentioned was that respondents could overcompensate in their responses so as to avoid being identified as a faker. My results do not support this alternative explanation for the effect of warnings on personality scores. Since I included both faking and honest conditions in my study, I was able to show that a detection/consequence warning prompt does result in mean scores significantly lower than faked scores, and also that these lower scores did not differ from honest scores.

Unfortunately, the more applicant friendly warning was not as effective at reducing applicant faking behavior as a detection/consequence warning in this study's

context. It is possible that utilizing a reasoning warning in the selection context actually undermined the attempt to decrease applicant faking. The warning indirectly informs respondents that they have a choice when responding to be honest or distort their responses. This could have made participants who had not considered distorting their responses aware that they could, resulting in an increase in faking behavior for some respondents. Rees and Metcalfe (2003) found that a significant proportion of respondents said they had no problem faking during the application process, so informing applicants they have the opportunity to fake could result in increased faking behavior.

It is important to note that Pace et al. (2006) originally created this warning for use in a classification context in which applicants know they have a job, unlike this selection context. In that context, this warning may be more convincing to applicants. Accordingly, the reasoning warning might be used as a second-tier tool where organizations use the personality test's results to place the applicant in a best-fit job.

### *Procedural Justice*

I hypothesized that applicants in the Reasoning condition would perceive greater procedural justice than participants in the Detection/Consequence condition because of the more applicant-friendly tone of the Reasoning warning. My study shows no evidence of differences in perceptions of procedural justice between warning conditions. Perhaps the experimental manipulation did not replicate the selection context sufficiently to have participants form opinions about the fairness of the selection procedures.

Similarly, applicants may not form attitudes about the selection process until after the decision to hire is made. I modified my procedural justice measures to focus on the selection process such that no knowledge of the outcome was necessary to respond to the

measures. If applicants do not form a strong opinion about the procedures used when applying for a job until the outcome is known, this study's procedural justice measure could have missed the variance that would have been affected by utilizing different warnings.

A final possibility is that despite McFarland's (2003) findings, warnings on personality tests do not have any effect on applicant attitudes within the selection context. Rather, merely including a personality test in the selection battery could harm perceptions of procedural justice. I re-examined this study's data to look for differences between any of the four conditions examined, but found no differences between conditions for any of the procedural justice variables. Like McFarland's findings, participants in this study did report low perceptions of procedural justice (based on variable means, see Table 2), but of course, this is merely an observation, and more intensive research on applicant perceptions of personality tests should be performed.

#### *Impression Management*

I found no differences in self-reported impression management between the study's four conditions. This could be because the measure was not affected by the instruction prompts, which was to be expected. Participants were informed that the test for selection ended after they completed the personality test. As such, applicants likely did not respond to the impression management measure in a job selection mindset. It is also possible that impression management is more stable than some other personality traits so that a person who is prone to impression management would be unaffected by different testing contexts. I want to note that my Nursing sample demonstrated quite a bit of resistance to answering this part of the survey. Some wrote comments on their

response surveys and the experimenter overheard several respondents say they thought the items were too invasive and inappropriate. Many left items throughout the scale blank. From a methodological standpoint, the testing site could have been an issue. Since Nurse participants took the survey in their place of work, self monitoring could have affected responses such that those who completed the measure responded favorably, rather than honestly.

### *Limitations*

The main limitation of this study is the limited generalizeability of results due to the context of the test administration. As already mentioned, since the surveys in this study were not part of a true selection process, participant's responses may not reflect those that would be obtained in a high-stakes selection setting. However, the several significant mean differences between the Honest and Faking conditions of this study do show a difference in responses consistent with what previous research has shown between honest and faked responses, suggesting the results do simulate a real applicant setting, to an extent.

Another limitation relates to my Nursing sample. It is possible that the study's warning prompts could have different effects on non-nursing samples, such as those working in finance, marketing, or government, for example. The lack of significant results within the Student sample could have been due to industry-specific impacts of the warnings. Since the student sample was composed of participants working in a variety of fields as well as some unemployed participants, there could have been too much noise to pick up on the effects of the warning prompts. I reanalyzed the student sample responses by respondents' industry of employment (for those who were employed), but found no

significant mean comparisons. However, the small sample size of each condition within each industry ( $N < 11$ ) limited the possibility of finding significant differences. Thus, this study's results should be cautiously interpreted and applied outside the selection context for nurses.

Statistical power was also a concern in this study. Post hoc power tests showed Neuroticism, Extraversion, and Conscientiousness had sufficient power to detect significant differences (power  $> .80$ ), but Openness to Experience and Agreeableness had low power for the nursing sample. Power was very low for all personality variables in the student sample, leaving open the possibility that I missed effects that were in the data. Reliability for all the measures used was acceptable, so increasing my sample size would be the only remaining way to improve power. Due to organizational constraints, I was limited in the number of participants I could include in my Nursing sample. For the Student sample, however, I obtained data for 100 more participants, without any changes in my substantive findings nor much improvement in power.

#### *Future Research*

It would be interesting to see what effect the Detection/Consequence and Reasoning warnings would have in a real-world, high-stakes selection context. All four conditions could easily be inserted into a selection situation so applicants believed their responses would affect their chances of getting hired (but of course none could count in the selection process to ensure equitable treatment of applicants). Previous research included incentives within the study in an attempt to artificially create a selection context (the better applicants would receive money, for example). Awarding a financial incentive to the best performers in this study could have affected my findings. It is important to

remember, however, that the Honest and Faking manipulations did work as expected, so it is possible this study successfully simulated a selection context. An examination of the effect of financial or other incentives on results in a number of contexts would, however, benefit research on the contextual influences to responses to non-cognitive measures used for selection. Future research should also examine the effectiveness of a reasoning warning in reducing faking behavior in a classification context.

Byle and Holtgraves (2008) found that the type of design used had an effect on the magnitude of faking, with the between-subjects design having a smaller magnitude of faking than a within-subjects design when compared to honest scores. A within-subjects design where respondents complete two or more conditions would compliment this study in a number of ways. First, a within-subjects design would improve the likelihood of detecting significant experimental effects by increasing statistical power. Second, this design would permit control of pre-existing trait levels on the personality variables included, which I was unable to do. Third, within-subjects replication would examine whether the effects seen in this study represent a lower-bound for the effects of the Detection/Consequence warning, as well as further examine the effect of the Reasoning warning on responding.

In order to better assess why I found no significant differences between group means for the measures of procedural justice, it would be beneficial to expand this study to include a cognitive test. With a cognitive test (or some other component of a selection battery), there could be a condition in which the applicant takes no personality test. Such a design would enable the researcher to examine if it is merely the presence of a

personality test that affects the applicant's perception of procedural justice, as in McFarland (2003).

### *Conclusion*

This study answered Dwight et al.'s (2003) call for researchers to explore the effect of the integration of warnings into the selection process, providing three main additions to the personality literature. The central purposes of this study were to examine if the reasoning warning would be as effective as the Detection/Consequence warning in reducing applicant faking behavior, as well as examine if the Reasoning warning was more applicant-friendly in terms of procedural justice. The study adds to the wide range of research on applicant faking behavior by being the first to compare a detection/consequence warning to both faked and honest responses. Furthermore, it was the first to compare the effectiveness of the reasoning warning to both faked and honest responses, as well as responses obtained under a detection/consequence warning. Perhaps researchers have not studied alternative warnings to reduce applicant faking behavior because the Detection/Consequence warning is quite effective (at least within the selection context). In this study, a reasoning warning did not elicit mean trait scores that were significantly less socially desirable than faked responses. The reasoning warning appears to be inappropriate within this selection context, but still may show promise for other contexts (e.g. classification). Finally, my results suggest that explaining to respondents the purpose for including a personality test in the selection battery, as done through a reasoning warning, has no impact on respondent perceptions of procedural justice.

## References

- Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 22*(1), 1-26.
- 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*, 261-272.
- Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M. A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology, 54*, 387-419.
- Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. L. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment, 14*(4), 317-330.
- Byle, K. A., & Holtgraves, T. M. (2008). Integrity testing, personality, and design: Interpreting the Personnel Reaction Blank. *Journal of Business Psychology, 22*, 287-295.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2<sup>nd</sup> ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Colquitt, J.A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*, 386-400.
- Costa, P.T., & McCrae, R. R. (1992). *NEO PI-R professional manual*. Florida: Psychological Assessment Resources Inc.

- 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.
- Donovan, J. J., Dwight, S. A., & Schneider, D. (2008, April). Faking in the real world: Evidence from a field study. Paper presented at the 23<sup>rd</sup> annual conference of the Society of Industrial and Organizational Psychology. San Francisco, CA.
- Dwight, S. A., & Donovan, J. J. (2003). Do warnings not to fake reduce faking? *Human Performance, 16*(1), 1-23.
- Ellingson, J. E., Sackett, P. R., & Hough, L. M. (1999). Social desirability corrections in personality measurement: Issues of applicant comparison and construct validity. *Journal of Applied Psychology, 84*(2), 155-166.
- Gilliland, S. W. (1993). The perceived fairness of selection systems: An organizational justice perspective. *Academy of Management Review, 18*, 694-734.
- Gilliland, S. W. (1994). Effects of procedural and distributive justice on reactions to a selection system. *Journal of Applied Psychology, 79*, 691-701.
- Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.
- Hogan, J., Barrett, P., & Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology, 92*(5), 1270-1285.

- Holtz, B. C., Ployhart, R. E., & Dominguez, A. (2005). Testing the rules of justice: The effects of frame-of-reference and pre-test validity information on personality test responses and test perceptions. *International Journal of Selection and Assessment, 13*(1), 75-86.
- 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 and constructs and the effect of response distortion on those validities. *Journal of Applied Psychology, 75*, 581-595.
- Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9*(1/2), 152-194.
- Hunthausen, J. M., Truxillo, D. M., Bauer, T. N., & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology, 88*(3), 545-551.
- Jones, A., & Abraham, J. (2008, April). Assessing personality scores in applicant settings: A DIF analysis. Paper presented at the 23<sup>rd</sup> annual conference of the Society of Industrial and Organizational Psychology. San Francisco, CA.
- Judge, T. A., Higgins, C. A., Thoreson, C. J., & Barrick, M. R. (1999). The big five personality traits, general mental ability, and career success across the life span. *Personnel Psychology, 52*, 621-652.

- Komar, S., Brown, D. J., Komar, J. A., & Robie, C. (2008). Faking and the validity of conscientiousness: A Monte Carlo investigation. *Journal of Applied Psychology, 93*(1), 140-154.
- Leventhal, G. S. (1980). What should be done with equity theory? New approaches to the study of fairness in social relationship. In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research: 27-55*. New York: Plenum.
- Lind, E. A., & Tyler, T. (1988). *The social psychology of procedural justice*. New York: Plenum.
- Marshall, M. B., de Fruyt, F., Rolland, J-P., & Bagby, R. M. (2005). Socially desirable responding and the factorial stability of the NEO PI-R. *Psychological Assessment, 17*(3), 379-384.
- McFarland, L. A. (2003). Warning against faking on a personality test: Effects on applicant reactions and personality test scores. *International Journal of Selection and Assessment, 11*(4), 265-276.
- McFarland, L. A., & Ryan, A. M. (2000). Variance in faking across noncognitive measures. *Journal of Applied Psychology, 85*(5), 812-821.
- Mount, M. K., & Barrick, M. R. (1995). The big five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resource Management, 13*, 152-200.
- Mount, M. K., Witt, A., & Barrick, M. R. (2000). Incremental validity of empirically-keyed biographical scales over GMA and the big five personality constructs. *Personnel Psychology, 53*, 299-323.

- Mueller-Hanson, R., Heggstad, E. D., & Thornton, G. C. (2003). Faking and selection: Considering the use of personality from select-in and select-out perspectives. *Journal of Applied Psychology, 88*, 348-355.
- 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- 670.
- Pace, V. L., & Borman, W. C. (2006). The use of warning to discourage faking on noncognitive inventories. In R. L. Griffith & M. H. Peterson (Eds.), *A closer examination of applicant faking behavior* (pp. 283-304). Greenwich, Connecticut: Information Age Publishing.
- Pace, V. L., Borman, W. C., Penney, L. M., Xu, X., & Bearden, R. M. (2005, July). *Examination of the effectiveness of warnings for deterring faking on the navy recruiter assessment battery (Technical Report No. 506)*. Tampa, FL: Personnel Decisions Research Institutes, Inc. (PDRI).
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46*, 598-609.
- Rees, C. J., & Metcalfe, B. (2003). The faking of personality questionnaire results: Who's kidding whom? *Journal of Managerial Psychology, 18*, 156-165.
- Rosse, J. G., Miller, J. L., & Stecher, M. D. (1994). A field study of job applicants' reactions to personality and cognitive ability testing. *Journal of Applied Psychology, 79*(6), 987-992.

- 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*, 634-644.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*(2), 262-274.
- 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*, 966-974.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology, 80*(5), 607-620.
- Smith, D. B., Hanges, P. J., & Dickson, M. W. (2001). Personnel selection and the Five-Factor Model: Reexamining the effects of applicant's frame of reference. *Journal of Applied Psychology, 86*(2), 304-315.
- Stark, S., Chernyshekno, O. S., Chan, K.-Y., Lee, W. C., & Drasgow, F. (2001). Effects of the testing situation on item responding: Cause for concern. *Journal of Applied Psychology, 86*(5), 943-953.
- Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology, 81*(2), 134-141.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44*, 703-742.

- Topping, G. D. & O’Gorman, J. G. (1997). Effects of faking set on validity of the NEO FFI. *Personality and Individual Differences, 23(1)*, 117-124.
- Wolford, K., & Christainsen, N. D. (2008, April). Effects of self-coaching on faking of personality tests. Paper presented at the 23<sup>rd</sup> annual conference of the Society of Industrial and Organizational Psychology. San Francisco, CA.
- Wrenn, K., & Maurer, T. (2008, April). Extraversion as a predictor of procedural justice perceptions. Paper presented at the 23<sup>rd</sup> annual conference of the Society of Industrial and Organizational Psychology. San Francisco, CA.

## Appendices

## Appendix A

### International Personality Item Pool Items\*

1. Often feel blue
2. Feel comfortable around people
3. Am not interested in abstract ideas
4. Have a sharp tongue
5. Am always prepared
6. Dislike myself
7. Have little to say
8. Believe in the importance of art
9. Cut others to pieces
10. Pay attention to details
11. Rarely get irritated
12. Make friends easily
13. Do not like art
14. Have a good word for everyone
15. Waste my time
16. Am often down in the dumps
17. Keep in the background
18. Have a vivid imagination
19. Believe that others have good intentions
20. Find it difficult to get down to work
21. Have frequent mood swings
22. Am skilled in handling social situations
23. Avoid philosophical discussions
24. Suspect hidden motives in others
25. Get chores done right away
26. Seldom feel blue
27. Would describe myself as somewhat dull
28. Tend to vote for liberal political candidates
29. Get back at others
30. Carry out my plans
31. Feel comfortable with myself
32. Am the life of the party
33. Do not enjoy going to art museums
34. Respect others
35. Do just enough work to get by
36. Panic easily
37. Don't like to draw attention to myself

Appendix A (Continued)

- 38. Carry the conversation to a higher level
- 39. Accept people as they are
- 40. Don't see things through
- 41. Am not easily bothered by things
- 42. Know how to captivate people
- 43. Tend to vote for conservative political candidates
- 44. Insult people
- 45. Make plans and stick to them
- 46. Am very pleased with myself
- 47. Don't talk a lot
- 48. Enjoy hearing new ideas
- 49. Make people feel at ease
- 50. Shirk my duties

\* Goldberg et al. (2006)

## Appendix B

### Procedural Justice Questionnaire Items\*

1. Have you been able to express your views and feelings during these procedures?
2. Have you had an influence over the outcome arrived at by these procedures?
3. Have these procedures been applied consistently?
4. Have these procedures been free of bias?
5. Have these procedures been based on accurate information?
6. Do you feel able to appeal the outcome arrived at by these procedures?
7. Have these procedures upheld ethical and moral standards?

\* adapted from Colquitt (2001)

## Appendix C

### Selection Procedural Justice Scale Items\*

1. Doing well on this test means a person can do the nursing job well.
2. A person who scored well on this test will be a good nurse.
3. I could really show my skills and abilities through this test.
4. This test allowed me to show what my job skills are.
5. This test gives applicants the opportunity to show what they can really do.
6. I was able to show what I can do on this test.
7. The content of the test did not appear to be prejudiced.
8. The test itself did not seem too personal or private.
9. The content of the test seemed appropriate.
10. It would be clear to anyone that this test is related to the nursing job.
11. The content of the test was clearly related to the nursing job.

\* adapted from Bauer at al. (2001)

## Appendix D

### Impression Management Scale\*

1. I sometimes tell lies if I have to
2. I never cover up my mistakes
3. There have been occasions when I have taken advantage of someone
4. I never swear
5. I sometimes try to get even rather than forgive and forget
6. I always obey laws, even if I'm unlikely to get caught
7. I have said something bad about a friend behind his or her back
8. When I hear people talking privately, I avoid listening
9. I have received too much change from a salesperson without telling him or her
10. I always declare everything at customs
11. When I was young I sometimes stole things
12. I have never dropped litter on the street
13. I sometimes drive faster than the speed limit
14. I never read sexy books or magazines
15. I have done things that I don't tell other people about
16. I never take things that don't belong to me
17. I have taken sick-leave from work or school even though I wasn't really sick
18. I have never damaged a library book or store merchandise without reporting it
19. I have some pretty awful habits
20. I don't gossip about other people's habits

\* Paulhus (1991)