July 2018

Assessing and Preventing Applicant Faking on Personality Tests

Justin Feeney
The University of Western Ontario

Supervisor
Goffin, R.D.
The University of Western Ontario

Graduate Program in Psychology

A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of Philosophy

© Justin Feeney 2018

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Recommended Citation
https://ir.lib.uwo.ca/etd/5424

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact tadam@uwo.ca, wlswadmin@uwo.ca.
Abstract

This dissertation examined the measurement and prevention of applicant faking on personality tests. Study 1 compared how 12 different faking indices differentiated between the same people’s personality scores during a job application and non-applicant condition, and how these faking indices discriminated between separate groups of job applicants and non-applicants. We found that researchers and practitioners should assess applicant faking using multiple faking indices, including Idiosyncratic Item Responding, Blatant Extreme Responding, and Communal Impression Management, and when possible, Residualized Change Scores. Interestingly, Bogus Items—a common faking measure in the literature—were ineffective. Similarly, two covariance indices, were also ineffective and failed to predict applicant faking or relate to any other measure of applicant faking. The findings also suggest that we can disregard a few measures from future analysis because they are inferior to existing options, and offer no utility, which include Percent Agreement, Individual Change Scores, Within-Subject Correlations, and Within-Subject Variances of the Differences. Studies 2 and 3 examined the efficacy of three new faking dissuasion messages in reducing applicant faking compared to a traditional faking warning and a no-warning control group. In Study 2, we found no evidence for any of the dissuasion messages, including one adapted from the existing literature. In Study 3, we tested military recruits from the Canadian Armed Forces. We found some evidence that an Immediate Authentication Warning, which informed recruits that faking could be identified using reference checks and internal integrity checks, helped reduce faking. However, none of the other faking dissuasion messages were effective relative to the control group. We discuss several potential explanations for these conflicted findings.
Keywords

Applicant Faking; Deception; Personality Testing; Faking Warnings; Social Desirability
Co-Authorship Statement

Richard Goffin
Advisor and professor in Industrial/Organizational Psychology. Second author of the manuscripts in Chapters 1, 2, and 3.

Colin Kemp
Defense Scientist at DGMPRA in Ottawa, ON. Third author of the manuscript in Chapter 3.

Kabir Daljeet
Graduate student in Industrial/Organizational Psychology. Fourth author of the manuscript in Chapter 1 and fourth author of the manuscript in Chapter 2.

Shadi Beshai
Professor in Clinical Psychology at the University of Regina. Third author of the manuscript in Chapter 1 and third author of the manuscript in Chapter 2.
Acknowledgments

I would like to thank my supervisor, Richard Goffin, for his continued guidance throughout my graduate education. He has served as an outstanding mentor, helping me hone my skills in writing, research, and professionalism. I would also like to thank my collaborators. Defense scientist Colin Kemp provided access to the military recruits used in Study 2 and helped with the data collection. Kabir Daljeet helped with the tedious calculations and data analysis in Study 1. Shadi Beshai provided invaluable feedback and writing suggestions for Study 1 and helpful comments for Studies 2 and 3. I owe a special gratitude to Sarah Ridley, my fiancé, for her support and proofreading throughout the final stages of my PhD. Finally, I would like to thank my family for always being there throughout my many years of academic training.
Table of Contents

Abstract ........................................................................................................... i
Co-Authorship Statement .................................................................................. iii
Acknowledgments ........................................................................................... iv
Table of Contents ............................................................................................ v
List of Tables .................................................................................................... viii
Chapter 1 .......................................................................................................... 1
  1. General Overview ...................................................................................... 1
     1.1 Detecting and Assessing Applicant Faking ............................................. 4
     1.2 Determining How to Assess Applicant Faking ........................................ 9
     1.3 Preventing Applicant Faking: Faking Dissuasion Messages ................. 10
     1.4 Improving Faking Dissuasion Messages ............................................... 11
     1.5 References .......................................................................................... 13
Chapter 2 .......................................................................................................... 21
  2. Comparing the Efficacy of 12 Applicant Faking Indices in Discriminating between Job
     Applicant and Non-Applicant Personality Scores ....................................... 21
     2.1 Abstract ............................................................................................... 21
     2.2 Introduction .......................................................................................... 21
        2.2.1 Aims of the Present Study ............................................................... 23
        2.2.2 Positivity Bias: Social Desirability and Overclaiming ................. 23
        2.2.3 Response Patterns: Single Administration ..................................... 25
        2.2.4 Response Patterns: Multiple Administrations ............................... 27
        2.2.5 Differentiating Applicant and Non-Applicant Personality Scores .. 31
        2.2.6 Determining Faking Assessments for Future Research and Practice 32
     2.3 Method ................................................................................................. 33
        2.3.1 Participants and Procedure ............................................................. 33
        2.3.2 Measures ....................................................................................... 35
     2.4 Results ................................................................................................. 43
        2.4.1 Preliminary Analysis ..................................................................... 43
        2.4.2 Within-Subjects Faking Sensitivity ............................................... 45
        2.4.3 Between-Subjects Faking Sensitivity .............................................. 48
        2.4.4 Uninterested Applicants ................................................................. 50
        2.4.5 Supplemental Analysis: Covariance Indices ................................... 50
     2.5 Discussion ............................................................................................. 54
        2.5.1 Limitations .................................................................................... 58
        2.5.2 Conclusion ..................................................................................... 59
     2.6 References ............................................................................................. 60
Appendix A: Job Posting ................................................................................... 69
Appendix B: Bogus Item Scale ......................................................................... 71
Appendix C: Idiosyncratic Item Responding Scale .......................................... 72
Appendix D: Main Analyses by Faking Dissuasion Message Condition ............ 73
Chapter 3 .......................................................................................................... 83
  3. Faking Dissuasion Messages in the Applicant Screening Context ................ 83
     3.1 Abstract ............................................................................................... 83
3.2 Introduction ........................................................................................................... 84
  3.2.1 Present Study: Improving Faking Dissuasion Messages ................................ 85
  3.2.2 The Role of Verifiability .............................................................................. 88
  3.2.3 Applicant Reactions ..................................................................................... 89
3.3 Method .................................................................................................................. 90
  3.3.1 Participants ............................................................................................... 90
  3.3.2 Procedure ................................................................................................. 91
  3.3.3 Measures .................................................................................................... 92
3.4 Results .................................................................................................................. 95
  3.4.1 Preliminary Analyses: Job Manipulation Check ........................................... 96
  3.4.2 Preliminary Analyses: Did Applicants Read the Faking Dissuasion Messages? 96
  3.4.3 Main Analyses: Faking Dissuasion Message Efficacy ................................ 98
  3.4.4 Main Analyses: Faking Dissuasion Message Efficacy for Verifiable Items .... 98
  3.4.5 Main Analyses: Procedural Justice Perceptions and Test-Taking Anxiety ...... 99
3.5 Discussion ............................................................................................................. 102
  3.5.1 Limitations ............................................................................................... 108
  3.5.2 Conclusion ............................................................................................... 109
3.6 References .......................................................................................................... 110
Appendix A: Traditional Faking Warning ................................................................. 117
Appendix B: Immediate Authentication Warning .................................................... 118
Appendix C: Future Authentication Warning ........................................................... 120
Appendix D: Moral Suasion Appeal ....................................................................... 122
Appendix E: Job Posting ......................................................................................... 123
Chapter 4 .................................................................................................................. 125
4 Combatting Faking on Personality Tests during Military Recruitment .................. 125
  4.1 Abstract .......................................................................................................... 125
  4.2 Introduction ..................................................................................................... 126
    4.2.1 Selecting Military Recruits: Prevalence of Faking ...................................... 127
    4.2.2 Reducing Recruitment Faking .................................................................. 128
    4.2.3 Recruit Reactions to Faking Dissuasion Messages .................................... 131
  4.3 Method ............................................................................................................. 132
    4.3.1 Participants ............................................................................................. 132
    4.3.2 Measures ............................................................................................... 132
    4.3.3 Procedure ............................................................................................... 135
  4.4 Results .............................................................................................................. 137
    4.4.1 Preliminary Analyses .............................................................................. 137
    4.4.2 Faking Dissuasion Message Efficacy on OCEAN.20 .............................. 138
    4.4.3 Faking Dissuasion Message Efficacy on IPIP ...................................... 141
    4.4.4 Faking Dissuasion Message Efficacy on Faking Measures .................... 141
    4.4.5 Faking Dissuasion Message Efficacy on Recruit Reactions .................... 143
  4.5 Discussion ....................................................................................................... 143
    4.5.1 Limitations ............................................................................................. 149
    4.5.2 Implications and Conclusions ................................................................. 150
  4.6 References ...................................................................................................... 152
Appendix A: Immediate Authentication Warning: PowerPoint Presentation ............. 158
Appendix B: Future Authentication Warning: PowerPoint Presentation .................. 162
<table>
<thead>
<tr>
<th>Appendix C: Moral Suasion Appeal: PowerPoint Presentation</th>
<th>167</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix D: Traditional Faking Warning: PowerPoint Presentation</td>
<td>172</td>
</tr>
<tr>
<td>Appendix E: No Faking Warning: PowerPoint Presentation</td>
<td>177</td>
</tr>
<tr>
<td>Appendix F: Idiosyncratic Item Responding Scale</td>
<td>181</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>182</td>
</tr>
<tr>
<td>5. General Discussion</td>
<td>182</td>
</tr>
<tr>
<td>5.1. Future Directions in Faking Research</td>
<td>186</td>
</tr>
<tr>
<td>5.2. References</td>
<td>188</td>
</tr>
<tr>
<td>Curriculum Vitae</td>
<td>191</td>
</tr>
</tbody>
</table>
List of Tables

Table 1. Pairwise Comparisons of Personality Scores by Job Interest in Time 1 and Time 2 ..... 44
Table 2. Interested Applicants: Descriptive Statistics, Bivariate Correlations, and Internal Consistencies.......................................................................................................................... 46
Table 3. Within-Subject Faking Sensitivity.............................................................................. 49
Table 4. Between-Subjects Faking Sensitivity ........................................................................ 51
Table 5. Uninterested Applicants: Descriptive Statistics, Bivariate Correlations, and Internal Consistencies.................................................................................................................................. 52
Table 6. Pairwise Comparisons of Personality Scores by Job Interest in Time 1 and Time 2 ..... 97
Table 7. Faking by Job Interest and Faking dissuasion message condition.......................... 100
Table 8. Applicant Reactions by Job Interest and Faking dissuasion message condition ...... 101
Table 9. Review of Faking Dissuasion Message Studies since 2003 ...................................... 105
Table 10. Recruit Scores on the Trait Self-Descriptive Personality Inventory during Initial Recruitment.................................................................................................................................................. 139
Table 11. Recruit Scores on the OCEAN.20 Personality Inventory during the Application Simulation......................................................................................................................................... 140
Table 12. Recruit Scores on the IPIP Personality Inventory during the Application Simulation 142
Table 13. Recruit Scores on the Faking Measures during the Application Simulation .......... 144
Table 14. Recruit Reactions Scores to the Application Simulation........................................ 145
Chapter 1

1. General Overview

Personality is defined as a consistent pattern of thoughts, feelings, and behaviors. Assessment of personality has become common practice in pre-hire screening in organizational settings (Mischel, 2004; Rothstein & Goffin, 2006). A survey of 84 large organizations—with an average annual revenue of 45.7 billion dollars—found that 66% of the organizations used personality testing for “high potential” employees, and 57% used personality testing for senior executives (Church & Rotolo, 2013). Personality testing is now the second most common pre-employment assessment practice (Fallaw & Kantrowitz, 2013). Further, 74% of US-based HR professionals believe that personality testing can be useful (Fallaw & Kantrowitz, 2013; Society of Human Resource Management, 2011).

Personality testing is used by many organizations because scores from such testing predict the three pillars of job performance: (a) task performance, behaviors related to job-related tasks; (b) organizational citizenship behaviors, behaviors that help support the organization but that are outside of the immediate job-role; and (c) counterproductive work behaviors, the extent that an employee interferes with the performance of colleagues or their organization. If used correctly, personality tests predict between 9% and 16% of task performance (Tett, Jackson, & Rothstein, 1991), up to 13% of organizational citizenship behaviors (Borman & Motowidlo, 1997), and up to 40% of counterproductive behaviors such as theft, time fraud, or bullying (Marcus & Schuler, 2004).

The most favorable response to personality test items are often transparent to test-takers. For example, imagine an applicant had to respond to the following personality items from the International Personality Item Pool (Goldberg et al., 2006; Maples, Guan, Carter, & Miller,
2014): “I start tasks right away,” “I remain calm under pressure,” and “I love to help others.” For each item, it is obvious that answering “strongly agree” will appear more favorable to employers than answering “strongly disagree.” This is problematic to hiring organizations because applicants have incentive to answer dishonestly or to “fake” (Griffith, Chmielowski, & Yoshita, 2007). Faking is defined as deliberate provision of “inaccurate responses to personality items in a manner that they believe will increase their chances of obtaining valued outcomes, such as a favorable hiring decision” (Goffin & Boyd, 2009, p. 151). Between 30 and 50% of people endorse more favorable personality scores when applying for jobs in comparison to non-applicant self-reporting (Griffith et al., 2007), and up to 63% of applicants admit to “faking” on personality tests (Dwight & Donovan, 2003). Faking distorts the rank order of applicants and undermines hiring decisions (Goffin & Christiansen, 2003; Griffith et al., 2007). Fakers rise to the top of the job applicant pool, and as a result, companies hire them. This is especially the case for competitive job postings, where only a small number of people are hired (Donovan, Dwight, & Schneider, 2013). Thus, the successful applicants may not possess the traits they claim to have, and in turn, displace those that may be better suited for the position.

At first, these findings may seem counterintuitive. If applicants are faking, one may think that all applicants will increase their scores, and in turn, we should see mean shifts instead of rank order changes between honest and job application situations (J. Hogan, Barrett, & Hogan, 2007; R. Hogan, 2005). However, there are large individual differences in why and how applicants fake. Faking exists on a continuum (Donovan, Dwight, & Hurtz, 2003): some people respond honestly, answering the same way across situations; others fake a little, minimizing their negative traits; and still, others fake a lot, engaging in total dissimulation. These differences exist because applicants vary in their motivation to fake and perceived ability to fake, which are
influenced by a variety of factors including moral values and need to fake (Goffin & Boyd, 2009). This is problematic because the least honest applicants rise to the top of the applicant pool, and those who answer honestly or engage in lower levels of faking are severely disadvantaged (Griffith et al., 2007; Mueller-Hanson, Heggestad, & Thornton, 2003).

It is also important to qualify this review—faking does not always lead to more favorable scores. The change in scores is context dependent. For example, in clinical populations, some prospective patients may “fake bad”, making themselves look less favorable than their honest answers would portray (Holden & Book, 2011; Sullivan, 2001). Faking bad may lead to a clinical diagnosis that confers tangible benefits including workplace accommodations, insurance benefits, or government social assistance. People may also fake bad when they apply for a work position they are not really interested in (Boss, König, & Melchers, 2015). Imagine someone is encouraged by their spouse to apply for a job in a city they do not want to move—they could apply but sabotage their application by faking bad. In Switzerland, where mandatory military service is required, recruits often fake bad on personality tests to be selected out of the military (Boss et al., 2015).

Additionally, some people genuinely believe they have more desirable traits than they really demonstrate. In the social desirability literature (Paulhus, 1984, 1991), scholars differentiate between Impression Management, conscious dissimulation to appear more favorable, and Self-Deception, when the person believes what they are reporting. The purpose of this dissertation is to investigate motivated and intentional dissimulation where job applications attempt to appear more favorably to their prospective employers. As a result, faking bad and self-deception are not addressed in this dissertation.
In the pre-employment context, faking diminishes the validity of personality tests in hiring scenarios, which restricts the utility of test scores to make meaningful hiring predictions. For example, faking impairs convergent validity, as self-peer correlations are lower during job applications in comparison to responding in non-application scenarios (Robie, Taggar, & Brown, 2009). This suggests that applicant scores are less objectively verifiable, and in turn, should have less utility in predicting performance. Faking also diminishes the test-retest reliability of personality tests by introducing additional error variance into measurement (Hausknecht, 2010; Landers, Sackett, & Tuzinski, 2011). This suggests that applicant personality scores are unstable, and in turn, should have less utility in predicting important outcomes compared to non-applicant scores. In support of this, Peterson et al. (2011) found that faking weakens the criterion-related validity of personality tests in predicting counterproductive work behavior (M. H. Peterson, Griffith, Isaacson, O’Connell, & Mangos, 2011). As a result, faking can lead to hiring less optimal applicants, which is costly for hiring organizations.

To successfully protect organizations from applicants who fake, we need to appropriately assess applicant faking, and subsequently prevent it. In the next section, I review the most common faking assessments, discuss why further study is necessary, and describe how I plan to improve faking assessment. Once I identify the best faking assessments to capture variance related to applicant faking, I will use them to validate an innovative approach to dissuading applicants from faking.

1.1 Detecting and Assessing Applicant Faking

There are two broad ways to detect faking on personality tests: covertly assessing faking and assessing faking by analyzing inconsistencies or unlikely response patterns. The oldest, and most prevalent, approach is to covertly measure faking during the application process using a
Social Desirability Scale (Griffith & Peterson, 2008). In the job application context, Social Desirability Scales are used to “indicate a deliberate choice to alter responses to appear more socially favorable.” (Ellingson, Sackett, & Hough, 1999, p. 156). Social Desirability Scales measure how often applicants endorse improbable—but desirable—items such as “I’ve never forgotten to return something”, “I never swear”, and “I’ve never broken the law” (Paulhus, 1984). The use of a Social Desirability Scale in hiring assumes that when job applicants score favorably on social desirability items, job applicants are also faking on the personality test (Griffith & Peterson, 2008). There are several concerns with Social Desirability Scales. For example, scores from such scales tend not to correlate highly with scores from other faking assessments (Burns & Christiansen, 2006; Ellingson et al., 1999; Griffith & Peterson, 2008; Kuncel & Tellegen, 2009) and are confounded with legitimate personality traits such as extraversion (Uziel, 2010). Most detrimentally, items on Social Desirability Scales are obvious to test takers, and so are easy to fake (Alex, 1965).

In another attempt to covertly assess faking, researchers investigated if they could predict faking on personality tests by assessing the tendency to claim knowledge about fictitious topics. Presumably, applicants who mispresent their knowledge to appear more desirable are also faking on personality tests (Bing, Kluemper, Davison, Taylor, & Novicevic, 2011; Fan et al., 2012). There are two main approaches to assessing knowledge about fictitious phenomena: Bogus Items and Overclaiming. Bogus Items ask participants to endorse knowledge of things that do not exist, but sound job-related. For example, for a customer service position, applicants could be asked about their familiarity with point-of-sale machines (real) and customer first-order logic (fictitious). A more nuanced version of Bogus Items is Overclaiming general knowledge about broad topics including history, social science, and physics. Applicants rate their familiarity with
things that exist and things that do not. Statistically controlling for Overclaiming on fictitious things improves the criterion related validity of personality when predicting academic performance (Bing et al., 2011); however, and despite these promising results, Overclaiming fails to converge with other faking assessments in experimental settings (Feeney & Goffin, 2015), and is also confounded by legitimate personality traits (Dunlop et al., 2016).

The other main approach to assess applicant faking is to scrutinize applicants’ response patterns for inconsistencies. There are three measures of response patterns that can be used in employment settings as an index of faking: Idiosyncratic Item Responding, Covariance Indices, and Blatant Extreme Responding. Idiosyncratic items are determined by finding personality items that have different score distributions when filled out as an applicant than when completed as a non-applicant (Kuncel & Borneman, 2007; Kuncel & Tellegen, 2009), and creating a scale score across these idiosyncratic items. In theory, applicants who score more favorably on items that distinguish applicants from honest responders are more likely to be faking. The second approach is to examine the covariance between conceptually distinct personality items—such as “I work long hours at work” and “I enjoy being the life of the party”—but that are desirable to the target job (Burns & Christiansen, 2011). The underlying logic is that if an applicant is faking, the covariance between conceptually distinct but job-related traits—such as conscientiousness or extraversion—should increase when compared to non-applicants. This is because the faker is trying to appear favorable to the employer and will likely improve their scores on dissimilar items that are job-related. Finally, Blatant Extreme Responding is calculated by examining mean scores on favorable personality trait—such as conscientiousness or emotional stability (Birkeland, Manson, Kisamore, Brannick, & Smith, 2006)—where extreme scores may indicate faking (Landers et al., 2011).
Like social desirability scales and overclaiming, these faking indices are not without their conceptual concerns and methodological weaknesses (Burns & Christiansen, 2011). Chiefly, each index only captures improbable sets of response patterns (Burns & Christiansen, 2011), which can lead to misclassification of fakers as honest and honest responders as fakers. For example, it is possible for an applicant to be genuinely emotionally stable and conscientious, and therefore, have high covariance between the two measures. It would be equally possible for any given applicant to score high on idiosyncratic items for either trait, as well as have an extreme response profile. As a result, organizations may filter out potential high performers who are responding honestly, and, unintentionally, create concerns about the legal defensibility of their hiring practices (Catano, Wiesner, & Hackett, 2010).

There are also faking indices that examine consistency across personality testing sessions. Researchers can best use these indices when they have a set of the same people’s job applicant personality scores and non-applicant personality scores, where presumably, they are answering more honestly. The most intuitive and common faking index is to examine Individual Change Scores (Burns & Christiansen, 2011; Dilchert & Ones, 2011; McFarland, 2003), which assess the degree of change between job application and non-application conditions. The more an applicant changes their score in a favorable direction, the more the applicant is faking. A limitation of this faking index is that it is confounded by legitimate personality traits, where, for example, a difference score may be derived both from faking and the personality score itself (Burns & Christiansen, 2011). A modified approach is to calculate Residualized Change Scores, which controls for non-applicant personality scores, theoretically providing a pure index of applicant faking (Burns & Christiansen, 2011; Feeney & Goffin, 2015). This index is not confounded by legitimate personality scores, but its mean scores are difficult to interpret.
There are also three uncommon faking indices that have been used or suggesting by researchers. The first is Percent Agreement, which examines the number of items that an applicant answers exactly the same across two testing sessions (Gordon & Gross, 1978), such as answering “agree” to “I work hard” in both test sessions. This faking index is simple to calculate and gives a categorical score for each personality item—the applicant response is consistent in both conditions or not; however, Percent Agreement scores are insensitive to the magnitude of the difference. Furthermore, Percent Agreement scores do not differentiate between extreme faking, such as where an applicant answers strongly agree and strongly disagree to the same item across test sessions, or less severe faking, such as when an applicant switches from answering strongly agree to agree. As a result, the Percent Agreement faking index artificially restricts variance relating to applicant faking (Burns & Christiansen, 2011), limiting its ability to serve as an effective faking index. The second uncommon faking index is Within-Subject Correlations (Lautenschlager, 1986), which examines the correlations between applicant responses to personality items in job applicant and non-applicant conditions. This faking index is sensitive to changes in an applicant’s rank-order between the two response conditions, which is practical, because rank order changes influence hiring decisions (Goffin & Christiansen, 2003). However, this faking index is likely to demonstrate particularly poor reliability when there are small sample sizes, as small changes in scores could produce larges changes in rank order (Burns & Christiansen, 2011).

Finally, researches have also suggested in addition to Individual Change Scores, faking scholars can also assess the variance of those change scores. This faking index, known as Within-Subject Variance of Difference Scores, examines an applicant’s consistency when answering personality items across sessions (Burns & Christiansen, 2011). However, this
approach can be insensitive to mean shifts, because applicants can consistently shift their scores to be more favorable by a small amount, and have a small faking score. By contrast, someone else could have high variance in their difference scores, but not significantly improve their mean personality score. As a result, this faking index may be more prone to conflating reliability with faking than are other measures.

The common goal of these faking indices is that they attempt to identify people who exhibit substantial differences in personality between the job applicant and non-applicant conditions; however, it is unlikely that an organization would be able to obtain both job applicant and non-applicant responses from job applicants. Therefore, these faking indices may be better suited for the study of faking as a theoretical construct and to validate more practical faking assessments that can be used in employment settings (Feeney & Goffin, 2015).

1.2 Determining How to Assess Applicant Faking

In the first study in this dissertation, I address a gap in the current literature by examining the 12 reviewed faking assessments in a single investigation. To do so, I examine which faking assessments are best able to distinguish between job applicant and non-applicant personality scores, and to determine the best faking assessment for research and practice. Researchers often use faking assessments interchangeably, as they theoretically assess the same underlying construct; however, this idea remains untested with a motivated sample of job applicants (Bensch, Paulhus, Stankov, & Ziegler, 2017). Currently, researchers and practitioners employ different assessments to examine faking. This means that practitioners often use a single faking assessment (Griffith & Peterson, 2008), such as a Social Desirability Scale, potentially misclassifying applicants, and reducing the accuracy of hiring decisions (Donovan et al., 2013).
Researchers often use different faking assessments to study the same phenomenon because they use different study designs, yielding contradictory conclusions.

Thus, there is a clear need to examine the extent that different faking assessments measure the same underlying construct—applicant faking. If a faking assessment is deficient, it should be abandoned. If no single faking assessment stands out, then faking scholars may need more than one assessment to triangulate around the construct of faking to boost the credibility of classification. To address this limitation, I presented participants with a potential job opportunity, and investigated how well each faking assessment differentiated between people who were interested in the job and those who were not. I also compared how well each faking assessment discriminated between the same person’s personality scores as an applicant and as a non-applicant. Findings from Study 1 informed how to best assess applicant faking in Studies 2 and 3.

1.3 Preventing Applicant Faking: Faking Dissuasion Messages

In contrast to detecting faking, some scholars have developed techniques to prevent faking (Dilchert & Ones, 2011). Research suggests that, in general, faking dissuasion messages are among the most promising faking prevention techniques (Pace & Borman, 2006). Faking dissuasion messages are embedded into the instructions of personality tests and can take multiple forms. The most common method is to craft threatening faking dissuasion messages that warn of removal from the hiring process. But, faking dissuasion messages can also be written to gently persuade applicants not to fake, such as reminding applicants about the immorality of applicant faking (Goffin & Boyd, 2009; Pace & Borman, 2006; Uruena & Robie, 2011).

The most common faking dissuasion message is the faking warning, where best practice is to inform applicants that faking can be detected and will lead to immediate disqualification
from the selection process (Dilchert & Ones, 2011; Donovan et al., 2003; Pace & Borman, 2006). This approach can reduce faking by 30 to 50% (Converse et al., 2008; Dwight & Donovan, 2003; Fan et al., 2012; Goffin & Woods, 1995; Kuroyama, Wright, Manson, & Sablynski, 2010; Landers et al., 2011; McFarland, 2003; Robson, Jones, & Abraham, 2008; Vasilopoulos, Cucina, & McElreath, 2005), can reduce applicant faking in both applied and laboratory situations (Dwight & Donovan, 2003), and is inexpensive and easy to employ (Dilchert & Ones, 2011).

Despite these advantages, faking dissuasion messages suffer from a few limitations. First, they can create negative applicant reactions (Converse et al., 2008; Fan et al., 2012; McFarland, 2003). Faking dissuasion messages, by definition, alert applicants that their competition may fake, and may subsequently decrease the face validity of the personality tests (Fan et al., 2012). Negative applicant reactions should matter to organizations because they can reduce the likelihood that an applicant accepts a job offer or recommends your organization to others (McCarthy, Bauer, Truxillo, Anderson, et al., 2017). Second, faking dissuasion messages may be more effective for applicants with lower general mental ability than for applicants with higher general mental ability, partially confounding personality scores with variance from general mental ability (Vasilopoulos et al., 2005). Finally, applicants still fake after reading faking dissuasion messages—they just fake less often (Dwight & Donovan, 2003). Thus, it may be beneficial to improve faking dissuasion messages by using theory from other branches of psychology.

1.4 Improving Faking Dissuasion Messages

To improve the potency of faking dissuasion messages, I incorporate accountability theory (Lerner & Tetlock, 1999; Tetlock, Skitka, & Boettger, 1989) to more effectively guide
applicants to consider the possibility of being caught for faking. I also leveraged morality theory (Haidt, 2001) to better dissuade faking by appealing to their moral values about faking. The goal was to devise faking dissuasion messages that are more effective and reduce negative applicant reactions. In Study 2, I examined these new faking dissuasion messages using job applicants for a customer service position. It is important to acknowledge that Studies 1 and 2 use the same data. The two studies investigate different questions and it would be infeasible to write them up as a single publication. As result, they are presented separately. In study 3, I replicated the same faking dissuasion messages in a military setting to examine if the messages can reduce faking across organizational settings.
1.5 References


doi:10.1177/1073191117700268


doi:10.1207/s15327043hup1002_3


Chapter 2

2. Comparing the Efficacy of 12 Applicant Faking Indices in Discriminating between Job Applicant and Non-Applicant Personality Scores

2.1. Abstract

This study examined the assessment of applicant faking on personality tests. We compared how 12 different faking indices differentiated between the same people’s personality scores during a job application and a non-applicant condition, and how these faking assessments discriminated between separate groups of job applicants and non-applicants. We recruited workers from Amazon’s Mechanical Turk to participate in a multi-session research study, where we then presented an ostensibly real job opportunity and asked if they were interested. We had 349 Interested Job Applicants and 169 participants who were uninterested in the position (“Uninterested Applicants”). After the first session, we informed applicants that the job was fictitious and invited them to complete the same personality measures one week later, as non-applicants. The manipulation was successful, as Interested Applicants received more desirable personality scores than Uninterested Applicants, but only when they believed there was a real job. Our results suggest that scholars and practitioners should use multiple faking assessments to properly detect faking, including Residualized Change Scores, Blatant Extreme Responding, Idiosyncratic Item Responding, and Communal Impression Management. Implications for future research are discussed.

2.2. Introduction

Personality tests are commonly used by employee as selection tools and have utility in predicting performance; however, applicant faking supresses their potential (Converse, Peterson, & Griffith, 2009). Two-thirds of large corporations utilize personality testing when selecting for
high-level positions and surveys suggest that personality is one of the most commonly assessed constructs (Fallaw & Kantrowitz, 2013). Personality tests are popular because their scores predict the three pillars of job performance: task performance, organizational citizenship behaviors, and counterproductive work behaviors (Borman & Motowidlo, 1997; Marcus & Schuler, 2004; Tett et al., 1991). One obstacle to personality testing is that job applicants are often able to identify favorable responses and answer accordingly, inflating their scores and probability of being hired (Converse et al., 2009; Griffith et al., 2007; Holden, Lambert, D’Agata, & Book, 2017). Faking obstructs the psychometric properties of personality tests in job applications, as it decreases reliability, convergent validity, and criterion-related validity (Landers et al., 2011; M. H. Peterson, Griffith, Isaacson, et al., 2011; Robson et al., 2008). Therefore, it is critical that hiring organizations be able to thwart applicant faking to improve the validity of their hiring decisions.

Researchers have tackled the issue of applicant faking on personality tests for nearly a century, yet there is no consensus on best practices to detect faking (Alex, 1965; Zickar & Gibby, 2006). Some hiring organizations use faking dissuasion messages to prevent faking, which are test instructions that attempt to dissuade applicants from faking. Although the literature is generally supportive of the use of dissuasion messages, there are some inconsistent findings with null or trivial effects (Dwight & Donovan, 2003). One potential explanation for discrepant findings is that the extant literature on faking dissuasion messages assess faking differently across studies, including: Blatant Extreme Responding (Boss et al., 2015; Landers et al., 2011; Vasilopoulos et al., 2005), Bogus Items (Dwight & Donovan, 2003; Fan et al., 2012), Residualized Change Scores (Feeney, Daljeet, & Goffin, 2017), and Social Desirability Scales.
(Fan et al., 2012; McFarland, 2003). How can organizations draw firm conclusions from the current literature?

2.2.1. Aims of the Present Study

There are many techniques to assess applicant faking (Burns & Christiansen, 2011; Griffith & Robie, 2013); however, studies examining how such techniques compare to each other are scarce and non-exhaustive. The purpose of the current study is to answer critical questions regarding the assessment of applicant faking including: (1) which faking assessments best discriminate between different people’s job applicant and non-applicant personality scores; (2) which faking assessments best discriminate between the same people’s applicant and non-applicant personality scores; and (3) which faking assessments should be advised for future research and practice. Accordingly, results of the current study may help researchers and practitioners decide on a single faking assessment or combination of assessments that can be used in practice. In the following sections, we review the most common faking assessments, and then review literature on how these assessments relate to each other. Researchers in the area often devote little attention to how their assessment of faking may influence their results (Griffith & Peterson, 2008).

2.2.2. Positivity Bias: Social Desirability and Overclaiming

The most common method to assess applicant faking is to measure Socially Desirable Responding (Griffith & Peterson, 2008). There are many Social Desirability Scales (Blasberg, Rogers, & Paulhus, 2013), but in the selection context, they are generally used to “indicate a deliberate choice to alter responses to appear more socially favorable” (Ellingson et al., 1999, p. 156). These scales are embedded within a personality test to capture the tendency to endorse unusual—but desirable—items such as “I’ve never broken the law” (Li & Bagger, 2006).
Theoretically, if an applicant endorses many of these items, then they are likely mispresenting themselves. Some test publishers integrate these scales into their personality measures and offer regression formulas to “correct” for faking (Goffin & Christiansen, 2003).

There are several concerns with social desirability scales. First and most detrimental, evidence suggests that Socially Desirable Responding does not converge with other faking assessments (Burns & Christiansen, 2006; Ellingson et al., 1999; Kuncel & Tellegen, 2009). As a result, Socially Desirable Scales may misclassify honest responders as fakers or the reverse. Secondly, these items are obvious to test takers, and evading them can be easily taught (Alex, 1965). Third, Socially Desirable Responding positively correlates with personality traits such as extraversion (Li & Bagger, 2006; Uziel, 2010), and in turn, removing high scorers or adjusting their scores may confound personality measurement.

**Bogus Items and Overclaiming.** To resolve the shortcomings of Social Desirability Scales, scholars have measured the extent that applicants overstate their knowledge of fictitious phenomena (Paulhus, Harms, Bruce, & Lysy, 2003). Within this broad approach, there are two specific approaches to measuring knowledge about fictitious things: Bogus Items and Overclaiming. Bogus Items require participants to endorse their familiarity with things that do not exist, which often sound job-related (Dwight & Donovan, 2003; Fan et al., 2012). For example, for a financial investment position, an applicant could be asked about their familiarity with “exchange traded funds” (real) or “reverse mutual funds” (fictitious). The second approach is “Overclaiming”, which is claiming to have general knowledge about broad topics such as books, poems, history, and science, to both legitimate and fictitious items. For example, an applicant may be asked about their knowledge of Napoleon (legitimate) or Queen Alberta (fictitious). Applicants rate their familiarity with real and fictitious phenomena, and the more
they endorse fictitious phenomena, the more likely they are faking (Bing et al., 2011). There are a variety of methods to calculate Overclaiming (see Paulhus et al., 2003), but a typical approach is to average familiarly scores on items that assess fictitious phenomena (Bing et al., 2011; Feeney & Goffin, 2015). Statistically controlling for Overclaiming strengthens the veracity of personality when predicting university grades (Bing et al., 2011).

2.2.3. Response Patterns: Single Administration

An alternate approach to assessing faking is to examine response patterns to personality items within a single test administration (Burns & Christiansen, 2011). These response patterns can be used as faking indices in real hiring scenarios and do not require burdening respondents with additional items. They are also invisible to test-takers and, in theory, should be more difficult for applicants to evade. Three of these approaches include: Idiosyncratic Item Responding, Covariance Indices, and Blatant Extreme Responding.

**Idiosyncratic Item Responding.** Idiosyncratic Item Responding reflects shifts in response patterns on items that differentiate fakers and honest responders. Idiosyncratic Item Responding scales are produced in three steps (Kuncel & Borneman, 2007; Kuncel & Tellegen, 2009). In the first step, the distribution of scores for each personality item are compared in non-applicant and job applicant samples, and items that best distinguish between the two samples are identified. In the second step, an idiosyncratic item responding scale is produced by including personality items that have large mean differences across the two test administrations. In the third and last step, scale scores for each applicant are calculated. Idiosyncratic Item Responding scores have been used to successfully distinguish between people instructed to “fake good” and honest applicants (Holden et al., 2017; Kuncel & Borneman, 2007; Kuncel & Tellegen, 2009); however, this approach has some conceptual and practical barriers. It is unclear if idiosyncratic
items generalize across samples. Each organization may need to generate their own Idiosyncratic Item Responding scales, which would require access to both non-applicant and job applicant personality scores.

**Covariance Index.** Another approach is to examine the covariance between items that assess conceptually distinct traits, but that are perceived as desirable to the target job (Burns & Christiansen, 2011; Christiansen, Robie, Burns, & Speer, 2017). The covariance between traits that are conceptually distinct and desirable should increase for fakers, but not those who are answering honestly. This is because fakers will increase their scores across traits to look more desirable to hiring organizations, which will increase the covariance between the traits. In our literature review, only one study used a Covariance Index to measure faking (Christiansen et al., 2017), but it is a measure that would be easy and practical to conduct in applied settings; however, this approach is also not without its limitations. It is easy to imagine an applicant who genuinely scored high on two conceptually distinct, yet job-related personality traits (e.g., Extraversion and Conscientiousness), and in turn, some of the best applicants may be incorrectly identified as faking.

**Blatant Extreme Responding.** Blatant Extreme Responding is calculated by examining the frequency of extreme answers—*Strongly Agree* or *Strongly Disagree*—throughout the personality test. The assumption is that many extreme answers reflects faking (Landers et al., 2011) because most people show variation in their responses under normal conditions (Levashina, Weekley, Roulin, & Hauck, 2014). Blatant Extreme Responding has evidence of construct validity (Levashina et al., 2014). Blatant Extreme Responding also has an intuitive appeal: this approach is designed to catch the most severe misrepresentation. Removing applicants with high scores on Blatant Extreme Responding is also unlikely to deflate the
criterion-related validity of personality testing. This is because extreme scores can be detrimental to performance (Le et al., 2011). Logical concerns with Blatant Extreme Responding are that it is insensitive—detecting only extreme fakers—and it is easy for skilled applicants to evade (Landers et al., 2011).

2.2.4. Response Patterns: Multiple Administrations

In contrast to the previously reviewed approaches, scholars have developed faking indices that can be used when they have both job applicant and non-applicant personality scores (Burns & Christiansen, 2011). These faking indices include: Individual Change Scores, Residualized Change Scores, Percent Agreement, Within-Subject Correlations, Within-Subject Variance Of The Differences. The unifying goal of these faking indices is to identify applicants who exhibit differences between responding as a job applicant and when responding as a non-applicant, where there is presumably less motivation to fake. That being said, it is unlikely that an organization would have both sets of scores in a real hiring scenario, and in turn, these faking indices are ideal for researchers to study faking as a theoretical construct, or to validate more practical techniques (Feeney & Goffin, 2015). However, faking indices derived from repeated measures designs of applicant faking are touted as being superior both conceptually and empirically when compared to faking indices derived from a single test administration (Burns & Christiansen, 2011; Dilchert, Ones, Viswesvaran, & Deller, 2006; Feeney & Goffin, 2015; Gordon & Gross, 1978; Lautenschlager, 1986).

Individual Change Scores. Perhaps the most intuitive repeated measures faking index is calculated by assessing the difference in personality scores when the same people complete the personality measure as a job applicant and as a non-applicant. If applicant scores are more favorable than non-applicant scores, than the difference is likely due to faking—with larger
differences reflecting more faking. Individual Change Scores have been used in many investigations and exhibit strong psychometric properties (Dilchert et al., 2006; Feeney & Goffin, 2015; McFarland & Ryan, 2000; M. H. Peterson, Griffith, Converse, & Gammon, 2011). The major obstacle with Individual Change Scores is that the score, at least in theory, can be a result of faking or legitimate personality scores (Burns & Christiansen, 2011). For example, a difference score may reflect an applicant’s tendency to fake, but it also reflects their “room” to fake—the extent that an applicant is able to make themselves look more favorable than their honest score would permit. That is, someone with a favorable non-applicant score may have limited capacity to improve their scores, whereas the reverse is true for someone with a naturally unfavorable score. As a result, difference scores have limited use in theoretical studies of applicant faking because their relationships with other variables may be due to either faking or the legitimate personality trait that underlies them.

**Residualized Change Scores.** Residuals are computed by individually regressing applicant responses to each personality item on the same personality item when completed in a non-applicant condition. This process removes variance associated with honest scores, and in theory, the residual reflects a purer assessment of faking than Individual Change Scores—reducing inferential errors (Burns & Christiansen, 2011). Residualized Change Scores are computed by averaging residuals across items. Thus, a high score reflects more faking than does a low score (Feeney & Goffin, 2015); however, the limitation of these scores is that they are ambiguous. For the same reason that Residualized Change Scores are uncontaminated by legitimate personality variance, they have no discernable unit of measurement, and it is impossible to draw meaningful comparisons between scores on distinct traits (Burns &
Christiansen, 2011). Residualized Change Scores are also still possibly contaminated by measurement error associated with the personality scale filled out during the job application.

**Percent Agreement.** As a measure of consistency across repeated administrations, we can examine the percent of answers that stay the same. For example, agreement is achieved if respondents answer “strongly agree” to “I like to keep things clean and tidy” as an applicant or non-applicant (Gordon & Gross, 1978). Percent Agreement is sensitive to all sources of measurement error and is also insensitive to the magnitude of the disagreement—outcomes for each item are dichotomous (agreement or no agreement). If an applicant answered “5” as an applicant, a non-applicant score of “1” or “4” would generate the same disagreement outcome. In our search, Percent Agreement was rarely used in the faking literature. Thus, Percent Agreement is a limited version of Individual Change Scores, that is unable to capture magnitude of change.

**Within-Subject Correlation.**

Another approach to examining applicant faking between multiple test administrations is to examine the correlation between same person’s job applicant and non-applicant scores on the same personality dimension, such as conscientiousness (Burns & Christiansen, 2011; Lautenschlager, 1986). If an applicant is faking during a job application, the rank order of their responses should vary from when the same applicant filled the same questions without motivation to fake. For example, a job applicant may answer strongly agree to "I complete tasks successfully," "I keep my promises", and "I work hard," because each item may look favorable to a job. However, as a non-applicant, the same person may answer disagree, strongly agree, and strongly disagree to the same items. Thus, the Within-Subject Correlation between the two sets of responses would be low. By contrast, if someone completed the same personality test twice as a non-applicant, their responses are likely to remain similar both times, resulting in higher
Within-Subject Correlations. Thus, Within-Subject Correlations are inversely related to faking—where lower scores are more indicative of applicant faking. The limitation of this index is that it is insensitive to mean shifts. If an applicant increases their scores uniformly across items, they may be undetected as a faker, and in turn, it may be best used when paired with other faking assessments. Additionally, Within-Subject Correlations will be unstable with small sample sizes or shorter personality inventories (Burns & Christiansen, 2011), which may be a limitation for practitioners.

**Within-Subject Variance of the Differences.** The Within-Subject Variance of the Differences faking index is similar to, but more sensitive, than Within Subject Correlations (Lautenschlager, 1986). This measure “relies on computing the variance of the difference scores across test administrations at the individual level” (Burns & Christiansen, 2011, p. 366). For example, two job applicants may answer strongly agree to "I complete tasks successfully," "I keep my promises", and "I work hard," because each item may look favorable to a job. However, as non-applicants, one person may strongly agree, neutral, agree, and the other person might answer strongly disagree, neutral, agree. In contrast to the within-subject correlation, we would examine the difference for each response and calculate the variance of those difference scores. In this example, the second applicant would be identified as faking more than the first applicant, because there is more variation in their difference scores. The benefit of this measure is it indexes change in both rank-order and overall variability (Burns & Christiansen, 2011). This measure has some evidence of construct validity, although our review found only one study that used it (Lautenschlager, 1986); however, this index is imperfect, as it also conflates faking and measurement error. In the next section, we develop our specific research questions.
2.2.5. Differentiating Applicant and Non-Applicant Personality Scores

The main purpose of this study is to determine which faking assessment(s) should be advised for use in research and practice. We accomplished this using two approaches. First, we sought to examine how each faking assessment predicts the same people’s applicant and non-applicant scores on each personality dimension. We call this approach “Within-Subjects Faking Sensitivity.” A faking index will have higher Within-Subjects Faking Sensitivity if there is a significant difference in its correlation with the same people’s applicant and non-applicant personality scores. A faking index will have lower Within-Subjects Faking Sensitivity if it correlates similarly with both job applicant and non-applicant personality scores. Therefore, Within-Subject Faking Sensitivity scores range on a continuum from 0 (no sensitivity) to 1 (high sensitivity). Within-Subjects Faking Sensitivity scores were calculated separately for each personality dimension and then averaged for a composite score. If a faking assessment has low Within-Subjects Faking Sensitivity, it will have limited utility in screening out applicants who successfully fake during a job application—those who increase their scores on desirable personality traits relative to their non-applicant scores.

One potential concern is that faking assessments that utilize repeated measures designs (i.e., applicant vs. mon-applicant assessments) should demonstrate higher Within-Subjects Faking Sensitivity than faking assessments that are derived from applicant scores only. We argue that this is a feature of the Within-Subjects Faking Sensitivity score, not a limitation or inherent bias. The current “gold standard” of applicant faking measures—Residualized Change Scores—take personality scores from a job application and remove all shared variance with personality scores during non-applicant conditions (Feeney & Goffin, 2015). This leaves a residual, that in theory, represents pure faking. This is precisely why faking scholars favor Residualized Change
Scores and it makes sense that Residualized Change Scores should demonstrate high Within-Subjects Faking Sensitivity. Moreover, Within-Subjects Faking Sensitivity can still be used to determine which faking assessments should be used in practical settings when repeated measures designs are infeasible. Being able to identify applicants who distort their scores during a job application is, ultimately, the overarching goal of assessing applicant faking (Donovan et al., 2013).

**Research Question 1:** To what extent will the faking indices demonstrate Within-Subjects Faking Sensitivity?

The second approach to assess the efficacy of the faking assessments was to examine the extent that they discriminate between job applicants and non-applicants. We call this “Between-Subjects Faking Sensitivity”, and calculated it by comparing mean scores on each faking assessment between job applicants and non-applicants and then calculated Cohen’s D effect size estimates (Christiansen et al., 2017). In theory, if a faking assessment is capturing faking, then job applicants, who are presumably faking more than non-applicants, should score significantly higher on that measure than non-applicants. The more effective the faking assessment, the higher the effect size should be.

**Research Question 2:** To what extent will the faking indices demonstrate Between-Subjects Faking Sensitivity?

2.2.6. Determining Faking Assessments for Future Research and Practice

Once we identify faking assessments that demonstrate both Within-Subjects Faking Sensitivity and Between-Subjects Faking Sensitivity, it is also important to examine how these faking assessments relate to each other. For example, if two faking assessments demonstrate high faking sensitivity, but do not correlate highly with one another, then both faking assessments
need to be used to detect applicant faking. By contrast, if two faking assessments demonstrate have similar validity, but are also highly correlated, then the assessment that is most convenient can be used. Another concern is that many of the most promising faking assessments in the extant literature require repeated measures designs, such as Individual Change Scores, which are not useful in applied settings (Burns & Christiansen, 2011). Therefore, it is important to find practical faking assessments that can be used to approximate them for practitioners who use personality testing for hiring decisions. As a result, we also examine the bivariate correlations between each of our faking assessments.

**Research Question 3:** Which Faking Assessments should be recommended for future research and practice?

### 2.3. Method

#### 2.3.1. Participants and Procedure

We recruited 779 people for a two-session study on Amazon’s Mechanical Turk (mTurk). Participants were paid $2 for their time, which is consistent with the mTurk average rate of $1.17 per hour (Mason & Suri, 2012). All participants were presented with a fictitious job opportunity for a remote customer service position that would enable them to work from home—consistent with the work they do on mTurk. After viewing the posting (see Appendix A), applicants indicated if they were interested in the position (“Interested Applicants”) or not (“Uninterested Applicants”). Interested Applicants were informed that successful candidates would be recruited for a video-conferenced job interview. We modelled the job posting from a real customer-service job posting and the Customer Service Representatives section of the Occupational Information Network (O*NET; Listing #43-4051.00; N. G. Peterson et al., 2001). From our initial sample, 572 applicants completed both sessions (73.4%). We dropped applicants who did not participate
in both sessions as we would not be able to calculate many of our faking indices for these applicants. We subsequently dropped and additional 38 applicants for failing careless responding checks.

The resulting sample included 357 Interested Applicants ($M_{age} = 35.10$, $SD = 10.64$). The majority were female ($n = 223$), employed full time ($n = 225$) or part-time ($n = 71$). Our Interested Applicants reported a modal income of $30,000 to $39,999 USD ($n = 69$), and a modal education of a four-year university degree ($n = 135$). We also had 177 Uninterested Applicants that were used in our analyses ($M_{age} = 34.14$, $SD = 11.82$). The majority were female ($n = 96$), employed full-time ($n = 131$), the modal income was $20,000 to $29,000 USD ($n = 24$), and the modal education ($n = 57$) was a four-year university degree. Interested and Uninterested Applicants reported similar demographics, although the modal income was slightly lower for Uninterested Applicants than for Interested Applicants.

The present study emerged from an interest in comparing how four different faking dissuasion messages were able to reduce the extent that applicants faked on the personality assessment. As a result, applicants were assigned at random to one of the five conditions: an immediate authentication message ($n = 90$), which explained that answers could be verified using internal integrity check and references; a future authentication warning ($n = 117$) that mentioned that personality items would be verified against future job behavior; a moral suasion prompt ($n = 103$) that reminded applicants that faking was morally wrong; a traditional warning ($n = 119$) that emphasised both the risk of being identified and disqualified from the selection process; and finally, a no-warning control group ($n = 105$). Applicants in the four dissuasion message conditions read their respective personality test instructions prior to completing personality measurement, whereas the control group read standard instructions without a warning. As shown
in our preliminary analyses below, there were no mean differences in personality scores and the conclusions from the present study were generally consistent across conditions. As a result, we proceeded with comparing how the different faking assessments measured applicant faking across conditions.

The personality measure included a Social Desirability Scale and careless responding checks. Applicants also rated their endorsement of Bogus Items. As a reminder, many of the faking assessments in our study are derived from the personality scores. Upon completion of the first session, applicants were informed that we were not working with a real company. We informed all applicants—interested and uninterested—that they would have the opportunity to take part in a second session in exchange for payment. Applicants were invited to the second survey one week later to reduce carryover effects, which has been used in previous faking research in repeated measures designs (M. H. Peterson, Griffith, Isaacson, et al., 2011). At the beginning of the second session, we reminded them that the employment opportunity was fictitious and to respond as honestly as possible. Similar to the first session, participants in the second session completed measures of personality, which included a Social Desirability Scale and careless responding checks. For all applicants, we refer to the sessions as “Time 1” and “Time 2”. For the Interested Applicants, Time 1 scores are equivalent to job applicant scores and Time 2 are equivalent to non-applicant scores. For Uninterested Applicants, Time 1 and Time 2 scores are both equivalent to non-applicant scores.

2.3.2. Measures

**Personality.** Personality was assessed using the International Personality Item Pool (IPIP; Goldberg et al., 2006), which examines: Agreeableness, Conscientiousness, Emotional Stability, Extraversion, And Openness. These dimensions were assessed using 120 items on a 5-
point response scale (from 1 = “Strongly Disagree” to 5 = “Strongly Agree”). Internal consistencies for the five dimensions of personality range between .87 and .90 on the 120-item IPIP (Maples et al., 2014). Additionally, all five dimensions on the 120-item scale were found to correlate meaningfully with respective NEO Personality Inventory dimension, with coefficients ranging between .87 and .90 (Maples et al., 2014). We acknowledge that there are many other personality measures used in research and practice, but we used the five-factor model for greater parsimony with the broader faking literature (Birkeland et al., 2006).

**Careless Responding.** Following best practices in measuring careless responding (Meade & Craig, 2012), we asked applicants to answer six verifiable items (e.g., how much they agree with “The sum of 2+2 is 10.”) that were dispersed throughout the personality test. We also asked our participants whether we should include their data in our final analyses, with a guarantee that they would receive payment regardless of their answer. These approaches are generally sufficient in detecting careless responders.

**Faking.**

As we will review, many of the faking assessments in this study were derived from personality scores, which can be computed for specific traits or across traits. For example, we could create an Idiosyncratic Item Responding scale that consists of just idiosyncratic conscientiousness items, or we could create an Idiosyncratic Item Responding scale with idiosyncratic items from all five personality dimensions. In theory, the former scale should have an advantage when predicting faking on just Conscientiousness, which would be an unfair comparison to faking measures such as Impression Management or Bogus Items that supposedly predict faking more generally. Thus, when personality scores were required, we used personality scores on all five personality traits to ensure that some measures were not given an unfair
advantage in predicting faking. However, if a practitioner is only using one or two personality dimensions for hiring, they may be interested in crafting a measure from only these dimensions.

**Individual Change Scores.** Individual Change Scores measure the extent that someone changes their scores across repeated sessions (Burns & Christiansen, 2011). Individual Change Scores can be calculated for separately for personality traits (Feeney & Goffin, 2015; McFarland & Ryan, 2006). Individual Change Scores also demonstrate high internal consistency and convergent validity with other faking assessments, including self-reported faking admission (Feeney & Goffin, 2015). For all applicants, we subtracted scores on each personality item during Time 2 from the same item’s score during Time 1. We would expect Individual Change Scores should be positively related to applicant faking.

**Residualized Change Scores.** Residualized Change Scores are similar to Individual Change Scores, except that variance related to legitimate personality traits are removed (Burns & Christiansen, 2011). For each applicant, we calculated Residualized Change Scores by regressing each personality item score during Time 1 on the same personality item score during Time 2, and saved the residuals. These residuals reflect Time 1 personality scores without variance associated with Time 2 personality scores. We then computed the mean across these residuals, where higher scores reflect faking. In theory, Residualized Change Scores are a “pure” faking assessment that is no longer conflated with legitimate personality variance (Feeney & Goffin, 2015). This allows faking scholars to examine how faking relates to other traits. If a faking scholar examines how Individual Change Scores relate to other variables, the relationship could be driven by faking or the personality variables used to calculate them. By contrast, relationships between RCS and other traits are not confounded by applicant personality.
Residualized Change Scores have a high internal consistency (alpha = .86) and have been used successfully in other investigations (Feeney & Goffin, 2015).

**Bogus Items.** We crafted six Bogus Items that sound related to the customer service job, which included: “customer first-order logic”, “remote method invocation”, “remote procedure call”, “inverse body language”, “DCM”, and “customer service intelligence test”. Knowledge of these items were measured using 5-point response scales (0 = “Very Unfamiliar”, 4 = “Very familiar”). These items were randomly mixed with a scale of 24 legitimate customer service terms that were derived from a customer service training book (Leland & Bailey, 2006), and foils were sought out on multiple search engines to ensure they were not real customer service terms. This process to create Bogus Items is consistent with the extant literature (Dwight & Donovan, 2003; Levashina, Morgeson, & Campion, 2009). The full scale is included in Appendix B.

**Social Desirability.** We measured Social Desirability using the Bi-Dimensional Impression Management Index (Blasberg et al., 2013). This scale assesses two distinct forms of impression management: agentic, which involves “exaggerating one’s social or intellectual status” (e.g., “I’m usually the one to come up with the big ideas”; and communal, which involves “denying socially deviant impulses and claiming pious attributes” (e.g., “I don’t gossip about other people’s business” (Blasberg et al., 2013, p. 523). Both Agentic Impression Management and Communal Impression Management subscales are assessed using 10 items measured on a 5-point response scales (1 = “Strongly Disagree” to 5 = “Strongly Agree”). Internal consistencies exceeded .70 across two studies, and there is evidence of both convergent and discriminant validity (Blasberg et al.). These items were embedded randomly throughout the Personality assessment.
**Idiosyncratic Item Responding.** Idiosyncratic personality items are those that differentiate job applicant responses from non-applicant responses (Kuncel & Borneman, 2007). Prior studies identified idiosyncratic items by visually comparing score distributions for the same items in job applicant and non-applicant samples (Kuncel & Borneman, 2007; Kuncel & Tellegen, 2009). We preferred an empirical approach to visual inspections for the detection of idiosyncratic items, because we found eyeballing differences were unreliable. To identify idiosyncratic items, we examined mean differences for each item during Time 1 in the interested applicant and uninterested applicant samples. We selected the items with the 10 largest mean differences for our Idiosyncratic Item Responding scale. Items and data are reported in Appendix C. In theory, higher Idiosyncratic Item Responding scores should reflect higher levels of faking.

One potential concern with this approach is that when a scale is created and used within the same sample, the scale will capitalize on sample specific variance (Cureton, 1950). This is problematic if a practitioner plans to use the same scale for future job applicant pools without first validating it. Before an Idiosyncratic Item Responding scale can be generalized to future applicant pools, it must follow appropriate scale creation guideline—developed in one sample and validated in a second sample (Hinkin, 1995, 1998). As a result, findings from the present study could potentially overestimate the efficacy of Idiosyncratic Item Responding.

**Blatant Extreme Responding.** We calculated Blatant Extreme Responding by summing the frequency that applicants endorsed the most favorable response (i.e., strongly agree or strongly disagree) on each conscientiousness item during Time 1. Thus, for each item, an applicant received a score of 1 (extreme) or 0 (not extreme), and we calculated the sum of those extreme responses. This approach differs from the Blatant Extreme Responding used by Landers et al. (2011), who assigned scores (0, .25, .5, .75, 1) for each level of the Likert scale (1 to 5,
respectively) and then calculated the sum. One limitation of the Landers approach is that it is a transformation of the mean, and in turn, is logically conflated with legitimate personality scores. The main utility of measuring faking is to differentiate between those who are genuinely high on a trait and those who are not, so companies know who to hire and who to screen out—the Landers approach fails to do this. Our approach measures the tendency to answer with the most extreme answer, rather than just achieving a high overall score. This should lessen—but not eliminate—the chief limitation of the Landers approach to calculating Blatant Extreme Responding, which may flag the very people a firm wants to hire as fakers. Thus, we counted the frequency of extremely favorable responses across all five personality dimensions, where higher scores should reflect more faking.

**Covariance Indices.** We calculated two covariance indices based on prior studies (Christiansen et al., 2017). The first, Covariance Index-Personality examines the covariance between personality items, during Time 1, that are desirable to the job, but are also uncorrelated under non-applicant conditions. In theory, higher scores on Covariance Index-Personality reflects more faking. Increased covariance between job-related personality items during a job application suggests that applicants are responding to the conceptually distinct personality items as if they are measuring the same construct.

To calculate Covariance Index-Personality, we conducted an expert judgment panel to determine the job-relatedness of our personality items across all traits. Our sample consisted of 15 experts in industrial and organizational psychology—familiar with both personality measurement and personnel selection: two recent doctoral graduates, nine doctoral students, and four master’s students. Each expert provided a set of job-relatedness ratings (5-point scale from 1 = “Not at all Important” to 5 = “Extremely Important”) and were paid $10 for their assistance.
For job relatedness, one expert was dropped for having unacceptably low inter-rater reliability ($r < .50$). The ICC (3, k; Shrout & Fleiss, 1979) for job-relatedness ratings was .96. We subsequently selected all personality items that had a mean job-relatedness score higher than four, which was equivalent to “Important” in the expert judgment survey.

We then searched for items that were generally uncorrelated (-.1 to .1) in our Uninterested Applicant sample. Since these applicants were uninterested in the job, there is less concern that their data are contaminated by the potential of a job offer. We chose a range of -.1 to .1 as correlations below.10 are less than Cohen’s guidelines for a small effect size, suggesting that there is no relationship between the variables (Cohen, 1992). Given that we were interested in reducing the potential that the job application influenced responses, we examined the covariance of personality scores during Time 2. We found a total of 35 uncorrelated item pairs. Once we identified our uncorrelated item pairs, we standardized Time 1 personality scores separately for Interested and Uninterested Applicants, computed the cross-products for all item-pairs, and calculated the average. This procedure produced in a composite Covariance Index-Personality for each applicant (as describd by Christiansen et al., 2017). As done in Christiansen et al. (2017), we also derived a second index, Covariance Index-Impression Management from our Agentic and Communal Impression Management items using the same guidelines. We crafted a list of 57 uncorrelated item pairs, standardized Impression Management scores during Time 1 separately for Interested Applicants and Uninterested Applicants, computed the cross-products for all item-pairs, and calculated the average. For both Covariance Index-Personality and Covariance Index-Impression Management, with higher scores reflecting more faking.

**Percent Agreement.** To measure Percent Agreement, we examined the number of personality item scores that remained consistent across Time 1 and Time 2 for both our
Interested Applicants and Uninterested Applicants, and calculated the average (Burns & Christiansen, 2011; Lautenschlager, 1986). Thus, for each pair of responses, we coded if the answer remained consistent (1) or inconsistent (0), summed the number of items that remained consistent, and divided it by the number of personality items. The final indicator suggests the percentage of responses to our personality measure that changed between the two times, where lower scores theoretically reflecting more faking.

**Within-Subject Correlation.** An additional approach to assessing applicant faking is to determine the extent that the rank order of responses remain consistent across Time 1 and Time 2 (Burns & Christiansen, 2011). We evaluated rank-order consistency for each applicant by calculating the correlation between their two sets of personality scores separately for each dimension (Time 1 and Time 2). Next, we calculated the average Within-Subject Correlation across the five personality dimensions to receive a final composite score. Thus, each applicant received a single score ranging from -1 to 1, where negative or smaller Within-Subjects Correlations should indicate more faking.

**Within-Subject Variance of the Differences.** The final approach we used to measure applicant faking is examining the Within-Subject Variance of the Differences (Burns & Christiansen, 2011; Lautenschlager, 1986). To calculate this measure, we summed the variance for each personality dimension during Time 1 and Time 2, and then subtracted the covariance for the same personality dimension’s scores between the two times. The result is a measure of rank-order consistency for each personality dimension that is sensitive to changes in variance between the two personality assessments. We then calculated the average of the Within-Subject Variance of the Differences scores across personality dimensions to derive a single composite score. Theoretically, higher scores should suggest more faking.
2.4. Results

2.4.1. Preliminary Analysis

One potential concern is that this dataset was part of an investigation that examined the efficacy of four different faking dissuasion messages on personality assessment when compared to a no-warning control group. The faking dissuasion messages were ineffective, and in turn, mean personality scores did not significantly differ across conditions. As a result, for this investigation, we aggregated across faking dissuasion message conditions to analyze the faking measures. There are no theoretical reasons to believe the efficacy of faking assessments should vary based on ineffective faking dissuasion messages. Still, to test for this potential confound, we conducted a MANOVA with faking dissuasion condition as the independent variable and each of our 12 faking measures as the dependent variables to ensure that faking assessment scores did not differ as a function of the warning condition. The MANOVA analysis was insignificant, suggesting there were no differences in faking across the faking dissuasion messages, $F(48, 1893.42) = .89, p = .70$, Wilk’s $\Lambda = .92$. In the interest of transparency, we report the main within-subjects and between-subjects faking sensitivity analyses for each condition (Appendix D).

Many of our faking assessments are derived from change in personality scores overtime and one potential concern with our study design is carryover effects. If our job application manipulation was successful, we would expect that Interested Applicants would receive significantly more favorable personality scores than Uninterested Applicants during Time 1, but not Time 2—when applicants were told that the job was fictitious. We conducted a repeated measures MANOVA with job interest as the between subjects, time as the within-subjects
Table 1. Pairwise Comparisons of Personality Scores by Job Interest in Time 1 and Time 2

<table>
<thead>
<tr>
<th>Personality Dimension</th>
<th>Time 1</th>
<th></th>
<th>Time 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Interested Applicants</td>
<td>Uninterested Applicants</td>
<td>d</td>
<td>Interested Applicants</td>
</tr>
<tr>
<td></td>
<td>$M$</td>
<td>$SE$</td>
<td>95% CI</td>
<td>$M$</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>4.30</td>
<td>0.03</td>
<td>[4.25, 4.36]</td>
<td>4.03</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>3.95</td>
<td>0.02</td>
<td>[3.90, 4.00]</td>
<td>3.75</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>3.87</td>
<td>0.04</td>
<td>[3.80, 3.94]</td>
<td>3.60</td>
</tr>
<tr>
<td>Openness</td>
<td>3.47</td>
<td>0.03</td>
<td>[3.42, 3.52]</td>
<td>3.49</td>
</tr>
<tr>
<td>Extraversion</td>
<td>3.65</td>
<td>0.03</td>
<td>[3.59, 3.71]</td>
<td>3.44</td>
</tr>
</tbody>
</table>

Note. This table includes the estimated marginal means, standard errors, and 95% confidence intervals for each personality dimension in Time 1 and in Time 2. Ns for Interested and Uninterested Applicants were 357 and 177 respectively. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means.
variable, and the Big Five personality dimensions scores as the dependent variables. There were significant multivariate effects for both job interest, \( F(5, 528) = 3.82, p = .002, \) Wilk’s \( \Lambda = .97. \) and time, \( F(5, 528) = 41.33, p < .001, \) Wilk’s \( \Lambda = .72. \) Main effects were qualified by a significant multivariate interaction, \( F(5, 528) = 10.75, p < .001, \) Wilk’s \( \Lambda = .91. \) Univariate tests of the interaction effects were significant for: Agreeableness, \( F(1, 532) = 20.84, p < .001, \) \( \eta_p^2 = .04; \) Conscientiousness, \( F(1, 532) = 40.48, p < .001, \) \( \eta_p^2 = .07; \) Extraversion, \( F(1, 532) = 27.30, p < .001, \) \( \eta_p^2 = .05; \) and Emotional Stability, \( F(1, 532) = 44.11, p < .001, \) \( \eta_p^2 = .07. \) There were no significant effects for Openness. This is consistent with the literature, as job applicants fake the least on Openness (Birkeland et al., 2006). Pairwise comparisons of the estimated marginal means compared Interested Applicants and Uninterested Applicants for each personality dimension at Times 1 and 2. Interested Applicants received significantly more favorable personality scores on Conscientiousness, Emotional Stability, Extraversion, and Agreeableness than Uninterested Applicants during Time 1, but not during Time 2. The results of the pairwise comparisons are reported in Table 1.

### 2.4.2. Within-Subjects Faking Sensitivity

To explore our first research question about the extent that each faking measure differentiates between our Interested Applicants Time 1 and Time 2 personality scores, we calculated a Within-Subjects Faking Sensitivity index separately for each personality dimension, by subtracting how well each faking index predicted personality scores during Time 2, from how well the same faking index predicted the same personality score during Time 1. Table 2 presents the means, standard deviations, bivariate correlations, internal consistencies, of our faking and personality measures for the Interested Applicant sample (\( n = 357 \)). We then conducted Fisher’s
<table>
<thead>
<tr>
<th>Measure</th>
<th>$M$</th>
<th>$SD$</th>
<th>Bivariate Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BER</td>
<td>8.62</td>
<td>4.93</td>
<td>(.87)</td>
</tr>
<tr>
<td>2. BIMI-A</td>
<td>2.92</td>
<td>.59</td>
<td>.45 (.73)</td>
</tr>
<tr>
<td>3. BIMI-C</td>
<td>3.27</td>
<td>.77</td>
<td>.55 (.80)</td>
</tr>
<tr>
<td>4. Bogus</td>
<td>15.48</td>
<td>5.85</td>
<td>.14 (.83)</td>
</tr>
<tr>
<td>5. CVI-P</td>
<td>0.17</td>
<td>0.46</td>
<td>.18 (.84)</td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td>0.13</td>
<td>0.33</td>
<td>.28 (.93)</td>
</tr>
<tr>
<td>7. ICS</td>
<td>0.26</td>
<td>0.32</td>
<td>.30 (.75)</td>
</tr>
<tr>
<td>8. IIR</td>
<td>4.07</td>
<td>0.62</td>
<td>.72 (.75)</td>
</tr>
<tr>
<td>9. PCT</td>
<td>0.56</td>
<td>0.14</td>
<td>.04 (.82)</td>
</tr>
<tr>
<td>10. RCS</td>
<td>0.06</td>
<td>0.28</td>
<td>.73 (.93)</td>
</tr>
<tr>
<td>11. WSC</td>
<td>0.43</td>
<td>0.15</td>
<td>.10 (.92)</td>
</tr>
<tr>
<td>12. WSVD</td>
<td>1.34</td>
<td>0.72</td>
<td>.51 (.92)</td>
</tr>
<tr>
<td>13. T1-C</td>
<td>4.30</td>
<td>0.51</td>
<td>.74 (.92)</td>
</tr>
<tr>
<td>14. T2-C</td>
<td>3.95</td>
<td>0.59</td>
<td>.48 (.92)</td>
</tr>
<tr>
<td>15. T1-A</td>
<td>3.95</td>
<td>0.43</td>
<td>.51 (.92)</td>
</tr>
<tr>
<td>16. T2-A</td>
<td>3.73</td>
<td>0.49</td>
<td>.29 (.92)</td>
</tr>
<tr>
<td>17. T1-N</td>
<td>3.87</td>
<td>0.67</td>
<td>.68 (.92)</td>
</tr>
<tr>
<td>18. T2-N</td>
<td>3.43</td>
<td>0.76</td>
<td>.37 (.92)</td>
</tr>
<tr>
<td>19. T1-O</td>
<td>3.47</td>
<td>0.52</td>
<td>.28 (.92)</td>
</tr>
<tr>
<td>20. T2-O</td>
<td>3.45</td>
<td>0.57</td>
<td>.25 (.92)</td>
</tr>
<tr>
<td>21. T1-E</td>
<td>3.65</td>
<td>0.56</td>
<td>.60 (.92)</td>
</tr>
<tr>
<td>22. T2-E</td>
<td>3.38</td>
<td>0.64</td>
<td>.32 (.92)</td>
</tr>
</tbody>
</table>


Table 2. Interested Applicants: Descriptive Statistics, Bivariate Correlations, and Internal Consistencies
Table 2 Continued.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Bivariate Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>14</td>
</tr>
<tr>
<td>1. BER</td>
<td></td>
</tr>
<tr>
<td>2. BIMI-A</td>
<td></td>
</tr>
<tr>
<td>3. BIMI-C</td>
<td></td>
</tr>
<tr>
<td>4. Bogus</td>
<td></td>
</tr>
<tr>
<td>5. CVI-P</td>
<td></td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td></td>
</tr>
<tr>
<td>7. ICS</td>
<td></td>
</tr>
<tr>
<td>8. IIR</td>
<td></td>
</tr>
<tr>
<td>9. PCT</td>
<td></td>
</tr>
<tr>
<td>10. RCS</td>
<td></td>
</tr>
<tr>
<td>11. WSC</td>
<td></td>
</tr>
<tr>
<td>12. WSVD</td>
<td></td>
</tr>
<tr>
<td>13. T1-C</td>
<td></td>
</tr>
<tr>
<td>14. T2-C</td>
<td></td>
</tr>
<tr>
<td>15. T1-A</td>
<td>.25</td>
</tr>
<tr>
<td>16. T2-A</td>
<td>.40</td>
</tr>
<tr>
<td>17. T1-N</td>
<td>.52</td>
</tr>
<tr>
<td>18. T2-N</td>
<td>.63</td>
</tr>
<tr>
<td>19. T1-O</td>
<td>-09</td>
</tr>
<tr>
<td>20. T2-O</td>
<td>.01</td>
</tr>
<tr>
<td>21. T1-E</td>
<td>.28</td>
</tr>
<tr>
<td>22. T2-E</td>
<td>.36</td>
</tr>
</tbody>
</table>
r-to-z transformations for dependent correlations to determine if these differences were statistically significant (Steiger, 1980).

The formula we used to calculate these scores were:

\[ \text{Within-Subjects Faking Sensitivity} = \frac{r_{t1}}{r_{t2}} (\text{Faking Index, Time 1 Personality}) * r_{t2} (\text{Faking Index, Time 2 Personality}) \]

We then averaged the scores across personality dimensions to develop a composite Within-Subjects Faking Sensitivity score. As shown in Table 3, both Individual Change Scores (Within-Subjects Faking Sensitivity = .54) and Residualized Change Scores (Within-Subjects Faking Sensitivity = .46) performed the best discriminating between the same interested applicant’s Time 1 and Time 2 personality scores. However, both Idiosyncratic Item Responding (Within-Subjects Faking Sensitivity = .21) and Blatant Extreme Responding (Within-Subjects Faking Sensitivity = .22) performed reasonably well, which is informative for practitioners because these measures can be calculated without non-applicant scores. The results suggest that Within-Subjects Variance of the Differences, Communal Impression Management, Within-Subjects Correlations, and Percent Agreement discriminated between interested applicant’s Time 1 and Time 2 personality scores. Bogus Items, Covariance Index-Personality, Covariance Index-Impression Management, and Agentic Impression Management all performed poorly relative to other faking assessments.

2.4.3. Between-Subjects Faking Sensitivity

An additional approach to assess how well faking assessments capture faking on personality tests is to compare mean faking scores between Interested Applicants and Uninterested Applicants during Time 1. If a faking assessment is truly capturing variance relating to faking, Interested Applicants should receive higher scores on that assessment than
## Table 3. Within-Subject Faking Sensitivity

<table>
<thead>
<tr>
<th>Faking Index</th>
<th>Within-Subjects Faking Sensitivity ($r_d = r_{t1} - r_{t2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>1. ICS</td>
<td>.64***</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.54***</td>
</tr>
<tr>
<td>3. WSVD</td>
<td>.24**</td>
</tr>
<tr>
<td>4. BER</td>
<td>.26***</td>
</tr>
<tr>
<td>5. IIR</td>
<td>.24***</td>
</tr>
<tr>
<td>6. BIMI-C</td>
<td>.25***</td>
</tr>
<tr>
<td>7. BIMI-A</td>
<td>.10</td>
</tr>
<tr>
<td>8. CVI-IM</td>
<td>.07</td>
</tr>
<tr>
<td>9. CVI-P</td>
<td>.01</td>
</tr>
<tr>
<td>10. Bogus</td>
<td>.02</td>
</tr>
<tr>
<td>11. WSC</td>
<td>-.28***</td>
</tr>
<tr>
<td>12. PCT</td>
<td>-.38***</td>
</tr>
</tbody>
</table>

Note. $N = 357$. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management. We tested for the difference between two dependent correlations for each of the personality dimensions using two-tailed r-to-z transformations.

***p < .001, **p < .01, *p < .05
would Uninterested Applicants. We conducted a MANOVA with job interest as the independent variable and the 12 faking assessments as the dependent variables. The MANOVA revealed a significant multivariate effect, $F(12, 494) = 9.76, p < .001$, Wilk’s $\Lambda = .81$. Interested Applicants scored significantly higher on Blatant Extreme Responding, Agentic Impression Management, Communal Impression Management, Bogus Items, Individual Change Scores, Idiosyncratic Item Responding, Residualized Change Scores, and Within-Subject Variance of the Differences than did Uninterested Applicants. Interested Applicants had significantly lower Percent Agreement and Within-Subjects Correlations than did Uninterested Applicants. Covariance Index-Personality and Covariance Index-Impression Management did not differ significantly between Interested and Uninterested Applicants. Means, standard deviations, and effect sizes are reported in Table 4. The effect sizes provide an index of Between-Subjects Faking Sensitivity.

2.4.4. Uninterested Applicants

Finally, in Table 5, we present the means, standard deviations, bivariate correlations, and internal consistencies for all faking indices and personality measures for our Uninterested Applicants.

2.4.5. Supplemental Analysis: Covariance Indices

An anomalous finding is that neither the Covariance Index-Personality or Covariance Index-Impression Management performed well in our study. Both measures showed limited Within-Subjects Faking Sensitivity and neither measure demonstrated significant Between-Subjects Faking Sensitivity. One major difference between Christiansen et al.’s (2017) study and the current one is that in their study, the researchers used a “fake good” experimental design with university students, and these samples tend to have considerably higher levels of faking
Table 4. Between-Subjects Faking Sensitivity

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Interested Applicants</th>
<th>Uninterested Applicants</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
</tr>
<tr>
<td>1. IIR</td>
<td>4.07</td>
<td>.04</td>
<td>[4, 4.14]</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.06</td>
<td>.01</td>
<td>[.03, .08]</td>
</tr>
<tr>
<td>3. BIMI-C</td>
<td>3.27</td>
<td>.04</td>
<td>[3.19, 3.35]</td>
</tr>
<tr>
<td>4. ICS</td>
<td>.26</td>
<td>.02</td>
<td>[.23, .29]</td>
</tr>
<tr>
<td>7. BIMI-A</td>
<td>2.91</td>
<td>.03</td>
<td>[2.85, 2.97]</td>
</tr>
<tr>
<td>8. WSVD</td>
<td>1.33</td>
<td>.04</td>
<td>[1.26, 1.4]</td>
</tr>
<tr>
<td>10. CVI-P</td>
<td>.17</td>
<td>.03</td>
<td>[.11, .22]</td>
</tr>
<tr>
<td>11. WSC</td>
<td>.43</td>
<td>.01</td>
<td>[.42, .45]</td>
</tr>
<tr>
<td>12. PCT</td>
<td>.56</td>
<td>.01</td>
<td>[.54, .57]</td>
</tr>
</tbody>
</table>

Note. This table presents the estimated marginal means, standard errors, and confidence intervals from the MANOVA analysis that compares faking scores by job interest. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.

*p<.05,**p<.01,***p<.001.
Table 5. Uninterested Applicants: Descriptive Statistics, Bivariate Correlations, and Internal Consistencies

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. BER</td>
<td>6.39</td>
<td>4.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(88)</td>
</tr>
<tr>
<td>2. BIMI-A</td>
<td>2.74</td>
<td>.57</td>
<td>.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.74)</td>
</tr>
<tr>
<td>3. BIMI-C</td>
<td>2.85</td>
<td>.75</td>
<td>.42</td>
<td>.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.81)</td>
</tr>
<tr>
<td>4. Bogus</td>
<td>12.85</td>
<td>5.43</td>
<td>.01</td>
<td>.37</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.80)</td>
</tr>
<tr>
<td>5. CVI-P</td>
<td>.18</td>
<td>.55</td>
<td>-.19</td>
<td>-.14</td>
<td>-.05</td>
<td>.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td>.10</td>
<td>.42</td>
<td>-.03</td>
<td>-.42</td>
<td>-.07</td>
<td>-.04</td>
<td>.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. ICS</td>
<td>.09</td>
<td>.21</td>
<td>.28</td>
<td>.16</td>
<td>.16</td>
<td>.04</td>
<td>-.07</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. IIR</td>
<td>3.62</td>
<td>.74</td>
<td>.66</td>
<td>.54</td>
<td>.41</td>
<td>.21</td>
<td>-.31</td>
<td>-.20</td>
<td>.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. PCT</td>
<td>.61</td>
<td>.13</td>
<td>.06</td>
<td>-.11</td>
<td>-.04</td>
<td>-.15</td>
<td>.01</td>
<td>.16</td>
<td>-.30</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.82)</td>
</tr>
<tr>
<td>10. RCS</td>
<td>-.12</td>
<td>.25</td>
<td>.77</td>
<td>.42</td>
<td>.43</td>
<td>.10</td>
<td>-.33</td>
<td>-.08</td>
<td>.62</td>
<td>.81</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(.91)</td>
</tr>
<tr>
<td>11. WSC</td>
<td>.47</td>
<td>.14</td>
<td>.01</td>
<td>-.20</td>
<td>-.14</td>
<td>-.17</td>
<td>-.05</td>
<td>.22</td>
<td>-.30</td>
<td>-.09</td>
<td>.52</td>
<td>-.11</td>
<td></td>
<td></td>
<td></td>
<td>(.53)</td>
</tr>
<tr>
<td>12. WSVD</td>
<td>1.13</td>
<td>.54</td>
<td>.50</td>
<td>.13</td>
<td>.16</td>
<td>-.10</td>
<td>-.07</td>
<td>-.08</td>
<td>.36</td>
<td>.12</td>
<td>-.45</td>
<td>.27</td>
<td>-.47</td>
<td></td>
<td></td>
<td>(.66)</td>
</tr>
<tr>
<td>13. T1-C</td>
<td>4.03</td>
<td>.54</td>
<td>.75</td>
<td>.37</td>
<td>.46</td>
<td>.08</td>
<td>-.34</td>
<td>-.11</td>
<td>.24</td>
<td>.81</td>
<td>.19</td>
<td>.79</td>
<td>.01</td>
<td>.21</td>
<td></td>
<td>(92)</td>
</tr>
<tr>
<td>14. T2-C</td>
<td>3.91</td>
<td>.56</td>
<td>.62</td>
<td>.32</td>
<td>.39</td>
<td>.10</td>
<td>-.31</td>
<td>-.11</td>
<td>-.17</td>
<td>.71</td>
<td>.35</td>
<td>.53</td>
<td>.14</td>
<td>.02</td>
<td>.85</td>
<td>(91)</td>
</tr>
<tr>
<td>15. T1-A</td>
<td>3.75</td>
<td>.54</td>
<td>.54</td>
<td>-.04</td>
<td>.45</td>
<td>-.02</td>
<td>-.29</td>
<td>.20</td>
<td>.22</td>
<td>.41</td>
<td>.18</td>
<td>.62</td>
<td>.15</td>
<td>.10</td>
<td>.53</td>
<td>.42</td>
</tr>
<tr>
<td>16. T2-A</td>
<td>3.67</td>
<td>.56</td>
<td>.42</td>
<td>-.10</td>
<td>.41</td>
<td>-.02</td>
<td>-.24</td>
<td>.20</td>
<td>-.08</td>
<td>.30</td>
<td>.26</td>
<td>.40</td>
<td>.26</td>
<td>-.08</td>
<td>.41</td>
<td>.41</td>
</tr>
<tr>
<td>17. T1-N</td>
<td>3.60</td>
<td>.72</td>
<td>.63</td>
<td>.51</td>
<td>.43</td>
<td>.08</td>
<td>-.25</td>
<td>-.23</td>
<td>.13</td>
<td>.84</td>
<td>.21</td>
<td>.70</td>
<td>-.04</td>
<td>.10</td>
<td>.69</td>
<td>.64</td>
</tr>
<tr>
<td>18. T2-N</td>
<td>3.47</td>
<td>.75</td>
<td>.49</td>
<td>.39</td>
<td>.30</td>
<td>.04</td>
<td>-.19</td>
<td>-.21</td>
<td>-.29</td>
<td>.68</td>
<td>.33</td>
<td>.41</td>
<td>.04</td>
<td>-.02</td>
<td>.54</td>
<td>.65</td>
</tr>
<tr>
<td>19. T1-O</td>
<td>3.49</td>
<td>.50</td>
<td>.34</td>
<td>.08</td>
<td>.01</td>
<td>.00</td>
<td>-.17</td>
<td>.01</td>
<td>.17</td>
<td>.16</td>
<td>.04</td>
<td>.41</td>
<td>-.06</td>
<td>.03</td>
<td>.17</td>
<td>.11</td>
</tr>
<tr>
<td>20. T2-O</td>
<td>3.48</td>
<td>.51</td>
<td>.27</td>
<td>.10</td>
<td>.06</td>
<td>.01</td>
<td>-.16</td>
<td>.01</td>
<td>-.08</td>
<td>.13</td>
<td>.06</td>
<td>.27</td>
<td>.04</td>
<td>-.05</td>
<td>.15</td>
<td>.15</td>
</tr>
<tr>
<td>21. T1-E</td>
<td>3.44</td>
<td>.59</td>
<td>.47</td>
<td>.58</td>
<td>.08</td>
<td>.21</td>
<td>-.25</td>
<td>-.26</td>
<td>.20</td>
<td>.65</td>
<td>-.03</td>
<td>.61</td>
<td>-.07</td>
<td>.10</td>
<td>.42</td>
<td>.35</td>
</tr>
<tr>
<td>22. T2-E</td>
<td>3.36</td>
<td>.61</td>
<td>.36</td>
<td>.48</td>
<td>-.02</td>
<td>.16</td>
<td>-.23</td>
<td>-.30</td>
<td>-.18</td>
<td>.54</td>
<td>.08</td>
<td>.34</td>
<td>.05</td>
<td>.01</td>
<td>.34</td>
<td>.38</td>
</tr>
</tbody>
</table>

Table 5 Continued.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Bivariate Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>1. BER</td>
<td></td>
</tr>
<tr>
<td>2. BIMI-A</td>
<td></td>
</tr>
<tr>
<td>3. BIMI-C</td>
<td></td>
</tr>
<tr>
<td>4. Bogus</td>
<td></td>
</tr>
<tr>
<td>5. CVI-P</td>
<td></td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td></td>
</tr>
<tr>
<td>7. ICS</td>
<td></td>
</tr>
<tr>
<td>8. IIR</td>
<td></td>
</tr>
<tr>
<td>9. PCT</td>
<td></td>
</tr>
<tr>
<td>10. RCS</td>
<td></td>
</tr>
<tr>
<td>11. WSC</td>
<td></td>
</tr>
<tr>
<td>12. WSVD</td>
<td></td>
</tr>
<tr>
<td>13. T1-C</td>
<td></td>
</tr>
<tr>
<td>14. T2-C</td>
<td></td>
</tr>
<tr>
<td>15. T1-A</td>
<td>(.89)</td>
</tr>
<tr>
<td>16. T2-A</td>
<td>.89</td>
</tr>
<tr>
<td>17. T1-N</td>
<td>.31</td>
</tr>
<tr>
<td>18. T2-N</td>
<td>.20</td>
</tr>
<tr>
<td>19. T1-O</td>
<td>.30</td>
</tr>
<tr>
<td>20. T2-O</td>
<td>.30</td>
</tr>
<tr>
<td>21. T1-E</td>
<td>.11</td>
</tr>
<tr>
<td>22. T2-E</td>
<td>.01</td>
</tr>
</tbody>
</table>
than do field studies (Hough, 1998). The Covariance Index-Personality and Covariance Index-Impression Management might not be sensitive enough to discriminate between faking with more realistic estimates. To test this, we standardized our Individual Change Score measure and included only participants with measures larger than 1 to approximate a lab sample with much higher levels of faking. Independent samples t-tests showed that Interested Applicants now scored significantly higher than non-Interested Applicants on Covariance Index-Impression Management, \( t(76) = 2.19, p = .03, d = .50 \). There was a marginal difference for Covariance Index-Personality, \( t(76) = 1.82, p = .07, d = .42 \). These results suggest that Covariance Index-Personality and Covariance Index-Impression Management perform much better when discriminating among severe faking, more often seen in experiments that often overestimate the prevalence of faking.

2.5. Discussion

Our investigation had two primary purposes: (1) examine the extent that different faking indices are able to discriminate between personality scores from same people when they respond as an applicant and as a non-applicant (Within-Subjects Faking Sensitivity), as well discriminate between personality scores from interested applicant and non-applicant samples (Between-Subjects Faking Sensitivity); and (2) determine which faking indices should be recommended for future research and practice.

We identified seven faking assessments that demonstrated both Within-Subjects Faking Sensitivity and Between Subjects Faking Sensitivity, which included: Blatant Extreme Responding, Communal Impression Management, Individual Change Scores, Idiosyncratic Item Responding, Percent Agreement, Residualized Change Scores, and Within-Subjects Variance of the Differences. Overall, Residualized Change Scores performed consistently across both sets of
analyses, but was not the best within each analysis. This research is consistent with existing research that Residualized Change Scores are the “gold-standard” in assessing applicant faking (Feeney & Goffin, 2015). Unfortunately, Residualized Change Scores are an impractical measure that can only be used by faking scholars, but has limited utility for practitioners. By contrast, Idiosyncratic Item Responding had the highest Between-Subjects Faking Sensitivity and also had the fifth highest Within-Subjects Faking Sensitivity. Our findings suggest that Idiosyncratic Item Responding is the single best measure for practitioners.

That being said, using the best faking assessment does not necessarily preclude the use of other assessments. If two effective faking assessments do not overlap to the point that they are redundant, then it may be advised to use both assessments. For example, both Idiosyncratic Item Responding and Residualized Change Scores only explained a quarter of the variance in Communal Impression Management ($rs = .57$ and $.54$ respectively), which demonstrated both types of faking sensitivity. This provides some evidence that they are all assessing the same underlying construct, but there is also some divergence. As a result, Communal Impression Management and Idiosyncratic Item Responding should be used at the same time. Blatant Extreme Responding also demonstrated both Within-Subjects Faking Sensitivity and Between-Subjects Faking Sensitivity, but only explained 50% of the variance associated with Idiosyncratic Item Responding ($r = .72$) and 25% of the variance with Communal Impression Management ($r = .55$). As a result, we argue that Idiosyncratic Item Responding, Blatant Extreme Responding, and Communal Impression Management should all be used in conjunction to assess unique variance relating to applicant faking. In academic research designs, Residualized Change Scores should be used as well.
One concern is that Communal Impression Management, Idiosyncratic Item Responding, and Blatant Extreme Responding all have significant correlations with our Interested Applicants personality scores during Time 2, when the job opportunity was revealed as fictitious, suggesting these faking assessments may be contaminated by legitimate personality variance. These relationships also held for our Uninterested Applicants during Time 2. Therefore, there is the practical concern that each measure may potentially identify a faker as a non-faker, or the reverse, and a practitioner may unwittingly remove strong candidates from the applicant pool. Given the practicality and utility of each measure, we recommend that practitioners use all three faking assessments and exclude applicants who are identified as fakers on at least two of them. Filtering out job applicants without a valid basis for that decision could open a company up to human rights complaints or legal challenges (Catano, Wiesner, Hackett, & Methot, 2010).

Fortunately, Residualized Change Scores are calculated by removing variance associated with legitimate personality variance, and in turn, had weak or non-significant correlations with Interested Applicants personality scores during Time 2.

What about the other measures that showed faking sensitivity? Communal Impression Management outperformed Agentic Impression Management—higher faking sensitivity and correlations with other faking assessments. As a result, we recommend the use of Communal Impression Management instead of Agentic Impression Management for the detection of applicant faking. Percent Agreement, Within-Subjects Correlations and Within-Subject Variance of the Differences are all impractical since they require repeated measures designs and were inferior to Residualized Change Scores. There appears to be little reason for researchers to calculate these additional faking indices. Individual Change Scores were empirically similar to Residualized Change Scores ($r = .76$), but are theoretically less pure, and were much more
contaminated with Time 2 personality scores. Therefore, there is no reason to recommend the use of Individual Change Scores. Finally, Bogus Items showed Between-Subjects Faking Sensitivity, but not Within-Subjects Faking Sensitivity, and generally had weak to non-significant correlations with other faking assessments. As a result, we are unable to draw strong conclusions. It is possible that Bogus Items are measuring a form of faking that is truly unique, but is also possible that they are just an inferior faking assessment. More research is required to validate Bogus Items as a valid faking assessment.

Both the Covariance Index-Personality and Covariance Index-Impression Management performed poorly in our study when compared to existing studies (Christiansen et al., 2017). As reviewed in our results section, both measures performed considerably better when restricted to more extreme faking, more comparative to lab study environments. These results cast some doubt on the utility of these measures in field environments. However, there may also be some unique attributes of our sample that do not generalize to applications for more substantial jobs. Another major difference between our study and Christensen et al. is that they used an adjective checklist personality inventory and we used a more conventional personality inventory. It is possible that the use of different personality measures and different experimental designs account for the difference between the two studies. As a result, future research will need to examine the value of Covariance Index-Personality and Covariance Index-Impression Management more closely.

One anomaly is that we observed inter-trait correlations that are generally higher than reported in the personality literature (Lee & Ashton, 2004). One potential explanation is that mTurk workers tend to score higher on social desirability and may use a socially desirable response set (Paolacci & Chandler, 2014). This would explain why we found that favorable traits
such as Emotional Stability, Extraversion, and Conscientiousness were more correlated with
each other, whereas neutral traits such as Openness exhibited more typical inter-trait correlations.
Similarly, we found that Residualized Change Scores exhibited moderate to large correlations
with both Time 1 and Time 2 personality scores for Uninterested Applicants, but not for
Interested Applicants. This is difficult to tease apart. What do residuals represent is the absence
of real motivation to fake? What is Conscientiousness after controlling for Conscientiousness?
Perhaps the residuals now assess a social desirability response set, which may explain the low
correlations with between Residualized Change Scores and Openness. This interpretation
suggests that it may be inappropriate to use or attempt to interpret faking indices in unmotivated
conditions.

2.5.1. Limitations

This is a novel investigation that has provided some unique insight into the assessment of
applicant faking. However, the study is not without its limitations. The primary limitation is that
the dataset was part of a broader investigation into the effects of different faking dissuasion
messages. This meant that our applicants received different personality test instructions. That
being said, the faking dissuasion messages were ineffective and the finding in this investigation
were generally consistent across conditions. For example, Residualized Change Scores, Blatant
Extreme Responding, Idiosyncratic Item Responding, and Communal Impression Management
all performed similarity well at discriminating between the same people’s applicant and non-
applicant scores. A secondary limitation is that we used an mTurk sample for a low-complexity
job, and it is possible that these findings may not generalize to higher level management
positions. Still, the literature suggests that mTurk users provide high quality data when rigorous
data screening are conducted (Buhrmester, Kwang, & Gosling, 2011; Casler, Bickel, & Hackett,
Third, for each personality construct, including the Big 5 and social desirability, there are different measures that could be selected (c.f., Bensch et al., 2017). That being said, there is considerable evidence that the measures used converge with other common personality measures (Goldberg et al., 2006) and social desirability measures (Blasberg et al., 2013).

2.5.2. Conclusion

Our study has provided a novel contribution to the literature by providing guidance for the best faking measures to use in research and practice, and through illuminating how faking measures relate to each other. Our findings suggest that researchers and practitioners should assess faking using multiple faking indices including: Idiosyncratic Item Responding, Blatant Extreme Responding, Communal Impression Management, and when possible, Residualized Change Scores. Additionally, we also demonstrated that it is useful to assess faking both within subjects, and between subjects, using mean differences, in order to evaluate the merit of newer faking measures. As we illustrated, robust faking assessments such as Residualized Change Score, Idiosyncratic Item Responding, and Communal Impression Management were effective in both instances. If a measure only succeeds with one approach, it may be biased or limited in its ability to accurately assess faking.
2.6. References


Appendix A: Job Posting

**Remote Customer Service Representative (Job Posting)**

We're looking for a Remote Customer Service Representative!

We are a research team that works with a large international firm. This firm is looking for remote customer service representatives who will help customers with online inquiries and hold live chat sessions. If you are interested, your answers to our survey will be used as a preliminary screening for employment. The candidates with the strongest person-job fit will be considered for an interview and an employment opportunity.

**This person will be responsible for:**

- Communicating with customers by live web chat session or by e-mail to provide information about products or services, take or enter orders, cancel accounts, or obtain details of complaints.
- Following up with purchases to ensure that customers are satisfied and offering individualized promotions
- Diligently check to ensure that appropriate changes were made to resolve customers' problems
- Keeping records of customer interactions or transactions, recording details of inquiries, complaints, or comments, as well as actions taken.
- Completing contract forms, prepare change of address records, or issue service discontinuance orders, using remote web interface.
- Referring unresolved customer grievances to designated departments for further investigation
- Determining charges for services requested, collect deposits or payments, or arrange for billing
- Contacting customers remotely to respond to inquiries or to notify them of claim investigation results or any planned adjustments

**Consider applying if you:**
- Enjoy working from home with flexible hours.
- You enjoy communicating with customers and making customers feel valued
- You like writing e-mails and live online conversations
- Enjoy problem solving and helping customers with novel solutions
- Developing individualized promotions for previous customers and following up on previous purchases

**What We Offer:**
- Contractual services that pay you for answering customer e-mails and following up with customers
- The opportunity to learn and develop in a growing company
- Access to Health and Dental Benefits
- A profit-sharing program
- A positive and respectful work environment

If you are interested in this opportunity, your answers will be used for both research purposes and for employment screening. You will also still receive payment for completing the HIT.

If you are not interested in this opportunity, your answers will only be used for research purposes and you are still eligible to complete the HIT.

Are you interested in this opportunity?

- Yes, I am interested being considered for the remote customer service position and also receiving payment for completing the HIT
- No, I am not interested in this additional opportunity.
Appendix B: Bogus Item Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Never Heard of It</th>
<th>Not Sure</th>
<th>Very Familiar</th>
</tr>
</thead>
<tbody>
<tr>
<td>active listening</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>customer queue</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Access online communication portal for the International Customer Service Excellence Group (ICSEG)</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Remote Method Invocation (FOIL)</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>internal customer</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>first call resolution</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>speech recognition</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>reflection</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CRM</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Utilize murray-web system to locate effective customer service solutions</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>virtual call center</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>contact center</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>customer lifecycle</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>spend management</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Open body language</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Remote Procedure Call (FOIL)</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Closed body language</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>collaborative browsing</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CSR</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Inverse Body Language (FOIL)</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Customer Care</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>customer chain</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>external customer</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>DCM (FOIL)</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Customer Relationship Management</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Customer Service Culture</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>customer first-order logic (FOIL)</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>customer is always a customer</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>top-down customer service</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>customer service oriented company</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Read the manual</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Customer Service Intelligence Test (FOIL)</strong></td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
### Appendix C: Idiosyncratic Item Responding Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Interested Applicants (Time 1)</th>
<th>Uninterested Applicants (Time 1)</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I start tasks right away.</td>
<td>4.24 (0.84)</td>
<td>3.70 (1.06)</td>
<td>0.54</td>
</tr>
<tr>
<td>2. I need a push to get started.</td>
<td>4.15 (1.04)</td>
<td>3.66 (1.13)</td>
<td>0.50</td>
</tr>
<tr>
<td>3. I love excitement.</td>
<td>4.03 (0.98)</td>
<td>3.55 (1.19)</td>
<td>0.48</td>
</tr>
<tr>
<td>4. I have difficulty starting tasks.</td>
<td>4.17 (0.97)</td>
<td>3.69 (1.15)</td>
<td>0.48</td>
</tr>
<tr>
<td>5. I am able to stand up for myself.</td>
<td>3.78 (1.11)</td>
<td>3.31 (1.21)</td>
<td>0.47</td>
</tr>
<tr>
<td>6. I get irritated easily.</td>
<td>3.97 (1.07)</td>
<td>3.51 (1.15)</td>
<td>0.46</td>
</tr>
<tr>
<td>7. I find it difficult to get down to work.</td>
<td>4.22 (0.94)</td>
<td>3.80 (1.08)</td>
<td>0.42</td>
</tr>
<tr>
<td>8. I remain calm under pressure.</td>
<td>4.26 (0.84)</td>
<td>3.84 (1.00)</td>
<td>0.42</td>
</tr>
<tr>
<td>9. I love to help others.</td>
<td>4.43 (0.77)</td>
<td>4.02 (0.90)</td>
<td>0.42</td>
</tr>
<tr>
<td>10. I rarely get irritated.</td>
<td>3.48 (1.21)</td>
<td>3.06 (1.22)</td>
<td>0.41</td>
</tr>
</tbody>
</table>
Appendix D: Main Analyses by Faking Dissuasion Message Condition

Within-Subject Faking Sensitivity for Traditional Faking Warning Condition

<table>
<thead>
<tr>
<th>Faking Index</th>
<th>Within-Subjects Faking Sensitivity ($r_d = r_{t1} - r_{t2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>1. ICS</td>
<td>.79</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.72</td>
</tr>
<tr>
<td>4. BER</td>
<td>.40</td>
</tr>
<tr>
<td>5. BIMI-C</td>
<td>.39</td>
</tr>
<tr>
<td>6. IIR</td>
<td>.33</td>
</tr>
<tr>
<td>7. CVI-IM</td>
<td>.16</td>
</tr>
<tr>
<td>8. BIMI-A</td>
<td>.16</td>
</tr>
<tr>
<td>9. CVI-P</td>
<td>.11</td>
</tr>
<tr>
<td>10. BOGUS</td>
<td>.12</td>
</tr>
<tr>
<td>11. WSC</td>
<td>-.35</td>
</tr>
<tr>
<td>12. PCT</td>
<td>-.37</td>
</tr>
</tbody>
</table>

Within-Subject Faking Sensitivity for Immediate Authentication Warning Condition

<table>
<thead>
<tr>
<th>Faking Index</th>
<th>Within-Subjects Faking Sensitivity ($r_d = r_{t1} - r_{t2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>1. ICS</td>
<td>.81</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.73</td>
</tr>
<tr>
<td>3. BER</td>
<td>.45</td>
</tr>
<tr>
<td>4. BIMI-C</td>
<td>.50</td>
</tr>
<tr>
<td>5. IIR</td>
<td>.42</td>
</tr>
<tr>
<td>6. WSVD</td>
<td>.37</td>
</tr>
<tr>
<td>7. CVI-IM</td>
<td>.32</td>
</tr>
<tr>
<td>8. BOGUS</td>
<td>.26</td>
</tr>
<tr>
<td>9. BIMI-A</td>
<td>.28</td>
</tr>
<tr>
<td>10. CVI-P</td>
<td>.15</td>
</tr>
<tr>
<td>11. WSC</td>
<td>-.39</td>
</tr>
<tr>
<td>12. PCT</td>
<td>-.44</td>
</tr>
</tbody>
</table>

Note. $N = 75$. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
### Within-Subject Faking Sensitivity for Future Authentication Warning Condition

<table>
<thead>
<tr>
<th>Faking Index</th>
<th>C</th>
<th>A</th>
<th>N</th>
<th>O</th>
<th>E</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ICS</td>
<td>.45</td>
<td>.41</td>
<td>.45</td>
<td>.18</td>
<td>.42</td>
<td>.38</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.31</td>
<td>.26</td>
<td>.34</td>
<td>.09</td>
<td>.35</td>
<td>.27</td>
</tr>
<tr>
<td>3. WSVD</td>
<td>.10</td>
<td>.17</td>
<td>.15</td>
<td>.13</td>
<td>.16</td>
<td>.14</td>
</tr>
<tr>
<td>4. BER</td>
<td>.12</td>
<td>.10</td>
<td>.12</td>
<td>.04</td>
<td>.18</td>
<td>.11</td>
</tr>
<tr>
<td>5. BIMI-C</td>
<td>.11</td>
<td>.07</td>
<td>.14</td>
<td>-.05</td>
<td>.18</td>
<td>.09</td>
</tr>
<tr>
<td>6. IIR</td>
<td>.06</td>
<td>.07</td>
<td>.15</td>
<td>-.06</td>
<td>.12</td>
<td>.07</td>
</tr>
<tr>
<td>7. CVI-P</td>
<td>.05</td>
<td>.04</td>
<td>-.05</td>
<td>.06</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td>8. BIMI-A</td>
<td>.00</td>
<td>-.07</td>
<td>.03</td>
<td>.02</td>
<td>.03</td>
<td>.00</td>
</tr>
<tr>
<td>9. CVI-IM</td>
<td>-.06</td>
<td>-.01</td>
<td>-.06</td>
<td>.04</td>
<td>-.07</td>
<td>-.03</td>
</tr>
<tr>
<td>10. BOGUS</td>
<td>-.03</td>
<td>-.11</td>
<td>.01</td>
<td>-.06</td>
<td>.01</td>
<td>-.04</td>
</tr>
<tr>
<td>11. WSC</td>
<td>-.15</td>
<td>-.22</td>
<td>-.21</td>
<td>-.07</td>
<td>-.17</td>
<td>-.16</td>
</tr>
<tr>
<td>12. PCT</td>
<td>-.34</td>
<td>-.29</td>
<td>-.34</td>
<td>-.04</td>
<td>-.29</td>
<td>-.26</td>
</tr>
</tbody>
</table>

Note. \( N = 73 \). BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
**Within-Subject Faking Sensitivity for Moral Suasion Prompt Condition**

<table>
<thead>
<tr>
<th>Faking Index</th>
<th>Within-Subjects Faking Sensitivity ($r_d = r_{t1} - r_{t2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>1. ICS</td>
<td>.50</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.45</td>
</tr>
<tr>
<td>3. WSVD</td>
<td>.27</td>
</tr>
<tr>
<td>4. BER</td>
<td>.25</td>
</tr>
<tr>
<td>5. CVI-P</td>
<td>.23</td>
</tr>
<tr>
<td>6. IIR</td>
<td>.21</td>
</tr>
<tr>
<td>7. BIMI-A</td>
<td>.06</td>
</tr>
<tr>
<td>8. BIMI-C</td>
<td>.14</td>
</tr>
<tr>
<td>9. CVI-IM</td>
<td>.06</td>
</tr>
<tr>
<td>10. BOGUS</td>
<td>-.20</td>
</tr>
<tr>
<td>11. WSC</td>
<td>-.18</td>
</tr>
<tr>
<td>12. PCT</td>
<td>-.33</td>
</tr>
</tbody>
</table>

**Note.** $N = 60$. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
## Within-Subject Faking Sensitivity for Control Condition

<table>
<thead>
<tr>
<th>Faking Index</th>
<th>Within-Subjects Faking Sensitivity ($r_d = r_{t1} - r_{t2}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>C</td>
</tr>
<tr>
<td>1. ICS</td>
<td>.53</td>
</tr>
<tr>
<td>2. RCS</td>
<td>.39</td>
</tr>
<tr>
<td>3. IIR</td>
<td>.21</td>
</tr>
<tr>
<td>4. BIMI-C</td>
<td>.11</td>
</tr>
<tr>
<td>5. BER</td>
<td>.10</td>
</tr>
<tr>
<td>6. WSVD</td>
<td>.01</td>
</tr>
<tr>
<td>7. BIMI-A</td>
<td>-.01</td>
</tr>
<tr>
<td>8. CVI-IM</td>
<td>-.09</td>
</tr>
<tr>
<td>9. BOGUS</td>
<td>-.12</td>
</tr>
<tr>
<td>10. CVI-P</td>
<td>-.23</td>
</tr>
<tr>
<td>11. WSC</td>
<td>-.20</td>
</tr>
<tr>
<td>12. PCT</td>
<td>-.38</td>
</tr>
</tbody>
</table>

Note. $N = 70$. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
## Between-Subjects Faking Sensitivity: Traditional Warning Condition

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Interested Applicants</th>
<th>Non-Interested Applicants</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>BER</td>
<td>79</td>
<td>9.06</td>
<td>.53</td>
</tr>
<tr>
<td>BIMI-A</td>
<td>79</td>
<td>2.97</td>
<td>.07</td>
</tr>
<tr>
<td>BIMI-C</td>
<td>79</td>
<td>3.26</td>
<td>.09</td>
</tr>
<tr>
<td>BOGUS</td>
<td>17.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVI-P</td>
<td>78</td>
<td>.13</td>
<td>.04</td>
</tr>
<tr>
<td>CVI-IM</td>
<td>77</td>
<td>.17</td>
<td>.05</td>
</tr>
<tr>
<td>ICS</td>
<td>79</td>
<td>.29</td>
<td>.04</td>
</tr>
<tr>
<td>PCT</td>
<td>79</td>
<td>.55</td>
<td>.02</td>
</tr>
<tr>
<td>RCS</td>
<td>79</td>
<td>.09</td>
<td>.03</td>
</tr>
<tr>
<td>WSC</td>
<td>79</td>
<td>.43</td>
<td>.02</td>
</tr>
<tr>
<td>WSVD</td>
<td>79</td>
<td>1.32</td>
<td>.08</td>
</tr>
</tbody>
</table>

Note. This table presents the means, standard errors, and confidence intervals from the independent t-test analyses that compares faking scores by job interest. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
## Between-Subjects Faking Sensitivity: Immediate Authentication Warning Condition

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Interested Applicants</th>
<th></th>
<th>Non-Interested Applicants</th>
<th></th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
<td>n</td>
</tr>
<tr>
<td><strong>BIMI-A</strong></td>
<td>75</td>
<td>2.90</td>
<td>.07</td>
<td>[2.76, 3.04]</td>
<td>23</td>
</tr>
<tr>
<td><strong>BIMI-C</strong></td>
<td>75</td>
<td>3.22</td>
<td>.08</td>
<td>[3.05, 3.38]</td>
<td>23</td>
</tr>
<tr>
<td><strong>BOGUS</strong></td>
<td>75</td>
<td>15.23</td>
<td>.71</td>
<td>[13.84, 16.61]</td>
<td>23</td>
</tr>
<tr>
<td><strong>CVI-P</strong></td>
<td>74</td>
<td>.16</td>
<td>.04</td>
<td>[.08, .24]</td>
<td>23</td>
</tr>
<tr>
<td><strong>CVI-IM</strong></td>
<td>74</td>
<td>.11</td>
<td>.03</td>
<td>[.04, .17]</td>
<td>22</td>
</tr>
<tr>
<td><strong>ICS</strong></td>
<td>75</td>
<td>.23</td>
<td>.05</td>
<td>[.14, .32]</td>
<td>23</td>
</tr>
<tr>
<td><strong>IIR</strong></td>
<td>75</td>
<td>4.02</td>
<td>.07</td>
<td>[3.88, 4.16]</td>
<td>23</td>
</tr>
<tr>
<td><strong>PCT</strong></td>
<td>75</td>
<td>.54</td>
<td>.02</td>
<td>[.51, .58]</td>
<td>23</td>
</tr>
<tr>
<td><strong>RCS</strong></td>
<td>75</td>
<td>.02</td>
<td>.04</td>
<td>[-.05, .1]</td>
<td>23</td>
</tr>
<tr>
<td><strong>WSC</strong></td>
<td>75</td>
<td>.42</td>
<td>.02</td>
<td>[.38, .45]</td>
<td>23</td>
</tr>
<tr>
<td><strong>WSVD</strong></td>
<td>75</td>
<td>1.43</td>
<td>.10</td>
<td>[1.24, 1.62]</td>
<td>23</td>
</tr>
</tbody>
</table>

Note. This table presents the means, standard errors, and confidence intervals from the independent t-test analyses that compares faking scores by job interest. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
Between-Subjects Faking Sensitivity: Future Authentication Warning Condition

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Interested Applicants</th>
<th></th>
<th></th>
<th>Non-Interested Applicants</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
<td>n</td>
<td>M</td>
</tr>
<tr>
<td>2. BIMI A</td>
<td>73</td>
<td>2.98</td>
<td>.06</td>
<td>[2.86, 3.1]</td>
<td>36</td>
<td>2.66</td>
</tr>
<tr>
<td>5. CVI-P</td>
<td>71</td>
<td>.16</td>
<td>.07</td>
<td>[.02, .30]</td>
<td>34</td>
<td>.15</td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td>73</td>
<td>.12</td>
<td>.04</td>
<td>[.04, .20]</td>
<td>35</td>
<td>.09</td>
</tr>
<tr>
<td>7. ICS</td>
<td>73</td>
<td>.23</td>
<td>.03</td>
<td>[.17, .29]</td>
<td>36</td>
<td>.04</td>
</tr>
<tr>
<td>9. PCT</td>
<td>73</td>
<td>.58</td>
<td>.02</td>
<td>[.54, .62]</td>
<td>36</td>
<td>.61</td>
</tr>
<tr>
<td>10. RCS</td>
<td>73</td>
<td>.05</td>
<td>.03</td>
<td>[-.01, .11]</td>
<td>36</td>
<td>-.13</td>
</tr>
<tr>
<td>11. WSC</td>
<td>73</td>
<td>.45</td>
<td>.01</td>
<td>[.43, .47]</td>
<td>36</td>
<td>.48</td>
</tr>
<tr>
<td>12. WSVD</td>
<td>73</td>
<td>1.26</td>
<td>.07</td>
<td>[1.12, 1.4]</td>
<td>36</td>
<td>1.11</td>
</tr>
</tbody>
</table>

Note. This table presents the estimated marginal means, standard errors, and confidence intervals from the independent t-test analyses that compares faking scores by job interest. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
### Between-Subjects Faking Sensitivity: Moral Suasion Prompt

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Interested Applicants</th>
<th>Non-Interested Applicants</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>1. BER</td>
<td>60</td>
<td>8.89</td>
<td>.63</td>
</tr>
<tr>
<td>2. BIMI A</td>
<td>60</td>
<td>2.94</td>
<td>.08</td>
</tr>
<tr>
<td>3. BIMI C</td>
<td>60</td>
<td>3.40</td>
<td>.10</td>
</tr>
<tr>
<td>4. BOGUS</td>
<td>54</td>
<td>.17</td>
<td>.06</td>
</tr>
<tr>
<td>5. CVI-P</td>
<td>58</td>
<td>.09</td>
<td>.03</td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td>60</td>
<td>.27</td>
<td>.04</td>
</tr>
<tr>
<td>7. ICS</td>
<td>60</td>
<td>4.20</td>
<td>.08</td>
</tr>
<tr>
<td>8. IIR</td>
<td>60</td>
<td>.57</td>
<td>.02</td>
</tr>
<tr>
<td>9. PCT</td>
<td>60</td>
<td>.09</td>
<td>.03</td>
</tr>
<tr>
<td>10. RCS</td>
<td>60</td>
<td>.42</td>
<td>.02</td>
</tr>
<tr>
<td>11. WSC</td>
<td>60</td>
<td>1.34</td>
<td>.09</td>
</tr>
</tbody>
</table>

Note. This table presents the estimated marginal means, standard errors, and confidence intervals from the independent t-test analyses that compares faking scores by job interest. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
Between-Subjects Faking Sensitivity: Control Condition

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Interested Applicants</th>
<th>Non-Interested Applicants</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( n )</td>
<td>( M )</td>
<td>( SE )</td>
</tr>
<tr>
<td>1. BER</td>
<td>70</td>
<td>8.36</td>
<td>.61</td>
</tr>
<tr>
<td>2. BIMI A</td>
<td>70</td>
<td>2.81</td>
<td>.07</td>
</tr>
<tr>
<td>3. BIMI C</td>
<td>70</td>
<td>3.26</td>
<td>.09</td>
</tr>
<tr>
<td>4. BOGUS</td>
<td></td>
<td>16.23</td>
<td></td>
</tr>
<tr>
<td>5. CVI-P</td>
<td>67</td>
<td>.24</td>
<td>.07</td>
</tr>
<tr>
<td>6. CVI-IM</td>
<td>70</td>
<td>.16</td>
<td>.04</td>
</tr>
<tr>
<td>7. ICS</td>
<td>70</td>
<td>.29</td>
<td>.03</td>
</tr>
<tr>
<td>8. IIR</td>
<td>70</td>
<td>3.92</td>
<td>.09</td>
</tr>
<tr>
<td>9. PCT</td>
<td>70</td>
<td>.56</td>
<td>.02</td>
</tr>
<tr>
<td>10. RCS</td>
<td>70</td>
<td>.05</td>
<td>.03</td>
</tr>
<tr>
<td>11. WSC</td>
<td>70</td>
<td>.44</td>
<td>.02</td>
</tr>
<tr>
<td>12. WSVD</td>
<td>70</td>
<td>1.36</td>
<td>.08</td>
</tr>
</tbody>
</table>

Note. This table presents the estimated marginal means, standard errors, and confidence intervals from the independent t-test analyses that compares faking scores by job interest. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means. BER = Blatant Extreme Scores. BIMI-A = Agentic Impression Management. BIMI-C = Communal Impression Management. IIR = Idiosyncratic Item Responding. RCS = Residual Change Scores. PCT = Percent Agreement. ICS = Individual Change Scores. WSC = Within Subjects Correlation. WSVD = Within-Subject Variance of the Differences. CVI-P = Covariance Index-Personality. CVI-IM = Covariance Index-Impression Management.
Chapter 3

3. Faking Dissuasion Messages in the Applicant Screening Context

In this chapter, we investigated whether new faking dissuasion messages that incorporate accountability theory and morality theory can be used to reduce applicant faking. As existing faking dissuasion messages are associated with heightened negative reactions from applicants, we aimed to test messages that reduce faking without evoking such negative reactions.

3.1 Abstract

The applicant faking literature suggests that faking dissuasion messages—brief messages that informs applicants that faking can be detected—can be an effective tool at combatting faking on personality tests. Faking dissuasion messages can reduce faking by up to 50% (Fan et al., 2012). However, warnings are typically atheoretical and these threatening messages can cause applicants to feel negatively about the personality test (Converse et al., 2008; Goffin & Boyd, 2009). We tried to improve the efficacy of faking dissuasion messages, while minimizing negative applicant reactions, by leveraging the accountability and morality literatures. We conducted a study using Amazon’s Mechanical Turk, with 518 participants who believed they were considered for a real employment opportunity for a remote customer service position. Compared to participants not interested in the job, participants who were interested in the job had significantly more favorable personality scores. After informing participants that the position was fictitious and asking them to fill out the same personality measure as honestly as possible, the same previously interested participants received significantly less favorable personality scores. However, there were no differences in faking across the different faking dissuasion message conditions, suggesting the faking dissuasion messages were ineffective. After a review
of the literature, we speculated that faking dissuasion messages may be less effective in screening scenarios than when used later in the selection process.

**Keywords**

Personality assessment; job applications; faking; faking warnings; mTurk

### 3.2 Introduction

Imagine you are being considered for a coveted job—one that has everything you want—and the hiring company requests that you complete a personality test. The test asks about your punctuality—a trait that you know is highly desirable, but one that is a personal weakness of yours. Do you answer honestly or lie? Job applications generate incentive for applicants to “fake” to be more competitive in selection contexts. Faking has been defined as applicants deliberately providing “inaccurate responses to personality items in a manner that they believe will increase their chances of obtaining valued outcomes, such as a favorable hiring decision” (Goffin & Boyd, 2009, p. 151)

There is compelling evidence that a large percentage of applicants fake their responses to personality tests (Donovan et al., 2013; Holden et al., 2017; Jeong, Christiansen, Robie, Kung, & Kinney, 2017). Applicant faking on personality tests reduces the psychometric validity of such tests. For example, applicants that fake obtain reduced self-peer correlations on personality tests (Robie et al., 2009) and reduced criterion-related validity (M. H. Peterson, Griffith, Isaacson, et al., 2011). Reductions in psychometric validity may have grave consequences for organizations by reducing the accuracy of hiring decisions.

Several methods exist to mitigate the negative outcomes of applicant faking (Griffith & Robie, 2013). One popular technique is to include faking dissuasion messages, messages placed prior to or during the personality test, to dissuade individuals from faking on the test. The current
best practice is to inform applicants that faking can be detected and that detection will lead to immediate disqualification from the selection process (Dilchert & Ones, 2011). Several studies suggest that faking dissuasion messages may reduce the extent that each applicant fakes by between 30% and 50% (Fan et al., 2012; Landers et al., 2011; Robson et al., 2008). That being said, there are some studies that suggest faking dissuasion messages are much less effective in reducing faking (Fisher, Robie, Christiansen, & Komar, 2018; Vasilopoulos et al., 2005).

### 3.2.1 Present Study: Improving Faking Dissuasion Messages

One potential reason that faking dissuasion messages vary in their efficacy is that they tend to be atheoretical and do not leverage the psychological processes that underlie faking (Goffin & Boyd, 2009), thus limiting their potential applicability. See Appendix A for an example of a Traditional Faking Warning (TFW) adapted from the existing literature (Dwight & Donovan, 2003). Several scholars have suggested the development of new faking dissuasion messages should be derived from theory to improve their efficacy (Goffin & Boyd, 2009; Pace & Borman, 2006).

In the present study, we presented a job opportunity to participants and asked if they were interested in the position. We instructed participants that interested participants would be contacted if they were a good fit for the position and provided them a personality test. As reviewed in more detail later, we randomly assigned participants to different faking dissuasion conditions where participants read different test instructions. These instructions included traditional faking warnings and as well as a control group. They also included new faking dissuasion messages that capitalized on either accountability or morality theory.

We examined if we could bolster the efficacy of faking dissuasion messages by crafting them to incorporate core lessons from accountability theory (Lerner & Tetlock, 1999). The
accountability literature revealed two methods that might enhance the potency of faking dissuasion messages: applicants should feel more accountable when they believe (a) there is a credible process in place to check the validity of their answers, and (b) they will need to defend their answers on an item-by-item basis—rather than just their overall score.

We composed two faking dissuasion messages that capitalized on these methods. The first warning is an Immediate Authentication Warning (IAW; Appendix B), which recommends applicants answer honestly because the hiring organization can verify their responses. We inform the applicants that the organization can verify responses by contacting references, employing “internal methodological procedures” (the term we used to describe social desirability scales) and explain how these procedures work to boost credibility. The second warning is a Future Authentication Warning (FAW; Appendix C), which advises applicants to answer honestly because the hiring organization will compare their test responses to future behaviors during interviews or on-the-job performance. The FAW also tries to reduce faking by instructing applicants that answering honestly is in their best interest, as faking may lead to being hired for a position for which they are not qualified, and hence diminishing their career potential. Both the IAW and FAW informs applicants that the consequence of being detected is a one-on-one interview, where they will need to defend their answers.

In contrast to threatening faking dissuasion messages, scholars have also suggested appealing to the test-takers’ morality to reduce applicant faking (Goffin & Boyd, 2009; Pace & Borman, 2006; Robie, Brown, & Beaty, 2007). Goffin and Boyd (2009) postulated that morality is the first issue that test-takers consider when faking on personality test items. Therefore, the efficacy of faking dissuasion messages to reduce faking might be enhanced by appealing to applicants’ morality. Consistent with this assertion, Uruena and Robie (Uruena & Robie, 2011, p.
17) found that applicants who read the test instructions “dishonest or distorted self-descriptions are simply wrong and do not adhere to commonly accepted standards of behavior” scored lower on a measure of conscientiousness than applicants who did not read them ($d = .23$). These results suggest that being prompted to consider the morality of faking may lead to reductions in applicant faking. To enhance the efficacy of this approach, we created a moral suasion appeal (MSA; Appendix D) that utilizes theoretical work on morality, and principles from the persuasion literature, to increase its salience with applicants.

Morality researchers in the morality literature suggest that people use a special set of rules to make morality judgments as opposed to judgements about less sensitive issues (Haidt, 2001). People form moral judgements based on intuitive opinions that stem from their emotional reactions to events or thoughts, instead of reaching a position through rationality. Across several studies, people are unable to provide rational explanations for their moral judgments, but instead, their judgments tend to be predicted by negative emotional responses (Haidt, 2001; Haidt & Hersh, 2001; Haidt, Koller, & Dias, 1993). For example, people tend to be morally opposed to cleaning a toilet using a national flag, despite their inability to rationally defend why such behavior is wrong (Haidt et al., 1993). Subsequently, arguments that defend morality positions are often developed post-hoc rather than informing the morality decision, and in turn, people only generate arguments that support their position (Haidt, 2001). Accordingly, we may be able to reduce faking by prompting negative emotional reactions to faking on personality tests.

In addition to explaining that faking is morally wrong, we composed a MSA to elicit negative emotions toward faking and positive emotions toward answering honestly by utilizing affective messages. This is supported by the persuasion literature, which shows that arousing affective states can influence people’s cognitions and, in turn, affect desired outcomes (Cialdini...
Our intention was that prompting applicants to feel negative emotions about faking may get them to think negatively about faking, and therefore, persuade applicants to answer honestly. Additionally, pictures have been shown to help elicit negative and positive emotions when they cause viewers to think of unpleasant or pleasant thoughts respectively, so we also included photographs to illustrate the text of our MSA (Lang, Greenwald, Bradley, & Hamm, 1993; Schimmack, 2005).

**Research Question 1:** To what extent will applicants fake on a personality test when presented with one of our newly developed faking dissuasion messages (IAW, FAW, MSA) as compared to a traditional faking warning (TFW) or no warning (NW)?

### 3.2.2 The Role of Verifiability

One potential concern is that the efficacy of increased accountability may vary as a function of how verifiable applicants perceive each personality item to be. Goffin and Boyd (2009) suggest that applicants consider the risk of being caught for faking for each personality item. Therefore, the IAW and FAW should be most effective when the applicant perceives the personality item as verifiable, because the threat of verification should appear more credible than when the applicant thinks the item is unverifiable. For example, if an applicant were to respond to the item “I would be quite bored by a visit to an art gallery”, the applicant may not believe there is a credible method to verify a response to this item, because it assesses behavior outside of the workplace. By contrast, the item “I avoid making ‘small talk’ with people” is something work references are likely able to evaluate. Thus, we expect that the IAW, FAW, and TFW should be more effective for items that appear verifiable because the warnings will appear more credible.
**Research Question 2:** Will item verifiability moderate the efficacy of our faking dissuasion messages?

### 3.2.3 Applicant Reactions

Faking dissuasion messages that threaten applicants can have the unintended consequence of increasing test-taking anxiety (Converse et al., 2008). In addition, warnings increase the difficulty of filling out the personality test—especially for those with low GMA (Vasilopoulos et al., 2005)—and therefore, applicants may form negative reactions about the fairness or appropriateness of the test. Applicant reactions are important because applicants crystallize negative impressions about the organization, and this could, in turn, facilitate undesired outcomes such as discouraging others from applying to the organization (McCarthy, Hrabluik, & Jelley, 2009).

Both the IAW and FAW threaten applicants with multiple sources of verification and may instigate higher levels of test-taking anxiety and negative perceptions of procedural justice. For example, applicants may think it is unfair to use their references or on the job performance to verify their personality responses, or become anxious of future detection. By contrast, our MSA may lead to less anxiety and more favorable perceptions of procedural justice, because it omits threatening language. That being said, honest respondents might find the threats in the IAW or FAW reassuring, because they may believe that fakers will be detected and punished, making the personality test more fair and accurate. Similarly, the MSA may also alert applicants that the competition is faking and does not provide any corrective mechanism, which could weaken perceptions of accuracy and fairness. As a result, it is difficult to draw clear hypotheses of how the new faking dissuasion messages will affect applicant reactions. This investigation presents a
novel contribution to the literature because few studies have examined how faking dissuasion messages impact applicant reactions.

**Research Question**: To what extent, if any, will the IAW, FAW, or MSP influence test-taking anxiety and procedural justice?

### 3.3 Method

#### 3.3.1 Participants

A total of 779 participants were recruited via Amazon’s Mechanical Turk (mTurk) online marketplace for a two-session research study. Participants were paid $2 for their time, which is consistent with the mTurk average rate of $1.17 per hour (Mason & Suri, 2012). From our initial sample, 572 participants completed both sessions (73.4%). We also excluded responses from 38 applicants for careless responding, including failing attention checks and self-reported inattentiveness. These measures are reviewed in the materials section.

The final sample retained 357 participants ($M_{age} = 35.10, SD = 10.64$) who indicated they were interested in the position (“Interested Applicants”). The majority were female ($n = 223$), employed full time ($n = 225$) or part-time ($n = 71$). Interested Applicants had a modal income of $30,000 to $39,999 USD ($n = 69$), a modal education of a four-year university degree ($n = 135$). We had 79, 75, 73, 60, and 70 Interested Applicants in the TW, IAW, FAW, MSA, and NFW conditions respectively.

The final sample also retained 177 participants ($M_{age} = 34.14, SD = 11.82$) who were not interested in the position (“Uninterested Applicants”). The majority were female ($n = 96$), employed full-time ($n = 131$), the modal income was $20,000 to $29,000 USD ($n = 24$), and the modal education was a four-year university degree ($n = 57$). We had 40, 23, 36, 43, and 35 Uninterested Applicants in the TW, IAW, FAW, MSA, and NFW conditions respectively.
3.3.2 Procedure

Participants were recruited for a personality measurement study on mTurk and after signing their consent forms—but before they completed the questionnaires—we informed participants of a job opportunity. We told participants that a company was looking to hire contractual customer-service representatives who would be allowed to work from home. Participants were presented with the job posting (see Appendix E) and asked to indicate if they would like to be considered for the job. We informed them that if their answers suggested they were well-suited to the position, they would be selected for a video-conferred job interview. The job posting was modeled on an actual customer-service job posting, and the Customer Service Representatives section of the Occupational Information Network (Listing #43-4051.00; N. G. Peterson et al., 2001). Uninterested applicants were still permitted to complete the study and receive compensation, thus serving as a control group for our job application manipulation. In the preliminary analyses, we examine if our Interested Applicants received more favorable personality scores than did Uninterested Applicants.

During the first session (“Time 1”), all participants were randomly assigned to one of five experimental conditions: TFW, IAW, FAW, MSA, and NW. All of the faking dissuasion messages had a Reading Ease score and a Flesch-Kincaid Grade Level that were appropriate for a typical 13 to 15 year old (Flesch, 1948; Kincaid, Fishburne Jr, Rogers, & Chissom, 1975). Participants in the faking dissuasion conditions read their assigned message and were asked to fill out the personality and careless responding measures described below. Participants in the NW condition viewed a page that asked them to click “continue” to begin their personality assessment and careless responding measures. At the end of our survey, participants were informed that we were not working with a real company. We invited all participants to a follow
up study one week later for an additional dollar of compensation. At the beginning of the follow-up survey (“Time 2”), participants were reminded that there was no employment opportunity and were asked to respond, as honestly as possible, to our personality test and post-experiment survey.

3.3.3 Measures

Careless Responding. One potential concern with mTurk is that some people respond carelessly (Buhrmester et al., 2011). In accordance with Meade and Craig (2012), participants were asked to respond to six objectively verifiable items (e.g., how much they agree with “The sum of 2+2 is 10.”) placed throughout our personality test. Meade and Craig also found that some careless responders will admit at the end of the study that their data should not be used in final analyses, so long as there is no penalty for doing so. Consistent with this recommendation, we asked participants if we should include their results in the analyses of the present study.

Faking. We examined applicant faking using the four faking assessments discussed in Study 1.

Blatant Extreme Responding. Blatant Extreme Responding was calculated by summing the number of times participants endorsed the most favorable response (i.e., strongly agree) on each personality item during Time 1. Thus, applicants received scores of 1 (extreme) or 0 (not extreme) for each item, which were then summed to compute a composite index of Blatant Extreme Responding. In theory, Blatant Extreme Responding should be positively related to applicant faking. Blatant Extreme Responding has been used in previous investigations and is an effective measure of applicant faking (Landers et al., 2011; Levashina et al., 2014).

Idiosyncratic Item Responding. Idiosyncratic items differentiate job applicant responses from non-applicant responses (Kuncel & Borneman, 2007). We identified
Idiosyncratic items by selecting the 10 items with the largest mean differences between our Interested Applicant and Uninterested Applicant samples. Idiosyncratic Item Responding scale scores should be positively correlated with applicant faking. Example items include “I start tasks right away” and “I find it difficult to approach others.” The items from this scale are part of the personality measure and are scored during the job application session.

**Communal Impression Management.** We measured impression management using Communal Impression Management, which involves “denying socially deviant impulses and claiming pious attributes” (Blasberg et al., 2013, p. 523). The Communal Impression Management is assessed using 10 items measured on 5-point response scales (1 = *Strongly Disagree*, 5 = *Strongly Agree*). An example item is “I don’t gossip about other people’s business.” Internal consistencies exceeded .70 across two studies, and there is evidence of both convergent and discriminant validity (Blasberg et al.). These items were embedded randomly throughout the Personality assessment.

**Residualized Change Scores.** We measured Mean Shifts using Residualized Change Scores (Burns & Christiansen, 2011; Feeney & Goffin, 2015). Residualized Change Scores are calculated by individually regressing the same person’s score on each personality item during a job application, on the same person’s score during a non-applicant condition—and saving the residuals. These residuals reflect applicant scores without variance associated with honest scores. Residualized Change Scores are a more accurate faking index than Individual Change Scores because they are not contaminated by legitimate personality variance. If a faking index reflects underlying personality traits, then scores may be due to either faking or the personality trait (Burns & Christiansen, 2011), creating inferential errors.
**Faking Dissuasion Message Viewing Time.** To ensure that applicants read our faking discussion messages, we recorded the number of seconds each participant spent viewing them. The TFW, IAW, FAW, and MSA consisted of 76, 394, 334, and 228 words respectively, and are located in Appendices A through D.

**Job Relatedness and Verifiability.** We conducted an expert judgment panel to determine the job-relatedness and the verifiability of our personality items. Our sample consisted of 15 experts in industrial and organizational psychology—familiar with both personality measurement and personnel selection: two recent doctoral graduates, nine doctoral students, and four master’s students. Each expert provided a set of job-relatedness ratings (5-point scale: *Not at all Important* to *Extremely Important*) and verifiability (5-point scale: *Extremely Difficult* to *Extremely Easy*) and were each paid $10 CAD for their assistance. For job relatedness, scores for one expert were excluded for having unacceptably low inter-rater reliability ($r < .50$). The ICC (3, k; Shrout & Fleiss, 1979) for job-relatedness was .96. The mean ratings were: Conscientiousness, 4.01; Agreeableness, 3.19; Emotional Stability, 3.10; Extraversion, 3.00; and Openness, 1.90. For verifiability, we dropped four experts for having low inter-rater reliability ($r < .50$), suggesting some disagreement among our reviewers. The resulting ICC (3, k) for verifiability was .90. We created a 9-item item composite by including items that experts rated as being both job-related ($M > 4$) and verifiable ($M > 3.5$). The resulting scale was reliable ($\alpha = .86$) and included items from each personality dimension, except Openness.

**Personality.** Personality was assessed using the International Personality Item Pool (IPIP; Goldberg et al., 2006), which examines: Agreeableness, Conscientiousness, Emotional Stability, Extraversion, And Openness. These dimensions were assessed using 120 items on a 5-point response scale ($1 = Strongly Disagree, 5 = Strongly Agree$). The 120-item scale has strong
internal consistencies between .87 and .90 for the five dimensions (Maples et al., 2014). Additionally, all five dimensions on the 120-item scale were found to parallel their respective NEO Personality Inventory dimension with correlations between .87 and .90 (Maples et al., 2014).

**Procedural Justice Perceptions.** Procedural justice perceptions was assessed using three dimensions from the Selection Procedural Justice Scale (Bauer et al., 2001), which were: Job-Relatedness, Information Known, and Chances to Perform. The resulting measure contains nine items that use a 5-point response scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*). An example item from the job-relatedness scale is “Doing well on this test means a person can do the retail manager job well.” The Selection Procedural Justice Scale (SPJS) possesses excellent reliability ($\alpha = .88$) and predicts meaningful outcomes such as whether applicants would recommend the organization to others and their perceptions of organizational attractiveness (Bauer et al., 2001; McCarthy et al., 2013, 2009).

**Test-Taking Anxiety.** We assessed Test-Taking Anxiety using the comparative anxiety subscale from the Test Attitude Survey (Arvey, Strickland, Drauden, & Martin, 1990). This scale is measured using 10 items on a five 5-point response scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*). The scale has good evidence of reliability ($\alpha = .80$) and criterion-related validity (Arvey et al., 1990).

### 3.4 Results

Prior to assessing whether our faking dissuasion messages reduced faking, we assessed the strength of our manipulations. First, we examined if our fictitious job posting was successful in prompting applicants to fake. Second, we examined if our applicants read the faking dissuasion messages.
3.4.1 Preliminary Analyses: Job Manipulation Check

If our job manipulation was effective, we would expect that Interested Applicants received more favorable personality scores than Uninterested Applicants during Time 1, when the job opportunity was presented, but not during Time 2, when the job opportunity was revealed as fictitious. Repeated measures ANOVA revealed a significant interaction effect between job interest (Interested Applicants and Uninterested Applicants) and time (Time 1, and Time 2). On average, Interested Applicants scored significantly higher on Emotionality Stability, Conscientiousness, Agreeableness, and Extraversion than did Uninterested participants during Time 1, but not during Time 2, $F(5, 528) = 10.75, p < .001$, Wilk’s $\Lambda = .91$. Interested Applicants received significantly more favorable personality scores on Conscientiousness, Emotional Stability, Extraversion, and Agreeableness than Uninterested Applicants during Time 1, but not during Time 2. The results of the pairwise comparisons are reported in Table 6. These results suggest that our job manipulation was successful in inducing applicant faking.

3.4.2 Preliminary Analyses: Did Applicants Read the Faking Dissuasion Messages?

To examine if applicants read the faking dissuasion messages, we conducted a univariate ANOVA, with job interest and faking dissuasion condition as the between subject variables and reading time as the dependent variable. There was a main effect for faking dissuasion condition, $F(4, 522) = 49.15, p < .001$, $\eta_p^2 = .27$. Tukey’s B post hoc tests revealed that applicants in IAW condition spent longer reading the faking dissuasion message ($M = 150.32$) than applicants in the MSA ($M = 50.14s$) and FAW ($M = 61.09s$) conditions, who spent significantly longer
Table 6. Pairwise Comparisons of Personality Scores by Job Interest in Time 1 and Time 2

| Personality Dimension | Time 1 | | | Time 2 | | |
|-----------------------|--------|--------|--------|--------|--------|
|                       | Interested Applicants | Uninterested Applicants | d     | Interested Applicants | Uninterested Applicants | d     |
|                       | M     | SE    | 95% CI | M     | SE    | 95% CI | M     | SE    | 95% CI |
| Conscientiousness     | 4.30  | 0.03  | [4.25, 4.36] | 4.03  | 0.04  | [3.96, 4.11] | .52   | 3.95  | 0.03  | [3.89, 4.01] |
| Agreeableness         | 3.95  | 0.02  | [3.90, 4.00] | 3.75  | 0.04  | [3.68, 3.82] | .42   | 3.73  | 0.03  | [3.67, 3.78] |
| Emotional Stability   | 3.87  | 0.04  | [3.80, 3.94] | 3.60  | 0.05  | [3.55, 3.75] | .39   | 3.43  | 0.04  | [3.35, 3.51] |
| Openness              | 3.47  | 0.03  | [3.42, 3.52] | 3.49  | 0.04  | [3.42, 3.57] | -     | 3.45  | 0.03  | [3.39, 3.51] |
| Extraversion          | 3.65  | 0.03  | [3.59, 3.71] | 3.44  | 0.04  | [3.35, 3.52] | .38   | 3.38  | 0.03  | [3.31, 3.44] |

Note. This table includes the estimated marginal means, standard errors, and 95% confidence intervals for each personality dimension in Time 1 and in Time 2. Ns for Interested and Uninterested Applicants were 357 and 177 respectively. Cohen’s D were calculated using the descriptive statistics, which may vary slightly from the estimated marginal means.
reading the faking dissuasion message than applicants in the TFW \((M = 19.23\text{s})\) and NW conditions \((M = 1.79\text{s})\). The main effect for job interest was significant, \(F(1, 522) = 4.80, p = .03, \eta^2_p = .01\). Interested Applicants spent marginally more time reading the faking dissuasion message \((M = 63.95\text{s})\) than Uninterested Applicants did \((M = 49.06\text{s})\).

### 3.4.3 Main Analyses: Faking Dissuasion Message Efficacy

To examine the efficacy of the faking dissuasion message, we conducted a MANOVA with job interest and faking dissuasion message condition as the independent variables and Blatant Extreme Responding, Idiosyncratic Item Responding, Communal Impression Management, and Residualized Change Scores as the dependent variables. The MANOVA revealed a significant main effect of job interest, \(F(4, 521) = 17.78, p < .001, \Lambda = .880\). Interested applicants scored significantly higher on Blatant Extreme Responding, Idiosyncratic Item Responding, Communal Impression Management, and Residualized Change Scores \((Ms = 8.62, 4.08, 3.27, \text{ and } .06, \text{ respectively})\) than did Uninterested Applicants \((Ms = 6.46, 3.62, 2.83, \text{ and } -.12, \text{ respectively})\). There were no effects of faking dissuasion message condition \(, F(16, 1592.32) = 1.17, p = .28, \Lambda = .97\). The interaction was also nonsignificant, \(F(16, 1592.32) = .74, p = .76, \Lambda = .98\). Examination of the means across the faking dissuasion message conditions revealed no hint of a significant effect for Blatant Extreme Responding, Idiosyncratic Item Responding, Communal Impression Management, or Residualized Change Scores. Means, standard errors, and confidence intervals are reported in Table 7.

### 3.4.4 Main Analyses: Faking Dissuasion Message Efficacy for Verifiable Items

To examine if our faking dissuasion messages were more effective for personality items that are job-related and verifiable, we conducted a univariate ANOVA with job interest and faking dissuasion message condition as the independent variable, and Residualized Change
Scores for our 9-item verifiability composite as the dependent variable. The results indicated that there was a main effect for job interest, $F(1, 524) = 61.52, p < .001, \eta^2_p = .11$. Again, Interested Applicants ($M = .09$) scored significantly higher than did Uninterested Applicants ($M = -.17$). However, there was no main effect for Faking dissuasion message condition and the interaction was also non-significant.

### 3.4.5 Main Analyses: Procedural Justice Perceptions and Test-Taking Anxiety

Our final research question was concerning how job applicants would react to our new faking dissuasion messages. To answer this question, we conducted a MANOVA with job interest and faking dissuasion message condition as the independent variables, and our applicant reaction measures including test-taking anxiety and perceptions of procedural justice as the dependent variables. The analyses revealed that there was a significant main effect of job interest, $F(4, 508) = 5.11, p < .001$, Wilk’s $\Lambda = .96$, and Faking dissuasion message condition, $F(16, 1552.60) = 2.30, p = .002$, Wilk’s $\Lambda = .93$. There was also a significant multivariate interaction between job interest and Faking dissuasion message condition, $F(16, 1552.60) = 1.78, p = .03$, Wilk’s $\Lambda = .95$. See Table 8 for estimated marginal means, standard errors, and confidence intervals.
Table 7. Faking by Job Interest and Faking dissuasion message condition

<table>
<thead>
<tr>
<th>Faking Measure</th>
<th>Job Interest</th>
<th>Faking dissuasion message condition</th>
<th>M</th>
<th>Std. Error</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BER</td>
<td>Yes</td>
<td>TFW</td>
<td>9.06</td>
<td>0.55</td>
<td>[7.98, 10.15]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>8.21</td>
<td>0.57</td>
<td>[7.09, 9.32]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>8.59</td>
<td>0.57</td>
<td>[7.46, 9.72]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>8.89</td>
<td>0.63</td>
<td>[7.64, 10.13]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>8.36</td>
<td>0.59</td>
<td>[7.20, 9.51]</td>
</tr>
<tr>
<td>BER</td>
<td>No</td>
<td>TFW</td>
<td>5.43</td>
<td>0.78</td>
<td>[3.90, 6.95]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>7.56</td>
<td>1.02</td>
<td>[5.55, 9.57]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>6.73</td>
<td>0.82</td>
<td>[5.13, 8.34]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>7.11</td>
<td>0.75</td>
<td>[5.64, 8.58]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>5.48</td>
<td>0.83</td>
<td>[3.85, 7.11]</td>
</tr>
<tr>
<td>IIR</td>
<td>Yes</td>
<td>TFW</td>
<td>4.13</td>
<td>0.07</td>
<td>[3.98, 4.27]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>4.02</td>
<td>0.08</td>
<td>[3.87, 4.17]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>4.12</td>
<td>0.08</td>
<td>[3.97, 4.27]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>4.20</td>
<td>0.08</td>
<td>[4.03, 4.37]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>3.92</td>
<td>0.08</td>
<td>[3.76, 4.07]</td>
</tr>
<tr>
<td>IIR</td>
<td>No</td>
<td>TFW</td>
<td>3.58</td>
<td>0.10</td>
<td>[3.38, 3.78]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>3.71</td>
<td>0.14</td>
<td>[3.44, 3.98]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>3.66</td>
<td>0.11</td>
<td>[3.45, 3.88]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>3.70</td>
<td>0.10</td>
<td>[3.50, 3.89]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>3.45</td>
<td>0.11</td>
<td>[3.23, 3.67]</td>
</tr>
<tr>
<td>BIMI-C</td>
<td>Yes</td>
<td>TFW</td>
<td>3.26</td>
<td>0.09</td>
<td>[3.09, 3.43]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>3.22</td>
<td>0.09</td>
<td>[3.04, 3.39]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>3.24</td>
<td>0.09</td>
<td>[3.07, 3.42]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>3.40</td>
<td>0.10</td>
<td>[3.21, 3.60]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>3.26</td>
<td>0.09</td>
<td>[3.08, 3.44]</td>
</tr>
<tr>
<td>BIMI-C</td>
<td>No</td>
<td>TFW</td>
<td>2.76</td>
<td>0.12</td>
<td>[2.52, 3.00]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>2.71</td>
<td>0.16</td>
<td>[2.40, 3.02]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>2.87</td>
<td>0.13</td>
<td>[2.62, 3.12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>2.98</td>
<td>0.12</td>
<td>[2.75, 3.21]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>2.84</td>
<td>0.13</td>
<td>[2.59, 3.10]</td>
</tr>
<tr>
<td>RCS</td>
<td>Yes</td>
<td>TFW</td>
<td>0.09</td>
<td>0.03</td>
<td>[0.03, 0.15]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>0.21</td>
<td>0.03</td>
<td>[-0.04, 0.08]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>0.05</td>
<td>0.03</td>
<td>[-0.01, 0.11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>0.09</td>
<td>0.03</td>
<td>[0.03, 0.16]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>0.05</td>
<td>0.03</td>
<td>[-0.01, 0.11]</td>
</tr>
<tr>
<td>RCS</td>
<td>No</td>
<td>TFW</td>
<td>-0.16</td>
<td>0.04</td>
<td>[-0.24, -0.08]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IAW</td>
<td>-0.05</td>
<td>0.06</td>
<td>[-0.16, 0.06]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FAW</td>
<td>-0.13</td>
<td>0.04</td>
<td>[-0.22, -0.05]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MSA</td>
<td>-0.07</td>
<td>0.04</td>
<td>[-0.15, 0.01]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NW</td>
<td>-0.18</td>
<td>0.05</td>
<td>[-0.26, -0.09]</td>
</tr>
</tbody>
</table>

Note. This table displays the estimated marginal means, standard errors, and 95% confidence intervals for BER, BIMI-C, and RCS by Job Interest and Faking dissuasion message condition. These results were generated from a MANOVA that revealed only a significant main effect of Job Interest.
Table 8. Applicant Reactions by Job Interest and Faking dissuasion message condition

| Condition | Interested Applicants | | | | | | | | Uninterested Applicants | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Test-Taking Anxiety | PJ – Job Relatedness | PJ – Information Known | PJ – Chances to Perform |  | Test-Taking Anxiety | PJ – Job Relatedness | PJ – Information Known | PJ – Chances to Perform |  | M | SE | 95% CI | M | SE | 95% CI | M | SE | 95% CI | M | SE | 95% CI | M | SE | 95% CI |
| TFW | 3.17 | 0.13 | [2.91, 3.42] | 3.05 | 0.11 | [2.83, 3.28] | 3.49 | 0.11 | [3.27, 3.71] | 2.81 | 0.12 | [2.57, 3.05] | 3.32 | 0.18 | [2.97, 3.67] | 2.96 | 0.16 | [2.65, 3.27] | 3.26 | 0.15 | [2.96, 3.57] | 2.60 | 0.17 | [2.27, 2.94] |
| IAW | 3.19 | 0.13 | [2.94, 3.45] | 3.09 | 0.12 | [2.86, 3.31] | 3.42 | 0.11 | [3.20, 3.64] | 2.61 | 0.12 | [2.36, 2.85] | 3.73 | 0.23 | [3.27, 4.19] | 2.52 | 0.21 | [2.12, 2.93] | 3.07 | 0.20 | [2.68, 3.47] | 2.57 | 0.22 | [2.13, 3.00] |
| FAW | 3.05 | 0.13 | [2.79, 3.31] | 3.11 | 0.12 | [2.88, 3.34] | 3.51 | 0.11 | [3.28, 3.73] | 3.13 | 0.13 | [2.88, 3.38] | 2.83 | 0.19 | [2.45, 3.20] | 2.54 | 0.17 | [2.21, 2.88] | 3.16 | 0.17 | [2.83, 3.48] | 2.36 | 0.18 | [2.00, 2.72] |
| MSA | 3.10 | 0.15 | [2.81, 3.39] | 3.05 | 0.13 | [2.80, 3.31] | 3.44 | 0.13 | [3.19, 3.69] | 3.00 | 0.14 | [2.73, 3.28] | 2.83 | 0.17 | [2.50, 3.16] | 2.85 | 0.15 | [2.55, 3.15] | 3.21 | 0.15 | [2.92, 3.50] | 2.70 | 0.16 | [2.38, 3.02] |
| NW | 3.39 | 0.13 | [3.13, 3.66] | 2.85 | 0.12 | [2.61, 3.08] | 3.23 | 0.12 | [3.00, 3.46] | 2.64 | 0.13 | [2.39, 2.90] | 3.77 | 0.19 | [3.40, 4.14] | 2.34 | 0.17 | [2.01, 2.67] | 3.28 | 0.16 | [2.96, 3.60] | 2.59 | 0.18 | [2.23, 2.94] |

Note. This table displays the estimated marginal means, standard errors, and 95% confidence intervals for each applicant reaction measure by job interest and faking dissuasion message condition. PJ = Perceptions of Procedural Justice. These results were generated from a MANOVA analysis.
Univariate analyses revealed significant main effects for all three perceptions of procedural justice subscales including Job-Relatedness, $F(1, 511) = 17.20, p < .001, \eta_p^2 = .03$, Information Known, $F(1, 511) = 6.03, p = .01, \eta_p^2 = .01$, and Chances to Perform, $F(1, 511) = 7.51, p = .006, \eta_p^2 = .01$. Interested Applicants rated the personality test as having higher levels of Job-Relatedness, Information Known, and Chances to Perform ($Ms = 3.03, 3.42, \text{ and } 2.84$) than did Uninterested Applicants ($Ms = 2.64, 3.20, \text{ and } 2.56$, respectively). Univariate analyses revealed significant main effects of faking dissuasion message condition on test-taking anxiety, $F(4, 511) = 6.06, p < .001, \eta_p^2 = .05$, and perceptions of job relatedness, $F(4, 511) = 2.49, p = .045, \eta_p^2 = .02$. Pairwise comparisons of the estimated marginal means suggest that applicants in the FAW condition experienced significantly less test-taking anxiety ($M = 2.94$) than applicants in the IAW ($M = 3.46$) or NW ($M = 3.58$) conditions. These tests showed that applicants in the TFW condition had higher perceptions of job-relatedness ($M = 3.01$) than did applicants in the NW condition ($M = 2.60$). None of the univariate interaction tests were significant.

3.5 Discussion

Some studies demonstrated that faking dissuasion messages are one of the most effective tools to combat applicant faking on non-cognitive tests (Fan et al., 2012; Landers et al., 2011). However, the results from our study suggest that faking dissuasion messages might not be effective in all situations—as all four faking dissuasion messages did not reduce faking or influence applicant reactions.

As a partial test of our manipulations, we found that the time applicants spent reading the faking dissuasion messages corresponded to the length of those warnings—spending longer on lengthy warnings (e.g., IAW, FAW) and less time on the shorter ones (NW, TFW). The main
effect was also large ($\eta_p^2 = .27$), suggesting the faking dissuasion message condition was a primary determinant of viewing time. We also found that interested applicants had significantly more favorable personality scores than did uninterested applicants. Interested applicants also scored significantly more favorably when they thought there was a job (Time 1) than after they learned the job was fictitious (Time 2). These results suggest that our interested applicants faked to appear more favorable. In sum, interested applicants appeared to be attentive, indicated interest in the job, and faked accordingly, indicating that the failure of the faking dissuasion messages may not be due to our study design.

In their meta-analysis on faking dissuasion messages, Dwight and Donovan (2003) concluded that effective faking dissuasion messages must inform test-takers that faking can be identified and that faking will lead to consequences, such as being withdrawn from the selection process. Their recommendations have been successful in some studies (e.g., Fan et al., 2012). In our study, both the IAW and FAW explicitly informed applicants how faking could be identified and that being detected could lead to being withdrawn from the selection process. Yet, none of our faking dissuasion messages—including the TFW used in earlier studies—significantly reduced faking. There was also no evidence that the MSA was effective.

Job applicants are likely to fake when they believe it is advantageous for them to do so, and will not fake when they believe it will lead to negative consequences, such as being removed from the hiring pool (Goffin & Boyd, 2009). However, what about situations where the applicant believes they have little chance of being hired? In the current study, all applicants were told that they we were working with a company that would contact applicants with desirable results. One possibility is that our interested applicants viewed the situation as a competitive screening scenario with hundreds of other applicants, where they had little to lose by faking. Applicants
may have believed that the benefits of faking outweighed the risks of being caught faking. By contrast, applicants may perceive personality tests differently when encountered later in the selection process, when they have more to lose by being caught for faking. Thus, the utility of faking dissuasion messages may depend on job applicants’ perceptions of how competitive the selection process will be.

To probe this explanation, we conducted a post-hoc review of faking dissuasion message studies since the last published meta-analysis (Dwight & Donovan, 2003). As reviewed in Table 9, we found that these studies could be categorized into three methodological categories. In the first category, “Accusation”, faking dissuasion messages were successful in reducing faking when job applicants were told in real-time that they had been identified for faking based on how they had scored on preliminary personality and/or social desirability scales. These studies were instrumental in showing how technology can be used to bolster the efficacy of faking dissuasion messages. However, a primary concern with this approach is that it requires the provision of unique test instructions to only a subset of the applicant pool. More concerning, these warnings are generated based on response patterns that have not been fully validated in the faking literature. For example, Fan et al. (2012) issued faking warnings based on social desirability scores, which have been widely criticized in the literature (Griffith & Peterson, 2008). Issuing different test instructions, let alone when based on questionable criteria, may open an organization to potential legal challenges (Catano, Wiesner, & Hackett, 2010). This concern is amplified further if the warning is presented disproportionately to protected groups, which was not investigated in any of the studies in the Accusation category.
<table>
<thead>
<tr>
<th>Authors</th>
<th>Sample</th>
<th>Warning Design</th>
<th>Faking Design</th>
<th>Faking Assessment</th>
<th>Outcome</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present Study</td>
<td>349 mTurk applicants who believed they were applying for a job</td>
<td>Identification and consequences; Moral Suasion</td>
<td>Fake job that applicants thought was real</td>
<td>Residualized Change Score, Idiosyncratic Item Responding, Blatant Extreme Responding, Impression Management</td>
<td>Ineffective</td>
<td>Screening</td>
</tr>
<tr>
<td>Fisher et al. (2018)</td>
<td>275 Undergraduate Business Students</td>
<td>Identification and consequences</td>
<td>$20 for highest scorer; Fake Good</td>
<td>Difference Scores</td>
<td>Ineffective</td>
<td>Screening</td>
</tr>
<tr>
<td>Burns et al. (2015)</td>
<td>457 Undergraduate Students</td>
<td>Identification, consequences, Accusation</td>
<td>Repeated measures; Fake Good for Cash Prize</td>
<td>Mean personality scores controlled for “pre-warning” personality</td>
<td>Small effect sizes, with exception for “accusation” condition</td>
<td>Experimental Accusation &amp; In-test Accusation</td>
</tr>
<tr>
<td>Fan et al. (2012): Study #1</td>
<td>157 applicants for 10 staff positions at university</td>
<td>Accusation based on test scores</td>
<td>Single session, real applicants</td>
<td>Bogus Items and Social Desirability</td>
<td>Faking warning highly effective</td>
<td>In-test Accusation</td>
</tr>
<tr>
<td>Fan et al. (2012): Study #2</td>
<td>386 applicants for a student study group</td>
<td>Accusation based on test scores</td>
<td>Single session, real applicants</td>
<td>Bogus Items and Social Desirability</td>
<td>Faking warning highly effective</td>
<td>In-test Accusation</td>
</tr>
<tr>
<td>Landers et al. (2011)</td>
<td>16,779 internal applicants for promotion for initial sample; 488 applicants for retesting</td>
<td>Accusation based on test scores</td>
<td>Repeated sessions</td>
<td>Blatant Extreme Responding</td>
<td>Large initial effects, weakened overtime</td>
<td>In-test Accusation</td>
</tr>
<tr>
<td>Study</td>
<td>Sample Size</td>
<td>Methodology</td>
<td>Identified and Consequences</td>
<td>Measures</td>
<td>Findings</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------</td>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
<td>--------------------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Robie et al. (2009)</td>
<td>329 undergraduate students</td>
<td>Single session, cash prize “fake good”</td>
<td>Mean conscientiousness score; agreement with observer scores</td>
<td>Large effect</td>
<td>Experimental</td>
<td></td>
</tr>
<tr>
<td>Robson et al. (2008)</td>
<td>464 undergraduate students</td>
<td>Single session, cash prize “fake good”</td>
<td>Mean personality scores; Agreement with observer ratings</td>
<td>Medium to large effects; weakened convergent validity</td>
<td>Experimental</td>
<td></td>
</tr>
<tr>
<td>Converse et al. (2008)</td>
<td>293 undergraduate students</td>
<td>Single session, cash prize “fake good”</td>
<td>Faking not reported by warning condition; focused on validity outcomes</td>
<td>No improvement in convergent validity or incremental validity</td>
<td>Experimental</td>
<td></td>
</tr>
<tr>
<td>Vasilopoulos et al. (2005): Study #1</td>
<td>366 applicants for a federal law enforcement position</td>
<td>Real Applicants</td>
<td>Mean personality scores</td>
<td>Null to small effects</td>
<td>Screening</td>
<td></td>
</tr>
<tr>
<td>Vasilopoulos et al. (2005): Study #2</td>
<td>124 undergraduate students</td>
<td>Single session, cash prize “fake good”</td>
<td>Mean personality scores; Impression Management</td>
<td>Null to medium effects; no significant effect for Impression Management</td>
<td>Experimental</td>
<td></td>
</tr>
</tbody>
</table>
In the second category, “Experimental”, faking dissuasion messages tended to be successful. A major concern with this category is that lab scenarios produce significantly more faking than is found in field studies (Hough, 1998), and as a result, faking warnings have far more potential to decrease faking in simulation studies than in field environments. This is problematic because faking warnings may be able to produce artificially high effects sizes that would not be replicated with more realistic levels of faking. Even more problematic, participants in simulated studies may be subject to demand effects when they read warnings and constrict their faking more than a job applicant would. It may be apparent to some participants that the experimenter wants them to fake less. As a result, these studies may have oversold the potential for general, non-accusatory, faking warnings to thwart applicant faking.

In the third category, “Screening”, there are applicant scenarios with a high motivation and minimal downside to faking—such as during screening—where the applicant may view their chances of the desired outcome as small. For example, in a recent study (Fisher et al., 2018), undergraduate students were told that only the highest scorer would receive a cash prize. This low selection ratio meant that applicants had nothing to lose by being detected for faking, since they were likely aware that faking was their only chance to secure the cash prize. Similarly, in the present study, applicants were explicitly presented with a screening opportunity, where only high scorers would be contacted by the hiring organization. In our study, there was little to lose by being detected for applicant faking. In a job screening with real job applicants for a federal government position in the United States, Vasilopoulos et al. (2005) found significant, but small reductions due to faking dissuasion messages ($ds = .07$ to $.19$) overall. They also found that faking dissuasion messages were only effective for applicants with lower general mental ability—applicants with higher general mental ability were unaffected by the faking warning.
Taken together, these findings present some reason to caution the value of faking dissuasion messages as an effective tool in combatting applicant faking on personality tests.

Interestingly, the FAW appeared to reduce test-taking anxiety when compared to the IAW and NW, but not the TFW and MSA. One possibility is that the FAW reassured applicants that they would have the opportunity to demonstrate the personality on the job, providing them a sense of control over the outcome. By contrast, the IAW was threatening and placed the outcome on references and tricky test questions, which may have increased test-taking anxiety. The NW group was given no information about how test questions were verified and may also have provoked some test-taking anxiety. These results provided some preliminary support for including a preamble of how personality items are verified using future job behavior, as applicants may believe they are capable of performing those behaviors and it relieves them of the worry that they will be filtered using other approaches. By contrast, threatening warnings that place the outcome on other sources such as references or internal “integrity checks” may provoke test-taking anxiety. Similarly, providing no information about how information is verified may also allow applicants to ponder how their responses will be verified—increasing test-taking anxiety. The TFW increased perceptions of job-relatedness relative to the other faking dissuasion message conditions. One possibility is providing vague or general information about test verification actually appeared the most credible, whereas the other faking dissuasion messages were less compelling. That being said, the difference was small and may reflect an underlying difference in the conditions.

3.5.1 Limitations

This is a novel investigation that has provided some unique insight into the efficacy of faking dissuasion messages. However, there are a couple of limitations. The primary limitation is
that we examined the efficacy of faking dissuasion messages using a low-complexity job with mTurk users, and our findings may not generalize to more executive positions. That being said, mTurk often provide high quality data when rigorous data screening are conducted (Buhrmester et al., 2011; Casler et al., 2013; Landers & Behrend, 2015). The second limitation is that we only presented the faking dissuasion messages during initial screening, and our post-hoc literature review suggests that the timing of presentation is likely more important than the content of the faking dissuasion messages. Future studies will need to examine this more closely by manipulating when the faking dissuasion messages are presented.

3.5.2 Conclusion

Faking dissuasion messages may be effective when applicants are reporting factual and objectively verifiable information or when applicants truly believe that faking may cost them the job opportunity (Fan et al., 2012; Landers et al., 2011). However, in our study, faking dissuasion messages seem to have limited effect in job applications even when those items are rated as non-verifiable or when applicants believe the consequences of faking are outweighed by the benefits (e.g., early screening scenarios). This is problematic, since researchers have argued that personality tests should be used during screening rather than during final selection, in order to minimize the impact of applicant faking (Mueller-Hanson et al., 2003). Yet, this may be the type of situation when applicants are most likely to fake their responses. Future research will need to examine, more directly, if the efficacy of faking dissuasion messages is due to their content or to the characteristics of the assessment situation.
3.6 References


Appendix A: Traditional Faking Warning

**Workstyle Survey: Important Instructions**

The test you are going to complete calls for your honest responses. Be aware of the following two points:

1. The test contains questions that are designed to identify those who attempt to fake their responses. 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 may invalidate your results. Also, faking might result in you not being considered for the position.
Appendix B: Immediate Authentication Warning

Workstyle Survey: Important Instructions

You are about to fill out a work-style assessment that will ask you questions about things such as how friendly or outgoing you are at work. We’d just like to let you know that we can verify responses on the following work-style assessment to ensure everyone is being honest. This allows us to hire the most well-suited person for the job.

We use two methods to verify answers.

**Method 1: Reference Checking.**

- You’ll be asked to provide personal references. That is, people who are familiar with your general work-style. This could include employers, coworkers, or clients.
- A trained consultant can verify with your references, aspects of your work-style that can be observed by others.
  - For example, the consultant may call your references and ask about how talkative you are in large groups.
- If so, the trained consultant will look for consistent differences between your answers on the work-style assessment and the responses provided by your references.
- If we find many differences between your scores on this assessment and the answers provided by your references, this will lead to a one-on-one interview with the consultant.
- The consultant will request that you explain why discrepancies exist and defend your answers to any question where a discrepancy exists.
Method 2: Internal Integrity Checks.

We have included questions that assess whether you tend to provide answers that are unlikely to be honest. These questions have been developed and vetted by scientists and experts. For example, imagine you were asked, “How often do you follow the rules?” Your first thought might be that you ALWAYS follow the rules. However, have you ever broken even a small rule for a coworker or customer? Have you ever broken a rule you that you thought was outdated or served no real purpose?

- These questions are inserted throughout the work-styles assessment. So please think carefully before answering each item in this assessment, and answer honestly.
- If you are detected as answering dishonestly, this will lead to a one-on-one interview with the consultant. Here, you will need to explain your answers.

If you answer dishonestly and are unable to defend your answers, you will NOT get the job. However, if you answer honestly, you will have nothing to worry about.

Now, please provide two work references who we may contact. This can include previous employers, coworkers, or clients:

<table>
<thead>
<tr>
<th>Reference 1</th>
<th>Name</th>
<th>Phone Number</th>
<th>E-Mail Address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reference 2</th>
<th>Name</th>
<th>Phone Number</th>
<th>E-Mail Address</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Workstyle Survey: Important Instructions

You are about to fill out a work-style assessment that will ask you questions about things such as how friendly or outgoing you are at work. We’d just like to let you know that we are looking for applicants who consistently demonstrate certain workstyles and that we can verify responses on the work-style assessment to make sure everyone is being honest. This allows us to hire the most well-suited people for the job. It also reduces the chances someone is hired by answering dishonestly on the assessment. We can do this using a two-step process.

First step. If you are shortlisted and interviewed, the interviewer will make observations of some aspects of your work-style during the interview, such as how friendly or outgoing you are.

- A trained consultant will look for consistent differences between your answers on the work-style assessment and those provided by others.
- Many differences between your answers and the interviewers ratings will lead to a one-on-one interview with the consultant.
- The consultant will request that you explain why discrepancies exist between your responses and those provided by others.

Second step. If you are hired, your supervisor and peers will examine how your on-the-job workstyle relates to your job performance.

- If you are flagged as performing poorly in some aspects of your job because of your workstyle, you will be referred to a one-on-one interview with a qualified HR consultant to defend your original answers.
In both steps, if the consultant determines that you provided insufficient rationale for your answers on this workstyle survey, you will be removed from the hiring process or terminated immediately.

You should also think about the fact that lying during the work-style assessment may get in the way of your long-term interests.

- You will be unhappy if you get a job that requires a different work-style than your own
- In turn, you will likely perform poorly and not earn promotions or pay raises.

Thus, it is in your long-term best interest to answer the work-style assessment honestly.
Appendix D: Moral Suasion Appeal

**Workstyle Survey: Important Instructions**

You are about to fill out a work-style assessment that will ask you questions about things such as how friendly or outgoing you are at work. Please consider that providing dishonest answers is immoral. Doing this violates what most people consider moral behavior. Honest responses ensure that the most well-suited candidates are hired.

Now we want you to imagine yourself in the following situation. Think carefully about how you would **FEEL**:

- Imagine you’re a customer support worker for an internet company.
- The company just had a major internet outage
- The company is receiving more calls than usual.
- The customers are furious. Your supervisors are increasingly demanding.
- The customer queue is over 2 hours.
- Your boss expects you to work extra shifts. Your time off is cancelled.
- You’ve had to cancel plans to visit your family whom you haven’t seen in several years.
- Later, you learn the internet outage is because of a careless technician. He cut a major cable line. Instead of reporting his mistake, he simply covered it up. He hoped no one would discover that he was to blame.
- Now imagine that this technician only got the job because he answered dishonestly on a work-style measure such as this one. How upset would you be?

**In sum, honest responses ensure that the most well-suited candidates are hired.**
Remote Customer Service Representative (Job Posting)

We're looking for a Remote Customer Service Representative!

We are a research team that works with a large international firm. This firm is looking for remote customer service representatives who will help customers with online inquiries and hold live chat sessions. If you are interested, your answers to our survey will be used as a preliminary screening for employment. The candidates with the strongest person-job fit will be considered for an interview and an employment opportunity.

This person will be responsible for:

- Communicating with customers by live web chat session or by e-mail to provide information about products or services, take or enter orders, cancel accounts, or obtain details of complaints.
- Following up with purchases to ensure that customers are satisfied and offering individualized promotions
- Diligently check to ensure that appropriate changes were made to resolve customers' problems
- Keeping records of customer interactions or transactions, recording details of inquiries, complaints, or comments, as well as actions taken.
- Completing contract forms, prepare change of address records, or issue service discontinuance orders, using remote web interface.
- Referring unresolved customer grievances to designated departments for further investigation
• Determining charges for services requested, collect deposits or payments, or arrange for billing

• Contacting customers remotely to respond to inquiries or to notify them of claim investigation results or any planned adjustments

**Consider applying if you:**

• Enjoy working from home with flexible hours.

• You enjoy communicating with customers and making customers feel valued

• You like writing e-mails and live online conversations

• Enjoy problem solving and helping customers with novel solutions

• Developing individualized promotions for previous customers and following up on previous purchases

**What We Offer:**
- Contractual services that pay you for answering customer e-mails and following up with customers
- The opportunity to learn and develop in a growing company
- Access to Health and Dental Benefits
- A profit-sharing program
- A positive and respectful work environment

If you are interested in this opportunity, your answers will be used for both research purposes and for employment screening. You will also still receive payment for completing the HIT.

If you are not interested in this opportunity, your answers will only be used for research purposes and you are still eligible to complete the HIT.

Are you interested in this opportunity?

- [ ] Yes, I am interested being considered for the remote customer service position and also receiving payment for completing the HIT

- [ ] No, I am not interested in this additional opportunity.
Chapter 4

4 Combating Faking on Personality Tests during Military Recruitment

In this chapter, we investigate whether new faking dissuasion messages that incorporate accountability theory and morality theory can be used to reduce recruit faking on personality tests in a military environment, and whether reductions in applicant faking will come at the expense of negative applicant reactions.

4.1 Abstract

Military organizations use personality testing to select recruits that will perform better and who be less likely to leave the organization. However, military recruits fake on these tests to manipulate their chances of recruitment (Boss et al., 2015). We investigated the efficacy of three different faking dissuasion message conditions to reduce faking among military recruits that emphasize short-term accountability, long-term accountability, and morality. We also examined how these new warnings compare to a traditional warning and a no-warning control group. We tested 466 basic training recruits at the Canadian Armed Forces and asked them to engage in a selection simulation. We assigned groups of recruits to the different faking dissuasion message conditions and walked classes through the simulation. The results suggested that only the faking dissuasion message condition that emphasized short-term accountability, which threatened to detect fakers by contacting references and through the use of social desirability items, reduced applicant faking. None of the other messages had any effect when compared to a no-warning control group. We also found no differences in applicant reactions to the faking dissuasion messages. Implications for research and practice are discussed.

Keywords
Personality assessment; job applications; faking; faking warnings; military; recruitment
4.2 Introduction

Envision that you are a captain, commanding a platoon, fighting in a war-torn and treacherous area of a war-torn country. You have been captured by a group of insurgents and your platoon is planning a rescue mission. One of your trusted officers is about to take the lead in your absence. In this hypothetical situation, do you hope the officer is conscientious—detail-oriented, diligent, dutiful—or the opposite? This situation makes one thing clear: personality matters when selecting military recruits. But, what if a military recruit was selected for a critical position because they faked on the personality test, and were not actually suited for that position?

The purpose of this study was to examine the extent that faking dissuasion messages could reduce faking on personality testing by military recruits during the selection process. In this context, military recruits are to the military as job applicants are to civilian organizations. We developed three faking dissuasion messages that leveraged accountability (Lerner & Tetlock, 1999) and morality theory (Haidt, 2001)—with the intent of improving military selection decisions, where military recruits may eventually be placed in life and death situations.

Personality testing predicts important outcomes in military settings, including task performance, counterproductive work behavior, and leadership training (Bartone, Snook, & Tremble, 2002; Darr, 2011; Darr & Catano, 2016). In a meta-analysis, both Conscientiousness and Emotional Stability positively predicted military task performance (ρs = .35 and .28 respectively) and negatively predicted counterproductive work behavior (ρs = .28 and -.33 respectively; Darr, 2011). The researchers also found that Extraversion predicts task performance (ρ = .22) and that Agreeableness predicts counterproductive work behavior (ρ = -.21). In another investigation, Bartone et al. (2002) found that Conscientiousness incrementally predicted performance beyond general mental ability, gender, and social judgment in military leadership.
training. Overall, personality appears to be a strong predictor of military performance relative to other selection tools for lower-ranked military recruits (Darr & Catano, 2016).

Personality testing can also be used to identify military personnel who are more likely to leave the organization (Darr, 2011). Both Conscientiousness and Emotional Stability negatively predict turnover intentions (\( \rho_s = -0.28 \) and \(-0.32\) respectively)—meaning that people who rate more highly on these personality traits are less likely to leave their current career for another one. Personality traits can also predict which high school graduates are most likely to join the armed forces. In another study (J. J. Jackson et al., 2012), researchers found that military recruits were lower on Agreeableness \((d = -0.29)\) and Openness \((d = -0.15)\) than civilians, but higher on Emotional Stability \((d = 0.14)\). Taken together, recruit personality scores can be used to form prediction models of who will apply to the armed forces, and that the same traits—Conscientiousness and Emotional Stability—can be used to select and retain recruits who are a good fit for military service.

### 4.2.1 Selecting Military Recruits: Prevalence of Faking

Research suggests that military recruits fake during the selection stage. In a study of 918 conscripted recruits for the Swiss Armed Forces (Boss et al., 2015), self-reported military service motivation predicted personality scores during initial recruitment: Conscientiousness \((r = 0.50)\), Extraversion \((r = 0.45)\), Team Cooperation \((r = 0.29)\), and Stress Tolerance \((r = 0.51)\). Boss et al. then divided the recruits by their admission of faking. Not surprising, correlations between military service motivation and personality was highest for those who admitted to “faking good” and lowest for those who admitted to “faking bad.” Taken together, recruits who were unmotivated to serve in the military “faked bad” by intentionally scoring unfavorably—with the hope of being declared “unfit to serve”—and recruits who were motivated to serve “faked good”
by intentionally scoring favorably. This is troubling for military organizations that rely on voluntary service—such as Canada or the United States—because most recruits will be motivated, and in turn, be willing to fake good to join the armed forces. This means that the motivated recruits who are selected may be a poorer fit for service.

4.2.2 Reducing Recruitment Faking

There is little research on how the military can combat faking by recruits, but they can draw inferences from research on regular job applicants. The broader faking literature reveals that faking dissuasion messages are one of the simplest and most effective tool to thwart applicant faking in non-military settings; reducing faking by up to 50% (Fan et al., 2012; Landers et al., 2011). The Canadian Armed Forces (CAF) created their faking dissuasion message to utilize best practice, which includes the threats of identification and disqualification from the selection process (Dwight & Donovan, 2003). However, to our knowledge, the efficacy of faking dissuasion messages in reducing faking among military recruits has not been empirically tested.

There are some unique aspects to military recruits that may cause them to respond differently to faking dissuasion messages than civilian applicants. First, at least in Canada, the military rejects a low proportion of recruits (Office of the Auditor General, 2006). Theoretical models of faking suggest that applicants are likely to weigh the risks of being caught against the dividends provided by faking (Goffin & Boyd, 2009). That is, interested applicants may perceive a faking as advantageous in a competitive climate; however, in the military context, this means that recruits have less incentive to fake than would civilian applicants to more competitive positions. Assuming this is true, faking dissuasion messages that stress the risks of faking—identification and disqualification from the hiring process—should be a more serious deterrent to
military recruits than civilian applicants. Second, the military has a strong masculine culture that stresses conformity and loyalty (Dunivin, 1994). We speculate that faking dissuasion messages may carry more weight in a high conformity culture than for a civilian applicant to a regular organization.

Therefore, we examine the extent that the CAF’s faking dissuasion message is able to reduce faking among military recruits relative to a no-warning control group. If faking dissuasion message is effective, military recruits should receive less favorable personality scores than military recruits in the control group. Because the CAF warning is modelled off the existing literature, we refer to the CAF’s faking dissuasion message as a Traditional Faking Warning (TFW).

In addition, building on Study 2, we examined the effectiveness of two new faking dissuasion messages that increased accountability by describing a credible process by which personality scores could be verified and by making military recruits focus on the process of responding to the test, rather than the outcome (Lerner & Tetlock, 1999). The Immediate Authentication Warning (IAW) informed military recruits that the personality test contained test items that could detect faking and provided examples of an impression management scale. The IAW also informed military recruits that their responses to personality items could be verified by calling references that they provided during initial recruitment. Together, the IAW described two verification mechanisms to boost credibility that faking could be detected. The IAW also informed military recruits that if they were identified as faking, they would need to defend their responses to each personality item to a trained Recruiting Officer from the CAF. By having recruits consider if they could defend their answers to each personality item, in theory, they should fake less in cases where they would struggle to defend their choices.
The Future Authentication Warning (FAW) explained to recruits that their Recruitment Officer would compare their observed personality during training to their answers on the personality test. The FAW attempted to build credibility by illustrating that the personality test is assessing behaviors that can be verified by observing them. The FAW explained that any recruit identified as faking would need to defend their responses to each personality item to the Recruitment Officer. Military recruits were told that failure to do so would lead to disqualification from the selection process. This feature also aimed to prompt military recruits to carefully consider if they could defend their answers to each personality item rather than their overall test scores. Ideally, this should reduce faking. Finally, the FAW also included an educational component. It explained to recruits that their personality scores would be used for placement decisions and that faking may lead to placement in a position that was a poor fit for the recruit. We explained that being placed in a position with poor fit would lead to lower performance, and in turn, fewer opportunities for promotion.

Scholars have also postulated that morality is the first issue that test-takers consider when faking on personality test items, suggesting that “test-taking instructions that appeal to the test takers’ moral compass (by emphasizing that faking is a form of lying or cheating that violates most accepted standards of moral behaviour) might add to the success of existing faking warnings.” (Goffin & Boyd, 2009, p. 158). We tailored a moral suasion appeal (MSA) to the armed forces by having recruits consider a life-threatening combat situation. Then, we had recruits imagine how they would feel if an intelligence officer had information about that attack and forgot to relay the information to their commander—not from malice, but carelessness—because that person was placed into their position by faking. The goal of the MSA was to prompt recruits to feel negatively about faking, and in turn, have them answer more honestly. Moral
decisions are generally informed by affective responses rather than rational decision-making (Haidt, 2001). Thus, we also included pictures to facilitate thinking about the combat scenario to maximize the emotional response to the simulation (Lang et al., 1993; Schimmack, 2005).

**Research Question 1**: To what extent will recruits fake on a personality test when presented with faking dissuasion messages (IAW, FAW, MSA) as compared to no-warning control group?

### 4.2.3 Recruit Reactions to Faking Dissuasion Messages

Prior research suggests that faking dissuasion messages with threatening or negative wording increases test-taking anxiety (Converse et al., 2008). These messages also increase the cognitive demands of applicants (Vasilopoulos et al., 2005), making the personality tests less pleasant to complete. Subsequently, the faking dissuasion messages may cause negative reactions concerning the fairness or legitimacy of the personality test. Recruit reactions are particularly important because negative impressions can reduce the likelihood a candidate accepts job offers or recommends the organization to others (McCarthy, Bauer, Truxillo, Anderson, et al., 2017; McCarthy, Bauer, Truxillo, Campion, et al., 2017). The latter is of particular concern to the CAF, because they have documented problems recruiting and retaining talent (Office of the Auditor General, 2006). Each of the faking dissuasion messages alter the extent they increase accountability, educate recruits, and appeal to recruit morality, so the different faking dissuasion messages may invoke unique applicant reactions.

**Research Question 2**: To what extent will the faking dissuasion messages influence test-taking anxiety and perceptions of procedural justice when compared to the no-danger control group?
4.3. Method

4.3.1 Participants

We invited 535 recruits from the Canadian Armed Forces who were undergoing basic training to take part in our study. A total of 466 (87%) recruits agreed to participate. Consenting recruits provided their service numbers so that we could obtain demographic information and earlier test scores (e.g., cognitive ability) from their service files ($M_{age} = 24.50, SD = 5.59, 88$ females, $94.9\%$ regular force, and $5.1\%$ reserve). All recruits spoke English as their primary language except three, whose primary language was French but who were also fluent in English.

4.3.2 Measures

**Personality.** We measured personality using three measures. The first measure, used during initial recruitment, was the Trait Self-Descriptive Personality Inventory (TSD-PI; Darr, 2011). The TSD-PI is the CAF’s proprietary measure of the Big Five personality dimensions and has 75-items that examine Agreeableness, Conscientiousness, Emotional Stability, Extraversion, and Openness. Each item is measured using a 7-point scale ($1 = “Extremely Uncharacteristic”$ to $7 = “Extremely Characteristic”$). The TSD-PI is reliable ($\alpha$s $= .88$ to $.93$) and each of the five dimensions correlate with their respective dimension on the NEO (Darr, 2009). Overall correlations were not provided in the TSD-PI manual, but each item was validated against their respective NEO marker. The second measure, used during our experiment, was the 20-Item OCEAN.20 (O’Keefe, Kelloway, & Francis, 2012) developed in conjunction with the Canadian Armed Forces, which was derived from the full version of their proprietary TSD-PI (Darr, 2009, 2011). We included this measure for parsimony with the original TSD-PI. Accordingly, items used the same 7-point response scale. The scale is reliable ($\alpha$s $= .74$ to $.83$) and each of the Big 5 personality dimensions converges with their respective TSD-PI markers ($r$s $= .77$ to $.88$). Finally,
for consistency with Studies 1 and 2, we also measured the Big 5 personality dimensions using the International Personality Item Pool (IPIP; Goldberg et al., 2006). Similarly, we used the 120-item version (Maples et al., 2014), that uses a 5-point response scale (1 = *Strongly Disagree*, 5 = *Strongly Agree*). The 120-item scale has strong internal consistencies between .87 and .90 for the five dimensions. Additionally, all five dimensions on the 120-item scale were found to parallel their respective NEO Personality Inventory dimension with correlations between .87 and .90 (Maples et al., 2014).

**Faking.**

In this study, we were unable to use Residualized Change Scores, which have been demonstrated as the best faking assessment (Feeney & Goffin, 2015). Residualized Change Scores require a set of job applicant and non-applicant personality scores for the same people, which we were unable to obtain for this study. Instead, we used three practical assessments of applicant faking that were shown to demonstrate both Between-Subjects Faking Sensitivity and Within-Subjects Faking Sensitivity in Study 1.

**Blatant Extreme Responding.** We assessed faking using Blatant Extreme Responding (BER). We used the same approach as in Study 1, which required recoding personality items so that the most favorable answer for a given item is coded as 1, and the remaining answers are coded as 0. A composite BER score is calculated by totaling the number of extreme responses across all personality items used in the analysis. BER has been used in multiple investigations and has been demonstrated as an effective measure of applicant faking (Landers et al., 2011; Levashina et al., 2014). We derived two separate measures of BER from the OCEAN.20 and IPIP personality measures.
**Communal Impression Management.** We measured Communal Impression Management, which is the extent that applicants engage in “denying socially deviant impulses and claiming pious attributes” (Blasberg et al., 2013, p. 523). Communal Impression Management is assessed using 10 items measured on 5-point response scales (1 = “Strongly Disagree” to 5 = “Strongly Agree”). An example item is “I don’t gossip about other people’s business.” Internal consistencies exceeded .70 across two studies, and there is evidence of both convergent and discriminant validity (Blasberg et al.). These items were embedded randomly throughout the Personality assessment.

**Idiosyncratic Item Responding.** We also assessed faking using Idiosyncratic item responding, which differentiates interested applicant responses from uninterested applicant responses (Kuncel & Borneman, 2007). We created an Idiosyncratic Item Responding scale by selecting the 10 items with the largest mean differences between our CAF recruits and the Uninterested Applicant sample from Studies 1 and 2. Idiosyncratic Item Responding scale scores should be positively correlated with applicant faking. Example items include “I start tasks right away” and “I find it difficult to approach others.” These items are derived from the scores on the personality measure. One potential concern with developing and using a scale within the same sample is that it can capitalize on sample specific variance, and in turn, the findings may not generalize to other samples (Cureton, 1950). To address this concern, we also used an idiosyncratic item responding scale derived from another study, but the results were comparable. As a result, for simplicity, we report the one measure derived using this sample. The Idiosyncratic Item Responding scale is located in Appendix F.

**Procedural Justice Perceptions.** We assessed Procedural Justice Perceptions using three dimensions from the Selection Procedural Justice Scale (SPJS; Bauer et al., 2001), which
included Job-Relatedness, Information Known, and Chances to Perform. The composite scale had 9 items that used a 5-point response scale (1 = “Strongly Disagree” to 5 = “Strongly Agree”). Example items for Job-Relatedness, Information Known, and Chances to Perform were “Doing well on the test means a person can serve well for the CAF,” “I knew what to expect on the test,” and “I could really show my skills and abilities through the test,” respectively. The SPJS has strong evidence of internal consistency (α = .88) and validity (Bauer et al., 2001). The scale also demonstrated construct and criterion-related validity (Bauer et al., 2001; McCarthy et al., 2013).

**Test-Taking Anxiety.** We measured Test-Taking Anxiety using the Comparative Anxiety subscale from the Test Attitude Survey (Arvey et al., 1990). Test-taking anxiety was measured using 10 items on a five 5-point response scale (1 = “Strongly Disagree”, 5 = “Strongly Agree”). The Test Attitude Survey has evidence of internal consistency (α = .80) and criterion-related validity. The scale was previously validated using reactions to the Armed Services Vocational Aptitude Battery (Arvey et al., 1990; Hunter, Crosson, & Friedman, 1985).

### 4.3.3 Procedure

We ran the experiment with 13 different groups in a classroom setting, with a range of 29 to 52 recruits per class. We informed recruits about the study, stressed that participation was voluntary, and explained that their responses would not influence their military careers. Recruits within each classroom were assigned to the same condition, where they received either the TFW, IAW, FAW, MSA, or NFW personality test instructions. We used a random number generator to determine the order of the condition assignment and assigned our first 10 groups to these condition in order. We then assigned the last two groups to conditions with smaller samples. We used this approach to minimize differences in group size. Still, we had an unequal distribution of
recruits across faking dissuasion message conditions (ns: MSA = 107, IAW = 84, FAW = 54, TFW = 83, NW = 114). In a couple of instances, recruits joined the classroom too late to take part or were withdrawn during the simulation for military operations, creating unequal group sizes.

After recruits were seated, we provided a general overview of the experiment, including why personality measures are important for recruitment decisions into the Canadian Armed Forces. In order to comply with ethics, we minimized deception and informed all recruits that the CAF was looking to improve the accuracy of their personality test by piloting different sets of test instructions. However, we did not inform recruits of our hypotheses and the same script was used for all groups.

We instructed recruits in the four separate warning conditions to engage in a selection role-play exercise, where they would fill out the surveys as if they were initially trying to secure employment with the CAF. For each faking dissuasion message condition, the experimenter guided recruits through the study by displaying the roleplay instructions and warnings on PowerPoint slides, which were displayed on multiple monitors. The PowerPoint presentations included test instructions and faking dissuasion messages for the IAW, FAW, MSA, TFW, and NFW are presented in Appendices A through E. We used this approach to ensure that recruits were aware of the instructions before filling out the surveys and to keep recruits at a similar pace, so that we could better control the group. Next, these recruits were asked to fill out our personality measures, the OCEAN.20 and IPIP, as if they were trying to be recruited by the CAF. After completion, recruits were directed that the role-play was over and to fill out the remaining surveys honestly. Recruits in the no-warning group did not engage in the roleplay, were encouraged to respond as honestly as possible, and completed the personality test as they would
without the simulation. To encourage recruits to follow our instructions, we reminded recruits that their responses would be unassociated with their personnel file and would not influence their military service careers in any way. Recruits then completed a measure of procedural justice perceptions and test-taking anxiety. As before, the experimenter guided recruits through each measure—at a time—using PowerPoint slides and read instructions aloud. The experiment took 30 to 35 minutes per group.

4.4 Results

4.4.1 Preliminary Analyses

We did not have true random assignment, and as a result, we tested to see if mean personality scores varied between our faking dissuasion groups prior to our study. If there were differences prior to the manipulation, they could provide alternative explanations for any differences between our faking dissuasion message conditions. To test for this, we conducted a MANOVA with faking dissuasion message condition as the independent variable and the Big Five personality scores on the TSD-PI during initial recruitment (at baseline, and prior to the experimental manipulation) as the dependent variables. There was a significant multivariate effect for faking dissuasion message condition, $F(20, 1496.75) = 1.84, p = .01$, Wilk’s $\Lambda = .92$. The univariate analyses revealed significant main effects for Openness, $F(4, 455) = 3.51, p = .008$, $\eta_p^2 = .03$, and Emotional Stability, $F(4, 455) = 2.47, p = .04$, $\eta_p^2 = .02$. Tukey’s B post-hoc test revealed that recruits in the MSA and FAW conditions scored significantly higher on Openness at baseline (prior to group assignment) than did recruits in the TFW condition. Recruits in the FAW condition scored significantly higher on Emotional Stability than did recruits in the IAW condition. No other differences were significant, suggesting there were no systematic differences that would confound our main analyses. Applicants generally do not fake
on Openness (Birkeland et al., 2006), leaving only one significant difference that could complicate results. Rather than controlling for scores on Emotional Stability, which could influence all findings, we used this information to inform any differences between the IAW and FAW on Emotional Stability in the main analysis. All means, standard errors, and 95% confidence intervals of the TSD-PI scores are presented in Table 10.

4.4.2 Faking Dissuasion Message Efficacy on OCEAN.20

To examine the efficacy of our faking dissuasion messages, we conducted a MANOVA with faking dissuasion message condition as the independent variable and Big Five personality scores from the OCEAN.20 as the dependent variables. We found a significant multivariate effect of faking dissuasion message condition, $F(20, 1493.43) = 2.87, p < .001$, Wilk’s $\Lambda = .88$. The univariate analyses revealed significant main effects for: Openness, $F(4, 454) = 3.31, p = .01, \eta^2_p = .03$; Conscientiousness, $F(4, 454) = 5.25, p < .001, \eta^2_p = .04$; Agreeableness, $F(4, 454) = 3.85, p = .03, \eta^2_p = .03$; and Emotional Stability, $F(4, 454) = 2.73, p = .03, \eta^2_p = .02$. To further probe these effects, we conducted Tukey’s B post-hoc tests for each personality dimension. For Conscientiousness, recruits in the IAW condition scored significantly lower than recruits in the FAW, MSA, TFW, and NFW conditions. Further, recruits in the IAW condition scored significantly lower on Agreeableness than recruits in the CAF condition, but not lower than those in the NFW condition. Recruits in the NFW condition scored significantly lower on Agreeableness than recruits in the FAW condition. Finally, recruits in the IAW condition scored significantly lower on Openness than recruits in either the FAW or MSA condition, but not lower than in the NFW condition. All means, standard errors, and 95% confidence intervals of the OCEAN.20 scores are presented in Table 11.
Table 10. Recruit Scores on the Trait Self-Descriptive Personality Inventory during Initial Recruitment.

<table>
<thead>
<tr>
<th>Warning Condition</th>
<th>n</th>
<th>Trait Self-Descriptive Personality Inventory</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Openness</td>
<td>Conscientiousness</td>
<td>Extraversion</td>
<td>Agreeableness</td>
<td>Emotional Stability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>MSA</td>
<td>113</td>
<td>4.74</td>
<td>0.08</td>
<td>[4.58, 4.90]</td>
<td>5.59</td>
<td>0.06</td>
<td>[5.46, 5.71]</td>
<td>4.57</td>
<td>0.09</td>
<td>[4.39, 4.75]</td>
<td>5.85</td>
<td>0.05</td>
</tr>
<tr>
<td>IAW</td>
<td>85</td>
<td>4.53</td>
<td>0.09</td>
<td>[4.35, 4.71]</td>
<td>5.49</td>
<td>0.07</td>
<td>[5.35, 5.64]</td>
<td>4.47</td>
<td>0.11</td>
<td>[4.26, 4.68]</td>
<td>5.70</td>
<td>0.06</td>
</tr>
<tr>
<td>FAW</td>
<td>60</td>
<td>4.77</td>
<td>0.11</td>
<td>[4.55, 4.99]</td>
<td>5.41</td>
<td>0.09</td>
<td>[5.24, 5.58]</td>
<td>4.44</td>
<td>0.13</td>
<td>[4.19, 4.69]</td>
<td>5.78</td>
<td>0.07</td>
</tr>
<tr>
<td>TFW</td>
<td>86</td>
<td>4.39</td>
<td>0.09</td>
<td>[4.20, 4.57]</td>
<td>5.47</td>
<td>0.07</td>
<td>[5.33, 5.61]</td>
<td>4.37</td>
<td>0.11</td>
<td>[4.16, 4.58]</td>
<td>5.76</td>
<td>0.06</td>
</tr>
<tr>
<td>NFW</td>
<td>116</td>
<td>4.46</td>
<td>0.08</td>
<td>[4.30, 4.61]</td>
<td>5.51</td>
<td>0.06</td>
<td>[5.39, 5.63]</td>
<td>4.48</td>
<td>0.09</td>
<td>[4.30, 4.66]</td>
<td>5.85</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note. This table displays the estimated marginal mean personality scores on the Trait Self-Descriptive Personality Inventory prior to our experiment. Personality items are on a 7-point scale (1 = Extremely Uncharacteristic, 7 = Extremely Characteristic). MSA = Moral Suasion Appeal. IAW = Immediate Authentication Warning. FAW = Future Authentication Warning. TFW = Traditional Faking Warning. NFW = No warning.

a Emotional stability was significantly higher in the FAW condition than in the IAW condition.

b Openness was significantly higher in the FAW condition than in the TFW condition.
Table 11. Recruit Scores on the OCEAN.20 Personality Inventory during the Application Simulation

<table>
<thead>
<tr>
<th>Warning Condition</th>
<th>n</th>
<th>Openness</th>
<th>Conscientiousness</th>
<th>Extraversion</th>
<th>Agreeableness</th>
<th>Emotional Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SE</td>
<td>95% CI</td>
<td>M</td>
<td>SE</td>
</tr>
<tr>
<td>MSA</td>
<td>113</td>
<td>4.42</td>
<td>0.12</td>
<td>[4.18, 4.66]</td>
<td>5.19</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[4.18, 4.66]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IAW</td>
<td>84</td>
<td>3.80</td>
<td>0.14</td>
<td>[3.52, 4.08]</td>
<td>4.53</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[3.52, 4.08]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FAW</td>
<td>57</td>
<td>4.40</td>
<td>0.17</td>
<td>[4.07, 4.74]</td>
<td>4.98</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[4.07, 4.74]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFW</td>
<td>87</td>
<td>4.09</td>
<td>0.14</td>
<td>[3.82, 4.36]</td>
<td>5.03</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[3.82, 4.36]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NFW</td>
<td>118</td>
<td>4.23</td>
<td>0.12</td>
<td>[3.99, 4.46]</td>
<td>5.01</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[3.99, 4.46]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. This table displays the estimated marginal mean personality scores on the OCEAN.20 Personality Inventory prior to our experiment. Personality items were measuring using a 7-point scale (1 = Extremely Uncharacteristic, 7 = Extremely Characteristic). MSA = Moral Suasion Appeal. IAW = Immediate Authentication Warning. FAW = Future Authentication Warning. TFW = Traditional Faking Warning. NFW = No warning.
4.4.3 Faking Dissuasion Message Efficacy on IPIP

In addition, we examined the efficacy of our faking dissuasion messages by conducting a MANOVA with faking dissuasion message condition as the independent variable and Big Five personality scores from the IPIP as the dependent variables. We found a significant multivariate effect of faking dissuasion message condition, $F(20, 1447) = 1.85, p = .01$, Wilk’s $\Lambda = .92$. The univariate results suggest that only the univariate test assessing the effects of faking dissuasion message condition on Conscientiousness was significant, $F(4, 440) = 3.42, p = .01, \eta_p^2 = .03$. Tukey’s B post-hoc tests revealed that recruits in the IAW condition reported significantly lower scores on Conscientiousness than did the FAW, MSA, TFW, and NFW conditions. All means, standard errors, and 95% confidence intervals of the IPIP scores are presented in Table 12.

4.4.4 Faking Dissuasion Message Efficacy on Faking Measures

We also examined the extent that the faking dissuasion messages reduced scores on our faking measures. To do this, we ran another MANOVA with faking dissuasion message condition as the independent variable and Communal Impression Management, Blatant Extreme Responding, and Idiosyncratic Item Responding as the dependent variables. We found a significant multivariate effect of faking dissuasion message condition, $F(16, 1283.76) = 2.01, p = .01$, Wilk’s $\Lambda = .93$. There were significant univariate effects of faking dissuasion message condition on Blatant Extreme Responding derived from the OCEAN.20, $F(4, 423) = 4.26, p = .002, \eta_p^2 = .04$, and on Blatant Extreme Responding derived from the IPIP, $F(4, 423) = 4.37, p = .002, \eta_p^2 = .04$. Tukey’s B post-hoc tests revealed that recruits in the IAW condition had significantly lower scores on Blatant Extreme Responding derived from both the OCEAN.20 and the IPIP than recruits in the FAW and MFW conditions. There were no significant main effects of faking dissuasion message condition on Communal Impression Management or IIR.
Table 12. Recruit Scores on the IPIP Personality Inventory during the Application Simulation

| Warning Condition | n   | Openness | | Conscientiousness | | Extraversion | | Agreeableness | | Emotional Stability |
|-------------------|-----|----------|----------|--------------------|----------|----------------|----------|-------------------|-------------------|
|                   |     | M        | SE       | 95% CI             | M        | SE       | 95% CI             | M        | SE       | 95% CI             | M        | SE       | 95% CI             |
| MSA               | 113 | 3.23     | 0.04     | [3.16, 3.31]       | 3.76     | 0.04     | [3.68, 3.84]       | 3.47     | 0.04     | [3.38, 3.55]       | 3.51     | 0.04     | [3.44, 3.58]       | 3.55     | 0.04     | [3.47, 3.63]       |
| IAW               | 84  | 3.11     | 0.04     | [3.03, 3.19]       | 3.58     | 0.05     | [3.49, 3.68]       | 3.36     | 0.05     | [3.26, 3.45]       | 3.49     | 0.04     | [3.41, 3.47]       | 3.47     | 0.05     | [3.38, 3.56]       |
| FAW               | 57  | 3.22     | 0.05     | [3.12, 3.33]       | 3.76     | 0.06     | [3.64, 3.87]       | 3.54     | 0.06     | [3.42, 3.67]       | 3.48     | 0.05     | [3.38, 3.58]       | 3.69     | 0.06     | [3.57, 3.80]       |
| TFW               | 87  | 3.22     | 0.04     | [3.14, 3.30]       | 3.80     | 0.05     | [3.71, 3.89]       | 3.55     | 0.05     | [3.45, 3.64]       | 3.63     | 0.04     | [3.55, 3.70]       | 3.59     | 0.05     | [3.50, 3.68]       |
| NFW               | 118 | 3.20     | 0.04     | [3.13, 3.27]       | 3.77     | 0.04     | [3.69, 3.85]       | 3.47     | 0.04     | [3.39, 3.56]       | 3.50     | 0.03     | [3.43, 3.56]       | 3.53     | 0.04     | [3.46, 3.61]       |

Note. This table displays the estimated marginal mean personality scores on the IPIP Personality Inventory during our experiment. Personality items were measured using a 5-point scale (1 = Strongly Disagree, 5 = Strongly Agree). MSA = Moral Suasion Appeal. IAW = Immediate Authentication Warning. FAW = Future Authentication Warning. TFW = Traditional Faking Warning. NFW = No warning.
All means, standard errors, and 95% confidence intervals of the IPIP scores are presented in Table 13.

4.4.5 Faking Dissuasion Message Efficacy on Recruit Reactions

We examined the extent that the faking dissuasion messages influenced recruit reactions. To do this, we ran a MANOVA with faking dissuasion message condition as the independent variable, and Test Anxiety and three components of Procedural Justice Perceptions (Job Relatedness, Information Known, and Chances to Perform), as the dependent variables. The multivariate analysis did not approach significance, $F(16, 1354.03) = 1.27, p = .21$, Wilk’s $\Lambda = .96$. All means, standard errors, and 95% confidence intervals of the IPIP scores are presented in Table 14.

4.5 Discussion

We examined if faking dissuasion messages could thwart applicant faking among military recruits and if these messages engendered negative reactions. In contrast to Study 2, our results provided support for the IAW. The IAW reduced personality scores on a measure of conscientiousness relative to other faking dissuasion messages. The IAW also reduced BER scores – an index of faking – derived from both the OCEAN.20 and IPIP, suggesting that the IAW prevents recruits from responding with extreme responses (such as “7 = extremely agree”). However, the IAW did not appreciably reduce faking on other personality traits relative to other faking dissuasion messages, and also failed to reduce scores on two faking indices—Communal Impression Management and Idiosyncratic Item Responding. Results from Study 1 suggest that researchers and practitioners should aim for convergence across faking indices. Reducing faking on only one faking assessment is encouraging, but needs to be interpreted with caution. It is possible that the IAW is reducing one style of faking, but not others.
Table 13. Recruit Scores on the Faking Measures during the Application Simulation

<table>
<thead>
<tr>
<th>Warning Condition</th>
<th>$n$</th>
<th>OCEAN.20 BER</th>
<th>IPIP BER</th>
<th>IIR</th>
<th>BIMI-C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$M$</td>
<td>$SE$</td>
<td>95% CI</td>
<td>$M$</td>
</tr>
<tr>
<td>MSA</td>
<td>105</td>
<td>2.20</td>
<td>0.24</td>
<td>[1.73, 2.67]</td>
<td>18.28</td>
</tr>
<tr>
<td>IAW</td>
<td>76</td>
<td>1.47</td>
<td>0.28</td>
<td>[0.92, 2.03]</td>
<td>13.86</td>
</tr>
<tr>
<td>FAW</td>
<td>52</td>
<td>2.71</td>
<td>0.34</td>
<td>[2.04, 3.38]</td>
<td>19.98</td>
</tr>
<tr>
<td>TFW</td>
<td>83</td>
<td>2.46</td>
<td>0.27</td>
<td>[1.93, 2.99]</td>
<td>18.65</td>
</tr>
<tr>
<td>NFW</td>
<td>108</td>
<td>3.02</td>
<td>0.24</td>
<td>[2.55, 3.48]</td>
<td>21.82</td>
</tr>
</tbody>
</table>

Note. This table displays the estimated marginal means for our four faking measures. For each faking measure, higher scores reflect more faking. MSA = Moral Suasion Appeal. IAW = Immediate Authentication Warning. FAW = Future Authentication Warning. TFW = Traditional Faking Warning. NFW = No warning.
<table>
<thead>
<tr>
<th>Warning Condition</th>
<th>n</th>
<th>Test-Taking Anxiety</th>
<th>Job Relatedness</th>
<th>Information Known</th>
<th>Chances to Perform</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M      SE    95% CI</td>
<td>M      SE    95% CI</td>
<td>M      SE    95% CI</td>
<td>M      SE    95% CI</td>
</tr>
<tr>
<td>MSA</td>
<td>109</td>
<td>2.40 0.06 [2.29, 2.52]</td>
<td>2.90 0.07 [2.76, 3.05]</td>
<td>3.53 0.07 [3.4, 3.66]</td>
<td>2.39 0.07 [2.25, 2.53]</td>
</tr>
<tr>
<td>IAW</td>
<td>86</td>
<td>2.59 0.07 [2.46, 2.72]</td>
<td>2.98 0.08 [2.82, 3.14]</td>
<td>3.48 0.07 [3.34, 3.63]</td>
<td>2.53 0.08 [2.37, 2.69]</td>
</tr>
<tr>
<td>FAW</td>
<td>54</td>
<td>2.52 0.08 [2.36, 2.68]</td>
<td>3.01 0.10 [2.81, 3.21]</td>
<td>3.68 0.09 [3.5, 3.86]</td>
<td>2.56 0.10 [2.36, 2.77]</td>
</tr>
<tr>
<td>TFW</td>
<td>86</td>
<td>2.60 0.07 [2.47, 2.73]</td>
<td>2.76 0.08 [2.6, 2.92]</td>
<td>3.51 0.07 [3.37, 3.66]</td>
<td>2.35 0.08 [2.19, 2.51]</td>
</tr>
<tr>
<td>NFW</td>
<td>116</td>
<td>2.56 0.06 [2.45, 2.67]</td>
<td>2.89 0.07 [2.75, 3.03]</td>
<td>3.42 0.06 [3.29, 3.54]</td>
<td>2.49 0.07 [2.35, 2.62]</td>
</tr>
</tbody>
</table>

Note. This table displays the estimated marginal means for the recruit reaction measures. Recruit reactions were measured using a 5-point scale (1 = Strongly Disagree, 5 = Strongly Agree). MSA = Moral Suasion Appeal. IAW = Immediate Authentication Warning. FAW = Future Authentication Warning. TFW = Traditional Faking Warning. NFW = No warning.
It is also possible that the IAW is not reducing applicant faking, but instead, changing the way all applicants respond to the personality test. For example, even non-fakers might be less likely to engage in “extreme responding”—even when the extreme answer is most representative of them. Indeed, this has been a criticism of using faking warnings on personality tests (McFarland, 2003). As a result, the IAW may have utility in reducing applicant faking, but it is difficult to draw firm conclusions without convergent reductions in Communal Impression Management or Idiosyncratic Item Responding.

There are a few reasons the IAW showed some utility in the present study, but not in Study 2. First, CAF has high selection ratios, which changes the risk to reward ratio for faking. Faking is less likely to be advantageous since high scores are less essential to secure employment, and in turn, being detected as faking is more likely to cost applicants a job that they would have otherwise obtained. Theoretical models of faking suggest that applicants do weigh the risks of detection and the benefits of faking (Goffin & Boyd, 2009). By contrast, Study 2 presented a screening opportunity, where higher scores were essential just to obtain an interview, not even an offer of employment. As a result, the risk to reward ratio may have favored faking despite the faking dissuasion message. This would suggest that the efficacy of faking dissuasion messages may depend on where personality testing is used in the selection process. In the IAW condition, there were also built in mechanisms for verification of personality including reference checks. This threat may have seemed more credible in CAF context, where they were close to securing real employment, whereas it may have seemed less credible in an applicant screening scenario. References are typically one of the last steps in the selection process, as they take a substantial amount of time. In Study 2, where the IAW was ineffective, our mTurk applicants may have been aware of the improbability that references would be contracted. This last
mechanism may explain why the IAW was effective in Study 3, but the TFW was ineffective, which threatened disqualification without a credible mechanism.

Additionally, we expected the extra threat of verification in the IAW to produce negative testing reactions among the recruits, but found little evidence of increased negative reactions or lower ratings of procedural justice when compared to the other warning conditions. There are two potential explanations for why the IAW did not produce more adverse reactions. First, and in accordance with Goffin and Boyd’s (2009) faking decision tree model, applicants go through a serious of decisions when choosing to fake on each test item (e.g., considering the morality of faking; the chances of detection; whether faking will lead to disqualification; etc.). The CAF recruits may have experienced the same decision-making process during the application simulation, regardless of their condition, and in turn, the same cognitive burden or anxiety across faking dissuasion messages. Second—and more concerning—is that the recruits may not have experienced negative reactions because we had them engage in a simulation without real meaningful consequences. However, these findings are consistent with Study 2, which found that job applicants did not have more adverse reactions to more threatening warnings, even when they provided real contact information for references and thought there was a real job.

By contrast, the MSA and FAW failed to thwart applicant faking, although these results are still informative for the field. Several scholars have postulated that “softer” faking dissuasion messages that appeal to morality, educate, and have applicants consider long term consequences will reduce applicant faking—rather than threaten applicants (Goffin & Boyd, 2009; Pace & Borman, 2006; Uruena & Robie, 2011). The results of our studies do not support this speculation. In addition, we found that a faking dissuasion message, which capitalizes on short-term accountability and immediate consequences, are more effective than traditional warnings,
yet do not engender negative reactions in stimulated application scenarios. Together, these findings provide no incentive to further investigate these “softer” messages. Instead, future research on faking dissuasion messages should try to maximize short-term accountability by implementing more compelling descriptions of how faking can be detected. Our findings suggest that faking dissuasion messages may be unable to convince applicants to consider the morality of faking, or to be more concerned about the long-term consequences of faking beyond their dispositional inclinations.

Our findings suggest the MSA and FAW were ineffective in preventing faking, as other factors are more important to the decision to fake. Based on accountability theory, we expected that recruits in the FAW condition would worry about the long-term potential of verifiability and the need to defend their answers during a one-on-one observation; clearly, our findings do not support these ideas. We entertain three potential explanations for this finding. First, the recruits believed that the short-term reward of securing employment in the Canadian Armed Forces simply outweighed the risks of being detected at a future time. The recruits may also have believed that they could emulate the personality they portrayed, and in turn, genuinely viewed it as low-risk to fake (Goffin & Boyd, 2009). Second, the recruits—like applicants for large organizations—felt “faceless” and in turn, did not believe their faking would be noticed. Finally, as discussed earlier, recruits participated in a simulation and may not have been able to imagine the scenario as they would experience in a real job application. However, this is unlikely, as the recruits had recently applied for a position within the CAF, and therefore, should have been able to imagine the application scenario. Additionally, job applicant simulations with real employees tend to provide realistic estimates of faking when compared to student samples (Goffin, Jang, & Skinner, 2011).
We also found that moral prompts about the negative outcomes of faking did not reduce faking by the recruits. This finding mirrors those reported in Study 2. Goffin and Boyd (2009) suggest that the morality of faking is the first decision that applicants consider when answering personality items during the hiring process. It is possible that applicants may not consider morality first in the decision-making process; instead, applicants may view faking as the “norm” and believe that faking is normal. Indeed, the majority of applicants do fake (Donovan et al., 2003; Feeney & Goffin, 2015; Griffith et al., 2007; Holden et al., 2017).

Another possibility is that there may be external factors that are too powerful for test instructions to override. For example, if an applicant or recruit is applying for a position when they do not have enough money to support their family, they may view faking as a lesser evil than not providing for their family. Thus, morality may be relative to the applicant’s need for employment and external considerations. If this is the case, then test instructions appealing to morality are unlikely to be effective. Similarly, the moral consideration of applicant faking is dispositional rather than situational. For example, we know some people are higher in trait integrity than others, and that these traits predict workplace delinquency (Lee, Ashton, & de Vries, 2005). Therefore, virtuous applicants may be unlikely to fake, regardless of the test instructions, whereas others may not care about the morality of faking. Unfortunately, due to time constraints, we were unable to ask the recruits about their reactions after the simulation.

4.5.1 Limitations

The main limitation of our study is that we used a recruitment simulation with the recruits, and in turn, their results may not generalize to recruits or applicants undergoing the high incentive to perform in a real scenario. We used a simulation because both the CAF and our university comply with national ethics policies from the Social Sciences and Humanities
Research Council, which would forbid administering different test instructions to real job applicants, as some applicants may be disadvantaged relative to others as a function of our research. It was paramount that participants’ futures, or possible careers not be affected by the experimental outcomes as it might reward applicants in one condition, while punishing applicants in another condition. Therefore, every effort was made to re-create the initial application process—including auditory and visual cues present during the initial recruitment appointment (Aguinis & Bradley, 2014).

The second limitation is that the recruits were assigned to us in large groups (29 to 52) and we assigned groups (rather than individual recruits) randomly to our different experimental conditions. We also had to defer to military operations in some instances, which may have confounded our assignment. For example, our smallest group had several recruits removed in the early stage of our study for administrative purposes. The groups also had substantial variations in reading speed, and for logistical reasons, we moved at the average pace of recruits. This lead some recruits who were behind to skip sections or submit their package prematurely—which may have produced some systematic bias, especially for survey questions at the end of our study. Finally, we assumed that all recruits paid sufficient attention to and internalized instructions for each of the manipulations; however, findings suggesting the relative effectiveness of IAW condition in preventing faking provide partial evidence of the manipulation’s success.

4.5.2 Implications and Conclusions

Our findings have three primary implications for researchers and human resource practitioners. The first implication is that faking dissuasion messages that emphasize short-term accountability are the most effective at combatting applicant faking. This effect is best utilized by making applicants or recruits believe that there is a credible process to verify their answers,
such as threatening to check applicant responses with personal references they provided prior to
the personality test. The second implication is that there is little incentive to continue
investigating “softer” faking dissuasion messages, which emphasize educational or moral
suasion, as they may not appreciably reduce faking and may lead to more positive applicant
reactions compared to more threatening messages.
4.6 References


Appendix A: Immediate Authentication Warning: PowerPoint Presentation

**Overview**

- The Canadian Armed Forces uses workstyle assessments that help determine which recruits are the best fit for the Canadian Armed Forces.
- Workstyle assessments ask recruits a series of questions about their attitudes, thoughts, and behaviours.
- Workstyle assessments is another name for Personality Tests, which you have previously filled out for the Canadian Armed Forces.

**Overview**

- You are being invited to participate in a one hour research study.
- The study will examine how different sets of assessment instructions may influence the way military recruits respond to workstyle assessments.

**Overview**

- We begin with a broad overview of the study and then assign you to one of 5 different groups.
- This may involve a role-play exercise, where we will ask you to fill out workstyle assessments as if you were trying to be selected by the Canadian Armed Forces.
- Depending on your assigned group, there may not be a role-play exercise and we will ask you to fill out the workstyle measure as you normally would.
- Upon completion of the workstyle assessment, we will ask you to respond, as honestly as possible, to a variety of surveys that will assess topics such as how anxious you were during the assessment and how you perceived the workstyle assessment.

**Overview**

- We will also ask if you are comfortable providing your service number.
- The service number allows us to retrieve information from your personnel file.
- We will use this information to see how different assessment instructions influence your workstyle scores, and in turn, how these workstyle scores predict training outcomes.
- We will also look to see if scores on previous assessments can be used to predict how participants respond during this study.

**Overview**

- Providing your service number is completely voluntary and you can still participate today without providing your service number.
- Not providing your service number will not impact your career in anyway, will not be known to your supervisor, and will only be analysed at a aggregate level.
- If you do not wish to participate, you are free to exit the classroom for the duration of the study or stay at your desk to work on other tasks.
Overview

- Workstyle assessments have been shown to predict on-the-job performance across a wide variety of jobs and suitability for placement in the CAF.
- One concern with these assessments is that people may answer in a way that isn’t the most reflective of their true self.
- Our goal is to devise a set of test instructions that will help guide future military recruits to fill out the workstyle measures in a way that is the most self-reflective.
- The goal is to help the Canadian Armed Forces make more accurate recruitment/placement decisions.

Voluntary Participation

- Participation in this study is voluntary. You may decline to consent, refuse to answer any questions, or withdraw from the study at any time without any penalty or any effect on your career with the Canadian Armed Forces.
- Colin Kemp works with GDMPIRA, which operates at an arm’s length from operations.
  - Colin Kemp has no influence over your career and has no incentive to share your data with your chain of command or personnel in your career branch.
  - Thus, you can be confident that we will not breach confidentiality.
- Your supervisor will not be made aware of the responses you provide nor will they be placed in your personnel file.

Letter of Information and Consent

- Please take a moment to review the letter of information we have handed out.
- If you agree to participate in the study today, please sign the consent form that is on the next page.
- If you do not agree to participate, feel free to leave at this time.
- Please raise your hand if you have any questions.
- Please do not open the booklet until instructed to do so.

Freedom to Leave

- If you feel anxious or uncomfortable at any point during or after the role-play exercise, you may leave without consequence.
- We will not communicate your decision to participate to your supervisors.
- If you would like to discontinue your participation, you can simply discard your answers into the garbage bins at the front of the room.

Role-Play Exercise

- We would like you to engage in a role-play exercise.
- In this exercise, we will ask you to complete two assessments as though you have not yet been accepted into the Canadian Armed Forces.
- Why? We aim to improve the accuracy of the assessments used during recruitment. Thus, we need you to answer in a way that reflects how people answer during the recruitment process.
- Slides that are part of the role-play exercise will have a purple background.
- Remember: All of your answers today, including those given during the role-play, will only be used for research purposes and will not affect your career with the Canadian Armed Forces in any way.

ROLE-PLAY BEGINS

ROLE-PLAY EXERCISE
Instructions: Workstyle
- Imagine you are about to fill out two workstyle assessments that ask questions about how you typically think, feel, and act.
- Imagine they are very important assessments as they are used to help make decisions about recruitment into the Canadian Armed Forces.
- Imagine that the Canadian Armed Forces may verify responses on these workstyle assessments to ensure everyone is providing honest responses.
- We will now explain the different ways the Canadian Armed Forces may verify responses. We want you to imagine this is happening.

First Method of Verifying Responses: Reference Checking
- A Recruiting Officer may verify your responses on these workstyle assessments by communicating with the contacts you provided during your initial recruitment to the Canadian Armed Forces. By contacts, we are referring to the people you listed as potential references in your application to the Canadian Armed Forces.
- For example, if your answers on these assessments indicate you are highly attentive to detail and remain calm in stressful situations, the Recruiting Officer can call your contacts and ask them about your attention to detail and whether or not you remain calm under stress.
- If the Recruiting Officer finds several differences between your responses on these assessments and the information provided by your contacts, the director may schedule a one-on-one interview with you.
- In this interview, the Recruiting Officer may ask that you explain why your responses to these assessments are not similar to the information provided by your contacts.

Second Method of Verifying Responses: Internal Integrity Checks
- Testing experts have developed questions that assess the tendency to provide responses that are unlikely to be honest.
- Imagine a recruit was asked how much they agree with the following statement, "I always follow the rules."
- If the recruit responds quickly, without thinking, they may respond "strongly agree" because they tend to be a rule-abiding person. However, is this really accurate?
- For example, the recruit may occasionally cross the street when the "Don't walk" signal is lit. Similarly, the recruit may occasionally watch movies or listen to music they didn't properly purchase or rent.

Second Method of Verifying Responses: Internal Integrity Checks
- The reality is most people do break some rules. Thus, responding "strongly agree" would probably be inaccurate and indicate that the person is not being completely honest.
- Questions like this are inserted throughout the assessments. So please, think carefully before responding to each question in these assessments, and respond honestly.
- If your responses suggest that you are responding dishonestly, a Recruiting Officer may schedule a one-on-one meeting with you. In this meeting, you will be asked to explain your responses on the workstyle assessments.

Instructions: Workstyle Assessment #1
- Before filling out the assessment, please take a few minutes to carefully review the instructions for the first workstyle assessment.
- This assessment takes approximately 5 minutes to complete.
- Once you’re finished, please stop and wait for instructions.
- Please begin now.

If you respond dishonestly and are unable to satisfactorily explain your responses on the assessments, you may be ineligible for a position in the Canadian Armed Forces.
- However, if you respond honestly, you have nothing to worry about.
Instructions: Workstyle Assessment #2
- Before filling out the assessment, please take a few minutes to carefully review the instructions for the second workstyle assessment.
- This assessment takes approximately 15 minutes to complete.
- Once you’re finished, please stop and wait for instruction to continue.
- Please begin now.

Service Number Request Form
- We will request your service number in order to link your scores from today to some other information.
- This includes your scores on workstyle and aptitude tests during your recruitment and training performance data.
- We will only use this information to help improve the accuracy of the workstyle assessments that are filled out today.
- You may decline to provide your service number and still participate.
- Colin Kemp from DMPRA will link the data using your service number.
- Colin Kemp will remove the service number from the data file before providing the data to the researchers from Western or conducting any analysis.
- We will only look at overall trends in the data and are not interested in your individual responses.
- The data you provide today will only be used for research purposes.
- THIS WILL NOT AFFECT YOUR CAREER WITH THE Canadian Armed Forces IN ANY WAY.

Instructions: Follow-Up Measures
- Now that you’re finished completing the workstyle assessments, we would like you to stop role-playing and answer the questions naturally, as you would normally.
- As a reminder, your answers are strictly confidential and will only be used for research purposes.

Instructions: Feelings during Tests
- Now we would like you to fill out some additional measures.
- For each measure, we will review the instructions and prompt you when to start.
- Before answering, please pay careful attention to the different response categories, which are different for each measure.
- You will see a stop page when you finish each measure.
- Please wait until instructed to move forward.

END OF ROLE-PLAY
Appendix B: Future Authentication Warning: PowerPoint Presentation

Overview

- The Canadian Armed Forces uses workstyle assessments that help determine which recruits are the best fit for the Canadian Armed Forces.
- Workstyle assessments ask recruits a series of questions about their attitudes, thoughts, and behaviours.
- Workstyle assessments is another name for Personality Tests, which you have previously filled out for the Canadian Armed Forces.

Overview

- You are being invited to participate in a one hour research study.
- The study will examine how different sets of assessment instructions may influence the way military recruits respond to workstyle assessments.

Overview

- We begin with a broad overview of the study and then assign you to one of 5 different groups.
- This may involve a role-play exercise, where we will ask you to fill out workstyle assessments as if you were trying to be selected by the Canadian Armed Forces.
- Depending on your assigned group, there may not be a role-play exercise and we will ask you to fill out the workstyle measure as you normally would.
- Upon completion of the workstyle assessment, we will ask you to respond, as honestly as possible, to a variety of surveys that will assess topics such as how anxious you were during the assessment and how you perceived the workstyle assessment.

Overview

- We will also ask if you are comfortable providing your service number.
- The service number allows us to retrieve information from your personnel file.
- We will use this information to see how different assessment instructions influence your workstyle scores, and in turn, how these workstyle scores predict training outcomes.
- We will also look to see if scores on previous assessments can be used to predict how participants respond during this study.

Overview

- Providing your service number is completely voluntary and you can still participate today without providing your service number.
- Not providing your service number will not impact your career in any way, will not be known to your supervisor, and will only be analysed at a aggregate level.
- If you do not wish to participate, you are free to exit the classroom for the duration of the study or stay at your desk to work on other tasks.
Overview

- Workstyle assessments have been shown to predict on-the-job performance across a wide variety of jobs and suitability for placement in the CAF
- One concern with these assessments is that people may answer in a way that isn’t the most reflective of their true self
- Our goal is to devise a set of test instructions that will help guide future military recruits to fill out the workstyle measures in a way that is the most self-reflective
- The goal is to help the Canadian Armed Forces make more accurate recruitment/placement decisions

Voluntary Participation

- Participation in this study is voluntary. You may decline to consent, refuse to answer any questions, or withdraw from the study at any time without any penalty or any effect on your career with the Canadian Armed Forces
- Colin Kemp works with DMPRA, which operates at an arm’s length from operations
  - Colin Kemp has no influence over your career and has no incentive to share your data with your chain of command or personnel in your career branch
  - Thus, you can be confident that we will not breach confidentiality
- Your supervisor will not be made aware of the responses you provide nor will they be placed in your personnel file

Letter of Information and Consent

- Please take a moment to review the letter of information we have handed out
- If you agree to participate in the study today, please sign the consent form that is on the next page
- If you do not agree to participate, feel free to leave at this time
- Please raise your hand if you have any questions
- Please do not open the booklet until instructed to do so

Freedom to Leave

- If you feel anxious or uncomfortable at any point during or after the role-play exercise, you may leave without consequence
- We will not communicate your decision to participate to your supervisors
- If you would like to discontinue your participation, you can simply discard your answers into the garbage bins at the front of the room

Role-Play Exercise

- We would like you to engage in a role-play exercise
- In this exercise, we will ask you to complete two assessments as though you have not yet been accepted into the Canadian Armed Forces
- Why? We aim to improve the accuracy of the assessments used during recruitment. Thus, we need you to answer in a way that reflects how people answer during the recruitment process
- Slides that are part of the role-play exercise will have a purple background
- Remember: All of your answers today, including those given during the role-play, will only be used for research purposes and will not affect your career with the Canadian Armed Forces in any way

ROLE-PLAY BEGINS

ROLE-PLAY EXERCISE
Instructions: Workstyle

- Imagine you are about to fill out two workstyle assessments that ask questions about how you typically think, feel, and act.
- Imagine they are very important assessments as they are used to help make decisions about recruitment into the Canadian Armed Forces.
- Imagine that the Canadian Armed Forces may verify responses on these workstyle assessments to ensure everyone is providing honest responses.
- We will now explain the different ways the Canadian Armed Forces may verify responses. We want you to imagine this is happening.

We may verify responses using a two-step process:

First Step. If you are selected for an interview, a Recruiting Officer may make observations of your workstyle during the interview:

- The Recruiting Officer may consider whether your responses on both workstyle assessments are similar to their own observations during the interview.
- For example, the Recruiting Officer may make observations about how outgoing or socially confident you appear during the interview, and compare that to your answers on workstyle questions such as how much you agree with statements like “I am skilled in handling social situations” and “I feel comfortable around people.”
- If the Recruiting Officer finds several differences between your responses on these assessments and their own observations during the interview, they may schedule a follow-up meeting with you.
- In this meeting, the Recruiting Officer will ask that you explain why your responses to the workstyle assessments are similar to their observations.

Second Step. Observations about your workstyle during basic training may be obtained from your Recruiting Officer. This information may be used as part of the overall evaluation of your training performance (e.g., in your course report). If your performance does not meet expectations:

- Your Recruiting Officer may consider whether your responses to these workstyle assessments are similar to their own observations about your workstyle during training.
- If the Recruiting Officer finds several differences between your responses on these workstyle assessments and their observations, the Recruiting Officer may schedule a follow-up interview with you.
- In this interview, the Recruiting Officer may ask that you explain why your responses on the workstyle assessments are not similar to the observations of your workstyle during occupation training.

If you’ve failed to meet performance expectations and your Recruiting Officer determines that you did not provide a satisfactory explanation for your responses on this workstyle assessment, you may be discharged from the Canadian Armed Forces.

- You should also think about the fact that responding dishonestly during the workstyle assessments may get in the way of your long-term interests.
- Responding dishonestly could lead to being selected for a position you are not naturally suited to fulfill.
- You will be unhappy if you are selected for a position that requires a different workstyle than your own, and this may impact your career and safety.
• For example, infantry soldiers are expected to operate and maintain weapons and equipment in combat situations.
• To fulfill this role, an infantry soldier needs to remain calm under immense pressure.
• If you answer questions assessing how well you handle stress dishonestly, by exaggerating your ability to handle stress, you may be assigned as an infantry soldier.
• In this role, you may perform poorly and unsafely during combat situations.
• Therefore, it is critical that you answer the questions on the workstyle assessments honestly.

• Thus, if you respond dishonestly on these assessments, you may end up in a position that you’re not naturally suited for, and, as a result, you may perform poorly.
• This will, in turn, lead to limited career opportunities and possibly interfere with the safety of yourself and others.

Instructions: Workstyle Assessment #1
• Before filling out the assessment, please take a few minutes to carefully review the instructions for the first workstyle assessment.
• This assessment takes approximately 5 minutes to complete.
• Once you’re finished, please stop and wait for instructions.
• Please begin now.

Instructions: Workstyle Assessment #2
• Before filling out the assessment, please take a few minutes to carefully review the instructions for the second workstyle assessment.
• This assessment takes approximately 15 minutes to complete.
• Once you’re finished, please stop and wait for instruction to continue.
• Please begin now.

Service Number Request Form
• We will request your service number in order to link your scores from today to some other information.
  — This includes your scores on workstyle and aptitude tests during your recruitment and training performance data.
  — We will only use this information to help improve the accuracy of the workstyle assessments that are filled out today.
• You may decline to provide your service number and still participate.
• Colin Kemp from DQMPRA will link the data using your service number.
• Colin Kemp will remove the service number from the data file before providing the data to the researchers from Western or conducting any analyses.
• We will only look at overall trends in the data and are not interested in your individual responses.
• The data you provide today will only be used for research purposes.
• THIS WILL NOT AFFECT YOUR CAREER WITH THE Canadian Armed Forces IN ANY WAY.
Instructions: Follow-Up Measures

• Now that you’re finished completing the workstyle assessments, we would like you to stop role-playing and answer the questions naturally, as you would normally.
• As a reminder, your answers are strictly confidential and will only be used for research purposes.

Instructions: Feelings during Tests

• The first measure examines how you felt while filling out the workstyle assessments.
• Please fill this out, as honestly as possible, by circling the answer that is most true.

Instructions: Accuracy & Fairness

• This survey will ask you about how you felt about the accuracy and fairness of workstyle assessments during the role-play exercise.
• As a reminder, your answers will be kept confidential and will only be used for research purposes.
• Please answer as honestly as possible.

Instructions: Life events

• This set of questions asks you about how you think about some life events and opportunities.
• Please indicate your answer to each question by circling the answer you identify with.

Instructions: Test-Taking Ability

• Now we’re interested in how well you think you can take tests and look good to potential employers.
• Please answer these items as honestly as possible.
• Remember, your answers will only be examined at the group level and be used for research purposes.

Instructions: Post-Study Survey

• Finally, we would like to ask if you tried to make yourself look more positive or desirable when filling out the assessments during the role-play exercise.
• If so, we would also like to ask how and why you made yourself look more positive.
• Remember, your answers will only be examined at the group level and used for research purposes.
• Your honest answers here will help us improve the workstyle assessments.

Debriefing

• Thank you for participating in our study today!
• What exactly are we studying?
  – Some people try to provide an overly positive image of themselves during job applications when filling out workstyle assessments.
  – We want to see if different types of test instructions can influence applicants’ tendency to be honest or dishonest.
  – Participants were assigned to read different sets of test instructions and we will compare how people answered when they read different instructions.
• Questions?
Appendix C: Moral Suasion Appeal: PowerPoint Presentation

Overview
- The Canadian Armed Forces uses workstyle assessments that help determine which recruits are the best fit for the Canadian Armed Forces
- Workstyle assessments ask recruits a series of questions about their attitudes, thoughts, and behaviours
- Workstyle assessments is another name for Personality Tests, which you have previously filled out for the Canadian Armed Forces

Overview
- You are being invited to participate in a one hour research study
- The study will examine how different sets of assessment instructions may influence the way military recruits respond to workstyle assessments

Overview
- We begin with a broad overview of the study and then assign you to one of 5 different groups
- This may involve a role-play exercise, where we will ask you to fill out workstyle assessments as if you were trying to be selected by the Canadian Armed Forces
- Depending on your assigned group, there may not be a role-play exercise and we will ask you to fill out the workstyle measure as you normally would
- Upon completion of the workstyle assessment, we will ask you to respond, as honestly as possible, to a variety of surveys that will assess topics such as how anxious you were during the assessment and how you perceived the workstyle assessment

Overview
- We will also ask if you are comfortable providing your service number
- The service number allows us to retrieve information from your personnel file
- We will use this information to see how different assessment instructions influence your workstyle scores, and in turn, how these workstyle scores predict training outcomes
- We will also look to see if scores on previous assessments can be used to predict how participants respond during this study

Overview
- Providing your service number is completely voluntary and you can still participate today without providing your service number
- Not providing your service number will not impact your career in anyway, will not be known to your supervisor, and will only be analysed at a aggregate level.
- If you do not wish to participate, you are free to exit the classroom for the duration of the study or stay at your desk to work on other tasks
Overview

- Workstyle assessments have been shown to predict on-the-job performance across a wide variety of jobs and suitability for placement in the CAF
- One concern with these assessments is that people may answer in a way that isn’t the most reflective of their true self
- Our goal is to devise a set of test instructions that will help guide future military recruits to fill out the workstyle measures in a way that is the most self-reflective
- The goal is to help the Canadian Armed Forces make more accurate recruitment/placement decisions

Voluntary Participation

- Participation in this study is voluntary. You may decline to consent, refuse to answer any questions, or withdraw from the study at any time without any penalty or any effect on your career with the Canadian Armed Forces
- Colin Kemp works with DGMPRA, which operates at an arm’s length from operations
  - Colin Kemp has no influence over your career and has no incentive to share your data with your chain of command or personnel in your career branch
  - Thus, you can be confident that we will not breach confidentiality
- Your supervisor will not be made aware of the responses you provide nor will they be placed in your personnel file

Letter of Information and Consent

- Please take a moment to review the letter of information we have handed out
- If you agree to participate in the study today, please sign the consent form that is on the next page
- If you do not agree to participate, feel free to leave at this time
- Please raise your hand if you have any questions
- Please do not open the booklet until instructed to do so

Freedom to Leave

- If you feel anxious or uncomfortable at any point during or after the role-play exercise, you may leave without consequence
- We will not communicate your decision to participate to your supervisors
- If you would like to discontinue your participation, you can simply discard your answers into the garbage bins at the front of the room

Role-Play Exercise

- We would like you to engage in a role-play exercise
- In this exercise, we will ask you to complete two assessments as though you have not yet been accepted into the Canadian Armed Forces
- Why? We aim to improve the accuracy of the assessments used during recruitment. Thus, we need you to answer in a way that reflects how people answer during the recruitment process
- Slides that are part of the role-play exercise will have a purple background
- Remember: All of your answers today, including those given during the role-play, will only be used for research purposes and will not affect your career with the Canadian Armed Forces in any way

ROLE-PLAY BEGINS

ROLE-PLAY EXERCISE
Instructions: Workstyle

- *Imagine* you are about to fill out two workstyle assessments that ask questions about how you typically think, feel, and act.
- *Imagine* they are very important assessments as they are used to help make decisions about selection into the Canadian Armed Forces.
- Responding honestly to these assessments ensures that the most well-suited recruits are selected.
- Most people, including those at the Canadian Armed Forces and in the general public, consider responding dishonestly to workstyle assessments as morally wrong.

- To illustrate the importance of responding honestly to the workstyle assessments, we want you to imagine yourself in the following situation. Think carefully about how you would feel.
- *Imagine* your platoon has been navigating through a war-torn region.

- There have been several car bombs and violent explosions that have seriously harmed or killed some of your fellow soldiers over the past year.
- You’re in the middle of patrolling and your platoon encounters a massive explosion from an Improvised Explosive Device (IED).
- Your commanding officer discovers that two soldiers from your group are severely wounded from the explosion.

- Later, you learn that an intelligence officer had received information indicating there was an imminent threat of IEDs in your area.
- However, the intelligence officer forgot to inform your group’s commanding officer.
- Now, imagine the intelligence officer was only placed in his position because he responded dishonestly on the workstyle assessments you are about to respond to.
- In his responses to the workstyle assessments, he made it appear as though he was more conscientious and detail-oriented than he really is.
So, you can see that honest responses help ensure recruits are placed in positions in the Canadian Armed Forces that they are naturally suited for, and in turn, are likely to perform well in their positions. Providing dishonest responses is morally wrong and could result in tragedies like the one just described.

**ROLE-PLAY EXERCISE**

Instructions: Workstyle Assessment #1

- Before filling out the assessment, please take a few minutes to carefully review the instructions for the first workstyle assessment labelled.
- This assessment takes approximately 5 minutes to complete.
- Once you’re finished, please stop and wait for instructions.
- Please begin now.

**ROLE-PLAY EXERCISE**

Instructions: Workstyle Assessment #2

- Before filling out the assessment, please take a few minutes to carefully review the instructions for the second workstyle assessment.
- This assessment takes approximately 15 minutes to complete.
- Once you’re finished, please stop and wait for instruction to continue.
- Please begin now.

**ROLE-PLAY EXERCISE**

END OF ROLE-PLAY

Service Number Request Form

- We will request your service number in order to link your scores from today to some other information.
  - This includes your scores on workstyle and aptitude tests during your recruitment and training performance data.
  - We will only use this information to help improve the accuracy of the workstyle assessments that are filled out today.
- You may decline to provide your service number and still participate.
- Colin Kemp from DMPRA will link the data using your service number.
- Colin Kemp will remove the service number from the data file before providing the data to the researchers from Western or conducting any analyses.
- We will only look at overall trends in the data and are not interested in your individual responses.
- The data you provide today will only be used for research purposes.
- **THIS WILL NOT AFFECT YOUR CAREER WITH THE Canadian Armed Forces IN ANY WAY.**

Instructions: Follow-Up Measures

- Now that you’re finished completing the workstyle assessments, we would like you to stop role-playing and answer the questions naturally, as you would normally.
- As a reminder, your answers are strictly confidential and will only be used for research purposes.

**ROLE-PLAY EXERCISE**
Instructions

- Now we would like you to fill out some additional measures.
- For each measure, we will review the instructions and prompt you when to start.
- Before answering, please pay careful attention to the different response categories, which are different for each measure.
- You will see a stop page when you finish each measure.
- Please wait until instructed to move forward.

Instructions: Feelings during Tests

- The first measure examines how you felt while filling out the workstyle assessments.
- Please fill this out, as honestly as possible, by circling the answer that is most true.

Instructions: Accuracy & Fairness

- This survey will ask you about how you felt about the accuracy and fairness of workstyle assessments during the role-play exercise.
- As a reminder, your answers will be kept confidential and will only be used for research purposes.
- Please answer as honestly as possible.

Instructions: Life events

- This set of questions asks you about how you think about some life events and opportunities.
- Please indicate your answer to each question by circling the answer you identify with.

Instructions: Test-Taking Ability

- Now we're interested in how well you think you can take tests and look good to potential employers.
- Please answer these items as honestly as possible.
- Remember, your answers will only be examined at the group level and be used for research purposes.

Instructions: Post-Study Survey

- Finally, we would like to ask if you tried to make yourself look more positive or desirable when filling out the assessments during the role-play exercise.
- If so, we would also like to ask how and why you made yourself look more positive.
- Remember, your answers will only be examined at the group level and used for research purposes.
- Your honest answers here will help us improve the workstyle assessments.

Debriefing

- Thank you for participating in our study today!
- What exactly are we studying?
  - Some people try to provide an overly positive image of themselves during job applications when filling out workstyle assessments.
  - We want to see if different types of test instructions can influence applicants' tendency to be honest or dishonest.
  - Participants were assigned to read different sets of test instructions and we will compare how people answered when they read different instructions.
- Questions?
Appendix D: Traditional Faking Warning: PowerPoint Presentation

Overview

- The Canadian Armed Forces uses workstyle assessments that help determine which recruits are the best fit for the Canadian Armed Forces
- Workstyle assessments ask recruits a series of questions about their attitudes, thoughts, and behaviors
- Workstyle assessments is another name for Personality Tests, which you have previously filled out for the Canadian Armed Forces

Overview

- You are being invited to participate in a one hour research study
- The study will examine how different sets of assessment instructions may influence the way military recruits respond to workstyle assessments

Overview

- We begin with a broad overview of the study and then assign you to one of 5 different groups
- This may involve a role-play exercise, where we will ask you to fill out workstyle assessments as if you were trying to be selected by the Canadian Armed Forces
- Depending on your assigned group, there may not be a role-play exercise and we will ask you to fill out the workstyle measure as you normally would
- Upon completion of the workstyle assessment, we will ask you to respond, as honestly as possible, to a variety of survey questions such as how anxious you were during the assessment and how you perceived the workstyle assessment

Overview

- We will also ask if you are comfortable providing your service number
- The service number allows us to retrieve information from your personnel file
- We will use this information to see how different assessment instructions influence your workstyle scores, and in turn, how these workstyle scores predict training outcomes
- We will also look to see if scores on previous assessments can be used to predict how participants respond during this study

Overview

- Providing your service number is completely voluntary and you can still participate today without providing your service number
- Not providing your service number will not impact your career in any way, will not be known to your supervisor, and will only be analysed at an aggregate level.
- If you do not wish to participate, you are free to exit the classroom for the duration of the study or stay at your desk to work on other tasks
### Overview

- Workstyle assessments have been shown to predict on-the-job performance across a wide variety of jobs and suitability for placement in the CAF.
- One concern with these assessments is that people may answer in a way that isn’t the most reflective of their true self.
- Our goal is to devise a set of test instructions that will help guide future military recruits to fill out the workstyle measures in a way that is the most self-reflective.
- The goal is to help the Canadian Armed Forces make more accurate recruitment/placement decisions.

### Voluntary Participation

- Participation in this study is voluntary. You may decline to consent, refuse to answer any questions, or withdraw from the study at any time without any penalty or any effect on your career with the Canadian Armed Forces.
- Colin Kemp works with DCMPIA, which operates at an arm’s length from operations.
  - Colin Kemp has no influence over your career and has no incentive to share your data with your chain of command or personnel in your career branch.
  - Thus, you can be confident that we will not breach confidentiality.
- Your supervisor will not be made aware of the responses you provide nor will they be placed in your personnel file.

### Letter of Information and Consent

- Please take a moment to review the letter of information we have handed out.
- If you agree to participate in the study today, please sign the consent form that is on the next page.
- If you do not agree to participate, feel free to leave at this time.
- Please raise your hand if you have any questions.
- Please do not open the booklet until instructed to do so.

### Freedom to Leave

- If you feel anxious or uncomfortable at any point during or after the role-play exercise, you may leave without consequence.
- We will not communicate your decision to participate to your supervisors.
- If you would like to discontinue your participation, you can simply discard your answers into the garbage bins at the front of the room.

### Role-Play Exercise

- We would like you to engage in a role-play exercise.
- In this exercise, we will ask you to complete two assessments as though you have not yet been accepted into the Canadian Armed Forces.
- Why? We aim to improve the accuracy of the assessments used during recruitment. Thus, we need you to answer in a way that reflects how people answer during the recruitment process.
- Remember: All of your answers today, including those given during the role-play, will only be used for research purposes and will not affect your career with the Canadian Armed Forces in any way.
Instructions: Workstyle

- imagine you are about to fill out two workstyle assessments that ask questions about how you typically think, feel, and act
- imagine they are very important assessments as they are used to help make decisions about selection into the Canadian Armed Forces

Instructions: Workstyle

- Answer Honestly
- There are no ‘right’ or ‘wrong’ answers and it is important that you respond as honestly and accurately as you can
- Keep in mind your responses may be verified and individuals who do not respond honestly may be removed from the selection process

Instructions: Workstyle Assessment #1

- Before filling out the assessment, please take a few minutes to carefully review the instructions for the first workstyle assessment
- This assessment takes approximately 5 minutes to complete
- Once you’re finished, please stop and wait for instructions
- Please begin now

Instructions: Workstyle Assessment #2

- Before filling out the assessment, please take a few minutes to carefully review the instructions for the second workstyle assessment labelled “Workstyle Assessment #2”
- This assessment takes approximately 15 minutes to complete
- Once you’re finished, please stop and wait for instruction to continue
- Please begin now

Service Number Request Form

- We will request your service number in order to link your scores from today to some other information
- This includes your scores on workstyle and aptitude tests during your recruitment and training performance data
- We will only use this information to help improve the accuracy of the workstyle assessments that are filled out today
- You may decline to provide your service number and still participate
- Colin Kemp from DGMPRA will link the data using your service number
- Colin Kemp will remove the service number from the data file before providing the data to the researchers from Western or conducting any analysis
- We will only look at overall trends in the data and are not interested in your individual responses
- The data you provide today will only be used for research purposes
- THIS WILL NOT AFFECT YOUR CAREER WITH THE Canadian Armed Forces IN ANY WAY

END OF ROLE-PLAY
Instructions: Follow-Up Measures

- Now that you’re finished completing the workstyle assessments, we would like you to stop role-playing and answer the questions naturally, as you would normally
- As a reminder, your answers are strictly confidential and will only be used for research purposes

Instructions

- Now we would like you to fill out some additional measures
- For each measure, we will review the instructions and prompt you when to start
- Before answering, please pay careful attention to the different response categories, which are different for each measure
- You will see a stop page when you finish each measure
- Please wait until instructed to move forward

Instructions: Feelings during Tests

- The first measure examines how you felt while filling out the workstyle assessments
- Please fill this out, as honestly as possible, by circling the answer that is most true

Instructions: Accuracy & Fairness

- This survey will ask you about how you felt about the accuracy and fairness of workstyle assessments during the role-play exercise
- As a reminder, your answers will be kept confidential and will only be used for research purposes
- Please answer as honestly as possible

Instructions: Life events

- This set of questions asks you about how you think about some life events and opportunities
- Please indicate your answer to each question by circling the answer you identify with

Instructions: Test-Taking Ability

- Now we’re interested in how well you think you can take tests and look good to potential employers
- Please answer these items as honestly as possible
- Remember, your answers will only be examined at the group level and be used for research purposes
### Instructions: Post-Study Survey

- Finally, we would like to ask if you tried to make yourself look more positive or desirable when filling out the assessments during the role-play exercise.
- If so, we would also like to ask how and why you made yourself look more positive.
- Remember, your answers will only be examined at the group level and used for research purposes.
- Your honest answers here will help us improve the workstyle assessments.

### Debriefing

- Thank you for participating in our study today!
- What exactly are we studying?
  - Some people try to provide an overly positive image of themselves during job applications when filling out workstyle assessments.
  - We want to see if different types of test instructions can influence applicants’ tendency to be honest or dishonest.
  - Participants were assigned to read different sets of test instructions and we will compare how people answered when they read different instructions.
- Questions?
Appendix E: No Faking Warning: PowerPoint Presentation

Overview

- The Canadian Armed Forces uses workstyle assessments that help determine which recruits are the best fit for the Canadian Armed Forces.
- Workstyle assessments ask recruits a series of questions about their attitudes, thoughts, and behaviours.
- Workstyle assessments is another name for Personality Tests, which you have previously filled out for the Canadian Armed Forces.

Overview

- You are being invited to participate in a one hour research study.
- The study will examine how different sets of assessment instructions may influence the way military recruits respond to workstyle assessments.

Overview

- We begin with a broad overview of the study and then assign you to one of 5 different groups.
- This may involve a role-play exercise, where we will ask you to fill out workstyle assessments as if you were trying to be selected by the Canadian Armed Forces.
- Depending on your assigned group, there may not be a role-play exercise and we will ask you to fill out the workstyle measure as you normally would.
- Upon completion of the workstyle assessment, we will ask you to respond, as honestly as possible, to a variety of surveys that will assess topics such as how anxious you were during the assessment and how you perceived the workstyle assessment.

Overview

- We will also ask if you are comfortable providing your service number.
- The service number allows us to retrieve information from your personnel file.
- We will use this information to see how different assessment instructions influence your workstyle scores, and in turn, how these workstyle scores predict training outcomes.
- We will also look to see if scores on previous assessments can be used to predict how participants respond during this study.

Overview

- Providing your service number is completely voluntary and you can still participate today without providing your service number.
- Not providing your service number will not impact your career in anyway, will not be known to your supervisor, and will only be analysed at a aggregate level.
- If you do not wish to participate, you are free to exit the classroom for the duration of the study or stay at your desk to work on other tasks.
Overview

- Workstyle assessments have been shown to predict on-the-job performance across a wide variety of jobs and suitability for placement in the CAF
- One concern with these assessments is that people may answer in a way that isn't the most reflective of their true self
- Our goal is to devise a set of test instructions that will help guide future military recruits to fill out the workstyle measures in a way that is the most self-reflective
- The goal is to help the Canadian Armed Forces make more accurate recruitment/placement decisions

Voluntary Participation

- Participation in this study is voluntary. You may decline to consent, refuse to answer any questions, or withdraw from the study at any time without any penalty or any effect on your career with the Canadian Armed Forces
- Colin Kemp works with DGMPRA, which operates at an arm's length from operations
  - Colin Kemp has no influence over your career and has no incentive to share your data with your chain of command or personnel in your career branch
  - Thus, you can be confident that we will not breach confidentiality
- Your supervisor will not be made aware of the responses you provide nor will they be placed in your personnel file

Letter of Information and Consent

- Please take a moment to review the letter of information we have handed out
- If you agree to participate in the study today, please sign the consent form that is on the next page
- If you do not agree to participate, feel free to leave at this time
- Please raise your hand if you have any questions
- Please do not open the booklet until instructed to do so

Freedom to Leave

- If you feel anxious or uncomfortable at any point during or after the role-play exercise, you may leave without consequence
- We will not communicate your decision to participate to your supervisors
- If you would like to discontinue your participation, you can simply discard your answers into the garbage bins at the front of the room

Instructions

- You’ve been assigned to the group with no role-play exercise
- We’d like you fill out a workstyle assessment that assesses your attitudes, thoughts, and behaviours
- Please fill this measure out as honestly as possible
- As a reminder, your scores will not be placed on your personnel file or used against you in any way

Instructions: Workstyle Assessment #1

- Before filling out the assessment, please take a few minutes to carefully review the instructions for the first workstyle assessment
- This assessment takes approximately 5 minutes to complete
- Once you’re finished, please stop and wait until instructed to move on.
- Please begin now
Instructions: Workstyle Assessment #2

- Before filling out the assessment, please take a few minutes to carefully review the instructions for the second workstyle assessment labelled "Workstyle Assessment #2".
- This assessment takes approximately 15 minutes to complete.
- Once you're finished, please stop and wait for instructions to continue.
- Please begin now.

Instructions: Service Number Request Form

- We will request your service number in order to link your scores from today to some other information.
  - This includes your scores on workstyle and aptitude tests during your recruitment and training performance data.
  - We will only use this information to help improve the accuracy of the workstyle assessments that are filled out today.
- You may decline to provide your service number and still participate.
- Colin Kemp from DGMPSA will link the data using your service number.
- Colin Kemp will remove the service number from the data file before providing the data to the researchers from Western or conducting any analyses.
- We will only look at overall trends in the data and are not interested in your individual responses.
- The data you provide today will only be used for research purposes.
- THIS WILL NOT AFFECT YOUR CAREER WITH THE Canadian Armed Forces IN ANY WAY.

Instructions: Follow-Up Measures

- Now that you're finished completing the workstyle assessments, we would like you to answer the questions naturally, as you would normally.
- As a reminder, your answers are strictly confidential and will only be used for research purposes.

Instructions: Accuracy & Fairness

- This survey will ask you about how you felt about the accuracy and fairness of workstyle assessments.
- As a reminder, your answers will be kept confidential and will only be used for research purposes.
- Please answer as honestly as possible.

Instructions: Feelings during Tests

- The first measure examines how you felt while filling out the workstyle assessments.
- Please fill this out, as honestly as possible, by circling the answer that is most true.
Instructions: Life events

- This set of questions asks you about how you think about some life events and opportunities
- Please indicate your answer to each question by circling the answer you identify with

Instructions: Test-Taking Ability

- Now we're interested in how well you think you can take tests and look good to potential employers
- Please answer these items as honestly as possible
- Remember, your answers will only be examined at the group level and be used for research purposes

Instructions: Post-Study Survey

- Finally, we would like to ask if you tried to make yourself look more positive or desirable when filling out the assessments
- If so, we would also like to ask how and why you made yourself look more positive
- Remember, your answers will only be examined at the group level and used for research purposes
- Your honest answers here will help us improve the workstyle assessments

Debriefing

- Thank you for participating in our study today!
- What exactly are we studying?
  - Some people try to provide an overly positive image of themselves during job applications when filling out workstyle assessments
  - We want to see if different types of test instructions can influence applicants' tendency to be honest or dishonest
  - Participants were assigned to read different sets of test instructions and we will compare how people answered when they read different instructions
- Questions?
### Appendix F: Idiosyncratic Item Responding Scale

<table>
<thead>
<tr>
<th>Item</th>
<th>Interested Applicants (Time 1)</th>
<th>Uninterested Applicants (Time 1)</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I start tasks right away</td>
<td>4.24 ± 0.84</td>
<td>3.27 ± 0.92</td>
<td>0.97</td>
</tr>
<tr>
<td>2. I sympathize with the homeless</td>
<td>4.15 ± 0.96</td>
<td>3.23 ± 0.96</td>
<td>0.93</td>
</tr>
<tr>
<td>3. I fear for the worst</td>
<td>3.54 ± 1.24</td>
<td>2.61 ± 1.09</td>
<td>0.93</td>
</tr>
<tr>
<td>4. I love a good fight</td>
<td>4.06 ± 1.18</td>
<td>3.13 ± 1.27</td>
<td>0.93</td>
</tr>
<tr>
<td>5. I act without thinking</td>
<td>4.38 ± 0.82</td>
<td>3.51 ± 0.98</td>
<td>0.87</td>
</tr>
<tr>
<td>6. I insult people</td>
<td>4.61 ± 0.74</td>
<td>3.75 ± 1.05</td>
<td>0.86</td>
</tr>
<tr>
<td>7. I rush into things</td>
<td>3.98 ± 0.99</td>
<td>3.13 ± 0.98</td>
<td>0.85</td>
</tr>
<tr>
<td>8. I do not like art</td>
<td>4.40 ± 0.97</td>
<td>3.56 ± 1.09</td>
<td>0.84</td>
</tr>
<tr>
<td>9. I jump into things without thinking</td>
<td>4.10 ± 1.04</td>
<td>3.25 ± 1.04</td>
<td>0.84</td>
</tr>
<tr>
<td>10. I do not enjoy going to art museums</td>
<td>4.15 ± 1.15</td>
<td>3.31 ± 1.16</td>
<td>0.84</td>
</tr>
</tbody>
</table>
Chapter 5

5. General Discussion

This dissertation had two primary objectives: a) help improve the assessment of faking in research and practice; and b) develop practical tools to help reduce applicant faking on personality testing. The first goal addressed a major gap in the faking literature, where to my knowledge, no study has comprehensively compared common faking indices in a single study. Without such study, researchers and practitioners may choose faking assessments without knowing which ones best distinguish faking behaviours, and without knowledge that different faking assessments are, in fact, assessing the same construct.

The results from Study 1 suggest that, consistent with prior research, Residualized Change Scores was the most consistent faking index (Feeney & Goffin, 2015). This index was the second-best measure at distinguishing when between people’s applicant versus non-applicant personality profiles, and between job interested and uninterested responders. That being said, Residualized Change Scores were not the single best in either set of analyses, suggesting that it may be advantageous to use multiple faking indices. Second, Residualized Change Scores are impractical, requiring both job applicant and non-applicant personality scores, which employers are unlikely to have. Three other indices stood out as having utility in discriminating between applicants and non-applicants, which were Blatant Extreme Responding, Communal Impression Management, and Idiosyncratic Item Responding. All three indices may be easily used in applied settings. One concern is that these measures did not converge, as more than 50% of the variance in each faking index was unexplained by the others. In turn, correcting applicant scores or disqualifying applicants using only one of these measures would be inappropriate. Instead, we suggest a triangulation approach: we recommend the combined use of Idiosyncratic Item
Responding, Blatant Extreme Responding, and Communal Impression Management, and when possible, Residualized Change Scores.

Interestingly, our results suggested that Bogus Items—a common faking measure in the literature—were ineffective. This is problematic as it raises questions about the implications of any study that relied exclusively on Bogus Items as a measure of faking (c.f., Dwight & Donovan, 2003; Fan et al., 2012; Levashina et al., 2009). There have been other studies that cast doubt on the validity of Bogus Items. For example, Overclaiming (Bing et al., 2011; Paulhus et al., 2003), a close cousin of Bogus Items, assesses the tendency of applicants to misrepresent their general knowledge. In a prior investigation, overclaiming was ineffective in measuring applicant faking (Feeney & Goffin, 2015). Additionally, Overclaiming appears to be conflated with personality measures (Dunlop et al., 2016). Similarly, we found that two covariance indices (Christiansen et al., 2017) were also ineffective, as they failed to predict applicant faking or relate to any other measure of applicant faking. We speculate this is due to the reliance on non-applicant samples, where applicant faking tends to be much more severe (R. Hogan, 2005; Ones, Dilchert, Viswesvaran, & Judge, 2007).

The findings also suggest that researchers and practitioners should likely question the use of several faking indices, as they appear to perform less well than other existing options. For example, Percent Agreement and Individual Change Scores both performed well at discriminating between applicant and non-applicant personality scores, but were inferior to Residualized Change Scores, and were both confounded with legitimate personality variance. Within-Subject Correlations and Within-Subjects Variance of the Differences both demonstrated little utility in differentiating applicants from non-applicants, and also failed to converge with
their stronger alternatives. Given that both faking indices require a repeated measures design, they may not offer practicality nor validity.

An additional objective of this dissertation was to develop three faking dissuasion messages that were either more effective than standard warnings or perceived less negatively by job applicants and military recruits. Data from Study 2 revealed that none of the messages reduced faking or fostered more positive applicant reactions. By contrast, in Study 3, we found that the IAW was effective in reducing faking among military recruits, but that none of the other messages—including one used in prior studies—were effective relative to the control group. Overall, these results have several potential implications for the field.

The most plausible explanation for the reason the IAW worked in Study 3—but not Study 2—is that faking dissuasion messages are only effective when applicants believe that they have something to lose by being caught. Military recruits fill out the personality inventory well into the selection process, and in turn, would know they are close to placement. In their case, the risk of being caught for faking is considerably higher. The recruits also completed the personality assessments in military facilities, and in turn, they may have found the threat of detection more credible. By contrast, our job applicants in Study 2 were in the earliest stage of the selection process, and therefore, were aware that they needed to score favorably to pass screening. In this case, the benefits of faking may have outweighed the costs of detection. As a result, none of our faking dissuasion messages were able to dissuade job applicants from faking. This may present a conundrum for practitioners; the utility of personality tests is in the early screening stages, as they are inexpensive and easy to administer, but this is when applicants are most likely to misrepresent themselves—possibly weakening the validity of the personality scores.
The second major take-home is that warnings of immediate consequences are more effective than “soft messages” that try to gently dissuade people from faking. Scholars have suggested that the latter approaches may work (Goffin & Boyd, 2009; Pace & Borman, 2006; Uruena & Robie, 2011), and yet to-date, there is no compelling evidence that these messages are as persuasive as threats of immediate consequences. We know that the majority of applicants fake (Griffith et al., 2007; Holden et al., 2017), and perhaps most applicants already know this is occurring. This might explain why “soft messages” that appeal to morality or long-term consequences do not sufficiently dissuade applicants—they believe they will not secure employment unless they fake. By contrast, a faking dissuasion message that emphasizes immediate disqualification for faking with a credible explanation for how faking can be detected, targets applicants where their primary interest lays—securing employment. As a result, we suggest that future studies focus on teasing apart new ways to make that message resonate louder. The persuasion and decision-making literature may help improve the potency of these messages.

The third lesson is that people do not react strongly to faking dissuasion messages under motivated conditions. Across two studies, we found limited evidence that job applicants or military recruits responded negatively to our faking dissuasion messages. Yet, motivated responders’ major goal appears to be securing employment, and such, these responders may give little thought to how the faking dissuasion message reflects the hiring organization. These findings are consistent with the literature; three studies have found inconsistent findings in how applicants react to faking warnings (Converse et al., 2008; Fan et al., 2012; McFarland, 2003). It is possible that applicants are responding to some other aspect of the procedure or materials used in the testing, and in turn, this may explain why there are inconsistencies across studies.
The final and more general application from my dissertation is that Amazon’s Mechanical Turk can be used successfully to measure employment scenarios even when motivated job applicants are required. We employed an innovative design where mTurk users believed they were being screened for a real position. This allowed us to discern between both those who were interested and those who were not, providing unusually rich data. We were also able to recruit the same pool of interested applicants for a follow-up session, where they served as normal unmotivated respondents. As a result, we were able to assess the effect of our job application simulation both between samples and within the same sample, allowing for firmer conclusions. This type of research design may help benefit researchers in the future who would like to test this same sample. The one caution is that researchers may want to ask participants if they have previously participated in similar studies, so that the same type of deception is not being employed with the same people, and in turn, weakening the credibility of the manipulation.

5.1. Future Directions in Faking Research

Study 1 left a few major unanswered questions: 1) Are the faking assessment results generalizable to more competitive jobs and to different personality instruments? 2) How should an organization use these scores? 3) Once an organization triangulates applicant faking with multiple faking indices, what is the best approach to classifying an applicant as a faker or not? Unfortunately, the data from this dissertation do not address these questions. The results of Study 1 also prompt an additional question: why did some indices perform poorly in our investigation? Bogus Items, for example, are commonly used and assumed to measure faking. Study 1 yielded no evidence that faking on Bogus Items translates into faking on personality assessments. Future research will need to replicate these findings and seek to uncover why Bogus Items fail, before firm conclusions about their use may be reached.
Future research should also examine how faking prevention methods can be improved to reduce faking using multiple accurate faking assessments. For example, Jackson and Paunonen (1980) advocated for the creation of personality measures that are resistant to socially desirable responding. In the test creation process, items that are likely to receive a socially desirable response can be removed. In theory, this same approach could be used using Idiosyncratic Item Responding, Blatant Extreme Responding, and Residualized Change Scores. As a result, a publisher could produce personality tests that are resistant to multiple forms of faking, rather than just social desirability.

Finally, scholars have championed faking dissuasion messages as an effective tool to combat applicant faking. Studies 2 and 3 in this dissertation provide some reason for caution—faking dissuasion messages are not consistently effective and can not be viewed as a silver bullet. Future research will need to specifically test what factors moderate the efficacy of faking dissuasion messages. In a post-hoc literature review, we found that studies that accuse job applicants of faking during the test tend to be the ones that have the largest effects, whereas faking dissuasion messages tend to be ineffective in applicant screening scenarios or when applicants perceive that they have little to lose by faking. These findings tend to converge on risk serving as a potential moderator of when faking dissuasion messages are effective. Future research could test this directly and also seek to determine other contributing variables. Research in this area will better contribute to our understanding of applicant faking and provide guidance for practitioners in our field.
5.2. References


Curriculum Vitae
Justin Ryan Feeney

Education

Doctor of Philosophy, University of Western Ontario  
Industrial & Organizational Psychology  
2018

Master of Science, University of Western Ontario  
Industrial & Organizational Psychology  
2011

Honors Bachelor of Science with High Distinction, University of Toronto  
Psychology  
2009

Peer-Reviewed Publications


Scholarly Conference Presentations


*Honorable Mention for Top Student Submission*


*Featured Poster and Awarded SIOP’s John Flanagan Award*


---

**Research in the Press**

4. *Men’s Health* (Feb 2012). “3 Things You MUST Do Before a Job Interview” by Cassie Shortsleeve
Grants, Scholarships, and Awards

1. Dean’s Research Grant (Fall 2017-Winter 2018; $5000)
   University of Regina, Hill and Levene Schools of Business
2. Ontario Graduate Scholarship (2014-2015; $15,000)
   Ontario Ministry of Training, Colleges, and Universities
3. Joseph-Armand Bombardier Canada Graduate Scholarship Doctoral (2011-2014; $105,000)
   Social Sciences and Humanities Research Council
4. John Flanagan Award (2012; $1,500)
   Society for Industrial and Organizational Psychology (Top Student Contribution & Featured Poster)
5. Ontario Graduate Scholarship (2010-2011; $15,000)
   Ontario Ministry of Training, Colleges, and Universities
   University of Western Ontario (Student Competition Grant for thesis expenses)
7. Joseph-Armand Bombardier Canada Graduate Scholarship Master’s (2009-2010; $17,500)
   Social Sciences and Humanities Research Council
8. Douglas N. Jackson Memorial Award (2009; $500)
   University of Western Ontario (Incoming I-O psychology student with the highest research potential)
   Canadian Psychological Association

Teaching and Knowledge Translation

Courses Taught

1. Instructor, Staffing Organizations (Fall 2017 – Present)
   University of Regina
2. Instructor, Introduction to Organizational Behavior (Fall 2017 - Present)
   University of Regina
3. Sessional Lecturer, Research Methods in Psychology (Fall 2014 – Fall 2016)
   King’s University College at the University of Western Ontario
4. Sessional Lecturer, Introduction to Personality Theory (Fall 2015)
   King’s University College at the University of Western Ontario
5. Sessional Lecturer, Introduction to Industrial and Organizational Psychology (Summer 2013)
   University of Western Ontario

Service Contributions

University Service

1. Search Advisory Committee for the Associate Dean, Undergraduate Programs (2018).
   University of Regina, Business Administration.
2. Search Advisory Committee for the Associate Dean Research and Graduate Programs (2018).
   University of Regina, Business Administration.
4. Business Representative for the Executive of Council for the University of Regina (July 2018 to June 2020).
5. Business Representative, Market Supplements Committee, University of Regina Faculty Association (May 2018 – June 2019).
6. Lead Case Coach for the human resources team the Organizational Behaviour Case Competition (OBCC) at the Ted Rogers School of Management in Toronto, ON (December 2017 – January 2018). Team placed first overall in the competition. University of Regina.
7. SSRHC Canada Graduate Scholarship—Master’s Reviewer for the Faculty of Graduate Studies (January 2018). University of Regina.
8. Business Representative for the Faculty if Arts (Fall 2017 – Present). University of Regina.
10. Faculty and Student Liaison Committee (2011). University of Western Ontario.
11. Undergraduate Representative, Psychology Search Committee for Department Chair (2009). University of Toronto at Mississauga.

### Professional and Community Service

1. Mock Interviewer and Trainer (Fall 2017 - Present)  
   *Regina Open Door Society*
2. Student Representative, Board of Directors (2012 - 2014)  
   *Canadian Psychological Association*
3. Chair, Section for Students (2011 - 2014)  
   *Canadian Psychological Association*
4. Member, Professional Affairs Committee (2012 - 2013)  
   *Canadian Psychological Association*
5. Undergraduate Affairs Coordinator, Section for Students (2010 - 2012)  
   *Canadian Psychological Association*

### Peer Reviewing

7. Conference Reviewer (2015 - Present). *Society of Industrial and Organizational Psychology*

### Professional Affiliations

1. Society for Industrial and Organizational Psychology (2009 - Present)
2. Canadian Society for Industrial and Organizational Psychology (2009 - Present)
3. Canadian Psychological Association (2009 - Present)
4. Academy of Management (2010 - Present)
Industry Experience

Consulting

   - Conducted personnel assessments including employment interviews, cognitive ability testing, personality testing, emotional intelligence testing, and leadership analysis
   - Implemented evidence-based personnel selection procedures and screening tools
   - Programmed and implemented web-based personnel selection system
   - Composed comprehensive assessments and coaching reports for clients

   - Conducted literature reviews for large national partner on military-related topics
   - Consulted with management to ensure reports were consistent with client needs
   - Received reliability clearance from Public Works and Government Services Canada

3. Consultant, Everspring Farms (Summer 2015)
   - Conducted literature review to examine the efficacy of profit-sharing
   - Wrote executive report defining and analyzing profit-sharing programs
   - Developed recommendations to implement profit-sharing for the organization

   - Instructed workshop on operationalizing job performance and program evaluations
   - Provided consultation with managers and therapists concerning key criterion measures
   - Produced executive summary with practical recommendations for the board of directors

5. Student Practicum, Surrey Place Center, Toronto (2007 - 2008)
   - Scored and evaluated cognitive and neurological assessments
   - Wrote and examined case reports and files
   - Interviewed clients with dual-diagnosis of developmental delay and comorbid diagnosis