

IDENTIFYING FAKING ON SELF-REPORT PERSONALITY INVENTORIES:
RELATIVE MERITS OF TRADITIONAL LIE SCALES, NEW LIE SCALES, RESPONSE
PATTERNS, AND RESPONSE TIMES

by

Christine Elizabeth Lambert

A thesis submitted to the Department of Psychology

In conformity with the requirements for
the degree of Master of Science

Queen's University

Kingston, Ontario, Canada

(September, 2013)

Copyright © Christine Lambert, 2013

Abstract

The use of personality tests throughout Canadian society is based on the assumption that their results are valid. However, research has shown that individuals can, and do, fake their responses on personality inventories. Individuals may fake good, emphasizing their positive characteristics, or fake bad, emphasizing negative characteristics, in order to obtain a desired outcome. Recent research has provided support for a congruence model of faking, which states that schema-consistent responses are provided more quickly than schema-inconsistent responses. Faking successfully, without being detected by validity indices, requires balancing favourable and unfavourable responses, regardless of the faking schema a participant adopts. This demand results in cognitive fatigue over time, producing increasingly unbalanced response patterns. Two studies were conducted to evaluate the efficacy of the congruence and cognitive overload models of faking in detecting instructed faking, and to examine whether these models or the newly developed Faking Response Strategy Scales provide added value in detecting faking relative to currently established gold-standard measures. Results showed that all of the self-report scales examined—whether traditional or new—were valid detectors of faking, which supports their ongoing use. However, results highlighted the weakness of the Impression Management subscale of the Balanced Inventory of Desirable Responding, the current gold-standard in the field, in providing added value relative to other scales. Response latency data supported the congruence model of faking, but results for the cognitive overload model were mixed: Study 1 data supported the cognitive overload model, but time constraints introduced in Study 2 seem to have caused random responding, rather than increasing cognitive overload as was intended. Results supported a multidimensional model of faking, and show that adding measures of response latency and response pattern can enhance the ability of traditional measures to detect faking.

These findings have important theoretical and practical implications for methods of detecting faking and for the understanding of cognitive processes underlying faking.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Ronald Holden, for his endless patience and guidance. Thanks for sticking by me through all of the minor and major research setbacks, for reminding me that we learn most from our mistakes, and for never losing faith.

A big thank you to Spencer Arbuckle and Cara Chen for the endless hours they spent in the lab, which made this project possible.

Finally, thank you to my family and friends for their invaluable support.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iv
Table of Contents.....	v
List of Figures.....	vii
List of Tables.....	viii
Chapter 1 Introduction.....	1
Faking on Personality Inventories.....	1
Detecting Faking Using Response Patterns: The Cognitive Overload Model of Faking.....	2
Inducing Cognitive Overload Using Time Limits.....	3
Detecting Faking Using Response Times: The Congruence Model of Faking.....	5
Detecting Faking Using New Response Strategy Scales.....	6
Objectives.....	9
Chapter 2 Method.....	11
Participants.....	11
Study 1.....	11
Study 2.....	11
Materials.....	12
NEO-FFI (Costa & McCrae, 1992).....	12
New Faking Response Strategy Scales (Holden, Passey, Jay, & Book, 2009).....	12
Balanced Inventory of Desirable Responding (Paulhus, 1998).....	13
Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960).....	14
Jackson's (1984) Desirability Subscale.....	14
Holden Psychological Screening Inventory (Holden, 1996).....	14
Procedure.....	15
Study 1.....	15
Study 2.....	16
Chapter 3 Results.....	17
Manipulation Check.....	18
Traditional Measures.....	18
Discriminant Function Analysis.....	18
Stepwise Discriminant Function Analysis.....	19
New Faking Response Strategy Scales.....	22
Added Value of New Faking Response Strategy Scales.....	24

Item Order	25
Added Value of Item Order.	31
Restricted Response Times	32
Added Value of Response Times.....	36
Impression Management: An Industry Standard.....	36
Impression Management vs. Item Order	37
Impression Management vs. Response Times.....	38
Chapter 4 Summary of Findings	39
Traditional Measures	39
New Faking Response Strategy Scales	40
Response Patterns/Item Order: Efficacy of the Cognitive Overload Model.....	44
Time Limits.....	45
Response Latencies: Efficacy of the Congruence Model	45
Chapter 5 Discussion	49
Dimensionality of Faking	50
Support for the Congruence Model.....	51
Challenging the Cognitive Overload Model	51
Limitations	53
Review of Objectives	54
References.....	56
Appendix A Military Induction Instructions.....	61
Appendix B Faking Strategy Scales Developed by Holden, Passey, Jay, & Book (2009)	63
Competitive Self-Promotion	63
Social Sensitivity	64
Chameleonic Unpredictability	64
Strong-Mindedness	65
Showing Off.....	65
Appendix C Letter of Information, Informed Consent, & Debriefing Form	66
Appendix D An Example of the Preparation and Computer of Response Latencies	72

List of Figures

<i>Figure 1.</i> Frequency of correlations between positivity of response and item order for Study 1 honest condition.	29
<i>Figure 2.</i> Frequency of correlations between positivity of response and item order for Study 1 fake good condition.	29
<i>Figure 3.</i> Frequency of correlations between positivity of response and item order for Study 1 fake bad condition.	29
<i>Figure 4.</i> Frequency of correlations between positivity of response and item order for Study 2 honest condition.	30
<i>Figure 5.</i> Frequency of correlations between positivity of response and item order for Study 2 fake good condition.	30
<i>Figure 6.</i> Frequency of correlations between positivity of response and item order for Study 2 fake bad condition.	30
<i>Figure 7.</i> Mean adjusted item response latency by group and type of response.....	35

List of Tables

Table 1 <i>Faking Strategies Generated by Real-World Test Respondents, Faking-Instructed Undergraduates, and Literature Review</i>	8
Table 2 <i>New Faking Response Strategy Scales: Subscale Descriptions & Sample Items</i>	9
Table 3 <i>Alpha Coefficient Reliabilities</i>	17
Table 4 <i>Study 1: Classification by Individual Traditional Lie Scale</i>	19
Table 5 <i>Study 2: Classification by Individual Traditional Lie Scale</i>	19
Table 6 <i>Study 1: Group Means (SDs) by Faking Condition</i>	20
Table 7 <i>Study 2: Group Means (SDs) by Faking Condition</i>	21
Table 8 <i>Study 1: Group Means (SDs) by Faking Condition</i>	23
Table 9 <i>Study 2: Group Means (SDs) by Faking Condition</i>	23
Table 10 <i>Mean Correlation Between Positivity of Response & Item Order</i>	28
Table 11 <i>Study 1: NEO-FFI Scale Scores by Faking Condition</i>	32
Table 12 <i>Study 2: NEO-FFI Scale Scores by Faking Condition</i>	32
Table 13 <i>Study 1: Group Means (SDs) by Faking Condition</i>	35
Table 14 <i>Traditional Lie Scales: Individual & Stepwise DFAs</i>	41
Table 15 <i>New Faking Response Strategy Scales: Individual DFAs</i>	42
Table 16 <i>New Faking Response Strategy Scales: Stepwise DFAs</i>	43
Table 17 <i>Item Order & Response Latencies: Individual DFAs</i>	46
Table 18 <i>Item Order: Stepwise DFAs</i>	47
Table 19 <i>Response Latencies: Stepwise DFAs</i>	48

Chapter 1

Introduction

Faking on Personality Inventories

Self-report personality tests are used in many settings in Canadian society. They help individuals discover their vocational interests, assess the clinical status of forensic, counselling, and psychiatric patients, and evaluate the suitability of job applicants. However, although personality tests are developed to be valid across large samples, it is not uncommon for individual results to be invalid due to participant faking (Butcher et al., 1997; Rosse et al., 1998). Holden and Book (2012, p. 71) define faking as “intentional misrepresentation in self-report.” Participants are likely to fake results in high-stakes situations in an attempt to increase their chances of attaining a desired outcome. They may "fake good" by exaggerating their positive characteristics on an integrity assessment for a job application, or "fake bad" by underperforming in an assessment of academic abilities in order to qualify for additional support (Holden, 2007; Viswesvaran & Ones, 1999).

Accordingly, detecting and preventing faking on self-report personality inventories has become a matter of practical and scientific importance. In test development, many personality inventories include validity indices. The Minnesota Multiphasic Personality Inventory (MMPI-2), which is frequently used for screening applicants for jobs that have a direct effect on public safety or security, includes seven validity indices (Butcher, Morfitt, Rouse, & Holden, 1997). Other entire inventories have been developed to assess individuals' response styles, such as the

Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1998), the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960), and Jackson's Desirability Scale (1984).

Despite the widespread use of validity indices and response-style scales, some research suggests that there are inherent limitations in using self-report scale scores to detect participant dissimulation. Some fakers can avoid detection on standard validity scales by faking on the validity scales themselves (Book, Holden, Starzyk, Wasylkiw, & Edwards, 2006; MacNeil & Holden, 2006) using strategies such as role-playing (Kroger & Turnbull, 1975). The fact that lie scales themselves may be susceptible to faking points to the potential value of more objective methods of detection. Methods that could provide added value in detecting faking, relative to established inventories, might be particularly beneficial. Response latencies and response patterns over time are two alternative methods that, although associated with self-report, are based more in non-conscious or non-deliberate participant behaviour and thus may be less susceptible to faking (van Hooft & Born, 2012).

The current study evaluates the efficacy of a variety of new methods for detecting faking on personality inventories, and compares these alternative methods to the current gold-standard inventories in the field. New methods examined include item order, response times, and a newly developed inventory of faking response strategies.

Detecting Faking Using Response Patterns: The Cognitive Overload Model of Faking

The cognitive overload model of faking proposes that because the nature of faking can be quite sophisticated, and faking successfully (i.e., faking without being caught) requires balancing schema-consistent and schema-inconsistent answers, the process of faking is cognitively

demanding and will result in fatigue effects over time. Response pattern data have shown that as participants progress through a personality inventory, they give responses that are more extreme in a direction consistent with their faking schema (Holden, Fekken, & Cotton, 1991; Holden, Kroner, Fekken, & Popham, 1992). Holden and Book (2009) were able to use hybrid Rasch-latent class modeling to identify participants' faking schemas based on distinct response patterns. Participants following a fake-good schema gave more favourable responses over time, whereas those following a fake-bad schema gave less favourable responses over time. Participants responding honestly had response patterns unrelated to item order. This research shows the potential value of response patterns across items in identifying fakers.

The current study aims to develop a new way to operationally measure the influence of cognitive overload on the valence of participant response. If this new method proves useful in detecting participant faking, it would replicate the conclusion by Holden and Book (2009) that cognitive overload occurs during the faking process, and confirm the usefulness of this theory in detecting participant faking. Importantly, replicating these findings using a new statistical technique would also demonstrate that the significant findings by Holden and Book (2009) were not measure-specific, but are in fact reliable and of meaningful theoretical importance.

Inducing Cognitive Overload Using Time Limits

Because faking is believed to be cognitively demanding and take increasing amounts of time to perform successfully, researchers have hypothesized that constraining the amount of time participants have in which to respond may make them unable to fake successfully (Holden, Wood, & Tomashewski, 2001). Past research in our laboratory constrained response times on

Jackson's (1984) Personality Research Form; university undergraduates in the experimental condition were required to read and respond to each item in 6 seconds or less. Because this limited response time had no effect on whether participants could fake effectively, a second study was conducted with response times restricted further, to 2, 3, or 5 seconds (between participants). Items in the second study were presented via audiotape in order to control for inter-individual differences in reading speed. Even with these shorter time intervals, restricting response time had no effect on the influence of faking on scale validity: participants continued to be able to fake effectively (Holden, Wood, & Tomashewski, 2001).

A subsequent study was conducted in order to extend these findings to other personality inventories. In this study, participants responded to NEO-FFI and BIDR scale items, which were presented via audiotape. Participants were required to respond to each item in 1.5, 2.5, or 3.5 seconds, and reported that completing the task was a substantial challenge. Contrary to study hypotheses, these high levels of time pressure still did not influence the effects of faking on self- or peer-based criterion validity of any NEO-FFI scales (Holden, Wood, & Tomashewski, 2001).

Study 2 aims to extend the findings detailed above by constraining response times for NEO-FFI items still further. As in previous studies, participants will respond to each item via key press and their response times will be recorded by a computer. However, in Study 2, participants are required to both read and respond to NEO-FFI items in 2 seconds each. We hypothesize that being required to read and respond to inventory items under time limits will more effectively induce cognitive overload than responding to audiotaped items. We also control

for the effects of individual differences in reading speed by following response-latency standardization procedures recommended by Holden, Fekken, and Cotton (1991).

Detecting Faking Using Response Times: The Congruence Model of Faking

Research in our laboratory has provided key findings about the process of faking and how it might be detected, particularly about the value of using response time data. Adapted from schema theory, Holden and colleagues (Holden & Hibbs, 1995; Holden & Kroner, 1992; Holden, Kroner, Fekken & Popham, 1992; Holden & Troister, 2009) have found empirical support for a congruence model of personality test faking. According to this model, individuals who have adopted a faking schema will take relatively longer to produce schema-inconsistent responses than to produce schema-consistent responses. This model hinges on the assumption that individuals who have adopted a faking schema will also produce schema-inconsistent responses in order to avoid being identified as fakers by the standard validity indices built into personality inventories. Thus, individuals who are faking good (i.e., trying to present themselves in the best possible light) will still provide some negative answers about themselves, which will take relatively longer to be produced because they are schema-inconsistent. Similarly, individuals faking bad (i.e., attempting to convey an unfavourable impression of themselves) will still give some positive responses, which will take longer for them to produce. Whatever a participant's faking schema, schema-inconsistent responses take longer than schema-consistent responses.

In 2012, van Hooft and Born examined the congruence model of faking in a study using eye-tracking. Contrary to what would be predicted by the congruence model, they found no significant condition by item desirability interaction. Instead, they found that participants

responded significantly faster when they were instructed to fake good than when they were instructed to respond honestly. This finding supports an alternate theory (Hsu, Santelli, Hsu, 1989) that faking involves a primitive form of cognitive processing, based solely on semantic interpretation of item content, which takes less processing time than the self-referenced interpretation of item content that is required for honest responding.

However, we propose that the study by van Hooft and Born (2012) did not support the congruence model of faking because they did not aggregate data appropriately using the recommended double-standardization procedure developed by Holden and colleagues (Holden, 1998; Holden & Kroner, 1992; Holden et al., 1992; see Appendix D). The technique of Holden and colleagues involves separately aggregating positive and negative *answers* whereas van Hooft and Born mistakenly differentiate between positively and negatively keyed *items* (and not answers). Thus, we expect that once we appropriately distinguish item answers rather than item keying, the current study will support the congruence model of faking, refuting the findings of van Hooft and Born (2012).

Detecting Faking Using New Response Strategy Scales

Very little is known about the cognitive processes underlying faking. Research on faking makes assumptions about the nature of these processes, but little work has investigated the cognitive processes that are involved in faking successfully and unsuccessfully. Most current lie scales have been developed from theories about the cognitive processes underlying faking, but due to the difficulty in accessing these processes, these theories remain largely speculative.

One possible way to determine the cognitive processes underlying faking is simply to ask respondents what they do when they dissimulate. Though there is some debate over the matter, it seems quite possible that the people who actually do the dissimulation might have relevant information about the process of faking. In a recent study, researchers interviewed real-world test respondents, who had recently completed a personality questionnaire as part of a job application, a worker's compensation claim, a child custody evaluation, marital counselling, a mental health evaluation, or a legal assessment. Respondents were asked to describe how they decided on their answers, as well as to imagine tactics that they might use in order to fake to achieve favourable (fake good) and unfavourable (fake bad) outcomes. Researchers also interviewed participants who had been instructed to fake-good or fake-bad in a laboratory experiment, and asked them to describe how they generated their answers (Holden, Passey, Jay, & Book, 2009). Combining information from respondents and a literature review yielded a list of 11 response strategies that participants might use in order to fake-good or fake-bad within a selection context (see Table 1).

In a subsequent experimental study (Holden, Passey, Jay, & Book, 2009), participants were instructed to adopt 1 of these 11 response strategies. Discriminant function analysis revealed a two-factor structure, dominated by the two dimensions of Personal Effectiveness and Socialization. This work supports the notion that faking is a multidimensional entity, not a single, honesty-centred continuum, and may consist of even more than two dimensions (Holden & Evoy, 2005). The complex structure of faking has been confirmed by others (Kitching & Paulhus, 2009).

Table 1

Faking Strategies Generated by Real-World Test Respondents, Faking-Instructed Undergraduates, and Literature Review

1	Maximize selection chances.
2	Minimize selection chances.
3	Answer honestly.
4	Answer opposite to usual to minimize selection chances.
5	Answer opposite to usual.
6	Emphasize positive characteristics to maximize selection chances.
7	Emphasize negative characteristics to minimize selection chances.
8	Take into account the perspective of the person doing the selection in order to maximize selection chances.
9	Take into account the perspective of the person doing the selection in order to minimize selection chances.
10	Role-play as a typical incumbent in order to maximize selection chances.
11	Role-play as a typical incumbent in order to minimize selection chances.

Also included in the study just described were 57 experimental items selected or rationally written by the authors to have content relevant to the 11 response tactics (Holden, Passey, Jay, & Book, 2009). Thirteen items were selected from Snyder's (1974) Self-Monitoring scale, and 44 items were written by the authors. Responses to these items were correlated and subjected to principal components analysis. Results converged on five components, and experimental items were appropriately divided into five new faking response strategy scales: Competitive Self-Promotion, Social Sensitivity, Chameleonic Unpredictability, Strong-Mindedness, and Showing Off. These subscales each capture bipolar individual differences that

individuals who are faking, either good or bad, may reflect in their responses. Higher scores on each subscale indicate greater levels of endorsing each individual difference (see Table 2; see Appendix B for a complete list of items).

Table 2

New Faking Response Strategy Scales: Subscale Descriptions & Sample Items

Subscale	Personality Characteristic	Sample Items
Competitive Self-Promotion (CSP)	Tendency to consider one's self as superior to others and to compete interpersonally to ensure maximal personal gain	Losing is for others, not for me. When I want to, I can make other people admire my characteristics.
Social Sensitivity (SS)	Tendency to frequently experience and display care and concern for the well-being of others; wanting others to consider one sensitive interpersonally	The opinions of other people matter to me. I am an open and honest person.
Chameleonic Unpredictability (CU)	Tendency to react unpredictably, inconsistently, or randomly	I'm not always the person I appear to be. My behaviour is unpredictable.
Strong-Mindedness (SM)	Tendency to stick to one's opinions despite disagreement from others	I am not afraid to state my own opinion. I would not change my opinions (or the way I do things) in order to please someone else or to win favour.
Showing Off (SO)	Tendency to enjoy being the centre of attention	I would like to be a performer of some sort. I like to stand out in a crowd.

Objectives

The current study seeks to make the following novel contributions to the literature on detecting faking on personality inventories:

1. Evaluate the efficacy of the congruence model of faking (Holden, Kroner, Fekken & Popham, 1992) in detecting participant faking by using response time data.
2. Extend the findings of Holden and Book (2009) by developing a new way to operationally measure the impact of cognitive overload on faking and to use this measure to examine the efficacy of the cognitive overload model in detecting participant faking.
3. Further evaluate the cognitive overload model of faking by requiring participants to respond in a shortened period of time (Study 2), inducing excessive cognitive overload that may disrupt the process of faking
4. Examine the efficacy of the new faking response strategy scales developed by Holden, Passey, Jay, and Book (2009) in detecting participant faking.
5. Evaluate the added value that the cognitive overload model, the congruence model, and the new faking response strategy scales can provide in detecting participant faking relative to established gold-standard detection measures.

Thus, this study aims to make meaningful theoretical contributions by evaluating the accuracy of the cognitive overload and congruence models of faking. Any added value in detecting faking provided by these models or by the new faking response strategy scales may also be of practical value in detecting the presence of faking in real-life situations.

Chapter 2

Method

The current project includes data from two experimental studies, the first of which was conducted during the 2011-2012 academic year. Data for the second study were collected during the 2012-2013 academic year.

Participants

Study 1.

Two hundred and ninety-four undergraduate students from Queen's University were recruited to participate in this study using the introductory psychology course subject pool and by posting flyers around campus. Participants were compensated with one course credit (introductory psychology students) or \$15 for an hour of their time. The data from one participant were lost due to a computer malfunction, resulting in a total of 293 individuals (66 men and 227 women) being included in the analyses. The participants were between the ages of 17 and 24, with an average age of 18.82 years ($SD = 1.04$).

Study 2.

Three hundred undergraduate students (57 men and 243 women) from Queen's University were recruited to participate in this study using the introductory psychology course subject pool and by posting flyers around campus. Participants were compensated with one course credit (introductory psychology students) or \$15 for an hour of their time. The participants were between the ages of 17 and 28, with an average age of 19.22 years ($SD = 1.79$).

Materials.

All participants in both studies completed the 60-item NEO Five-Factor Inventory (NEO-FFI; Costa & McCrae, 1992), the new Faking Response Strategy Scales developed by Holden, Passey, Jay, and Book (2009), and a set of industry-standard and psychometrically strong validity scales [Paulhus' (1998) Balanced Inventory of Desirable Responding scales of Self-Deceptive Enhancement and Impression Management; the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960), Jackson's (1984) Desirability Scale, and the Holden Psychological Screening Inventory (Holden, 1996)].

NEO-FFI (Costa & McCrae, 1992).

The NEO-FFI is a 60-item self-report measure that assesses the 5-factor personality model (Neuroticism, Openness, Agreeableness, Conscientiousness, and Extraversion; 12 items per factor). Participants respond to all items using a 5-point scale (0 = strongly disagree; 4 = strongly agree). Higher scores on the NEO-FFI subscales are associated with greater levels of its factors. The NEO-FFI has coefficient alpha reliabilities above .73 and validities, based on correlations with spousal report, above .33 (Costa & McCrae, 1992).

New Faking Response Strategy Scales (Holden, Passey, Jay, & Book, 2009).

As described above, the New Faking Response Strategy Scales (NFRSS) were developed through principal components analysis of 57 experimental items selected to reflect 11 response tactics that individuals who faked in real-world situations reported using (Holden, Passey, Jay, &

Book, 2009). Five components emerged and were appropriately divided into 5 subscales each measuring a bipolar personality characteristic: Competitive Self-Promotion (15 items; coefficient alpha = .90), Social Sensitivity (11 items; coefficient alpha = .89), Chameleonic Unpredictability (6 items; coefficient alpha = .75), Strong-Mindedness (4 items; coefficient alpha = .61), and Showoff (5 items; coefficient alpha = .76). Higher scores on each subscale indicate greater levels of endorsing each personality characteristic (see Appendix B).

Balanced Inventory of Desirable Responding (Paulhus, 1998).

The Balanced Inventory of Desirable Responding (BIDR; most recently published as the Paulhus Deception Scales) measures an individual's tendency to give socially desirable responses on self-report inventories. It consists of 40 items, which are rated on a 7-point scale (1 = Totally disagree; 4 = Neutral; 7 = Totally agree). The BIDR contains two subscales: Self-Deceptive Enhancement (SDE), the tendency to unconsciously give unrealistically favourable self-descriptions, and Impression Management (IM), the tendency to dissimulate by giving unrealistically positive self-descriptions (Paulhus, 1998). SDE occurs at an unconscious level, and measures an individual's honest, though inaccurate, beliefs about him/herself. In contrast, IM measures a conscious effort to dissimulate or fake good. Higher scores indicate greater tendencies toward SDE and IM. The measure has an overall internal reliability of .83, and subscales have internal reliabilities of .70 (SDE) and .81 (IM) in a college student sample. The BIDR has high concurrent validity with the Marlowe-Crowne Social Desirability Scale ($r = .73$), discussed next.

Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960).

The Marlowe-Crowne Social Desirability Scale (MCSDS) consists of 33 items that were selected to have socially desirable content and low probability of occurrence (sample item: I never hesitate to go out of my way to help someone in trouble). Higher scores indicate that a respondent is presenting him/herself in an unrealistically favourable manner. The MCSDS has an internal reliability coefficient alpha of .88 in a sample of undergraduate students, and high concurrent validity as established through correlations with the MMPI validity scales (Marlowe & Crowne, 1960).

Jackson's (1984) Desirability Subscale.

Jackson's Social Desirability subscale (JSDS) of the Personality Research Form (Jackson, 1984) measures the extent to which an individual describes him/herself in desirable terms or presents a favourable picture of him/herself in response to personality statements. This favourable self-presentation may occur consciously or unconsciously, and may be accurate or inaccurate. Items are judged to be either True or False. Higher scores indicate a greater tendency for positive self-presentation. The Social Desirability subscale consists of 16 items and has an internal reliability coefficient of .68 in a college sample (Jackson, 1984).

Holden Psychological Screening Inventory (Holden, 1996).

The Holden Psychological Screening Inventory (HPSI) consists of 36 items designed to provide a very brief measure of three major dimensions of psychopathology: Psychiatric

Symptomatology (PS; 12 items), Social Symptomatology (SS; 12 items), and Depression (D; 12 items). The scale also provides a measure of Total Symptomatology, which is used as a validity index (extreme scores may indicate faking). Items are rated on a 5-point scale (1 = Never; 2 = Sometimes; 3 = Often; 4 = Very Often; 5 = Always), and the scale is scored so that higher scores indicate higher levels of psychopathology. Established median coefficient alpha reliabilities across clinical and non-clinical samples were .74 for Psychiatric Symptomatology, .73 for Social Symptomatology, .84 for Depression, and .83 for Total Symptomatology. Validities, based on correlation with roommate report, ranged between .28 for Psychiatric Symptomatology and .41 for Depression.

Procedure

Study 1.

Each experimental session began with obtaining written informed consent. Participants were then asked to answer the study materials as if they were being screened for military induction, under one of three conditions. Participants were randomly assigned to: 1) complete the measures under standard instructions; 2) fake answers to maximize their chances of being inducted (i.e., fake good); or 3) fake answers to minimize their chances of being inducted (i.e., fake bad; see Appendix A for Military Induction Instructions). All participants were warned of the presence of validity checks to detect faking, were asked to do their best to avoid being detected, and were given an incentive to do so: for each 25 participants, a \$50 prize was awarded to the participant who was farthest from activating the validity checks.

Following completion of the series of computer questionnaires, participants were asked to indicate their degree of compliance to their condition-specific faking instructions. Responses were made on a 9-point Likert-type scale ranging from 1 (Definitely did not comply with my instructions) to 9 (Always complied with my instructions). This served as a manipulation check to ensure that individuals were indeed responding to the questionnaire items in a manner that was consistent and in accordance with their instructions.

Study 2.

The procedure for Study 2 was identical to that described in Study 1, with one important exception: NEO-FFI items were administered under time constraints. Each written item appeared on the screen for exactly 2.0 seconds, along with the response scale. Participants were required to both read the item and select their response using a mouse before the item and response scale disappeared. The 2.0 second time limit was selected through pilot testing to determine the minimum presentation time that would allow participants to successfully read each item but also experience considerable pressure to do so quickly, thus inducing cognitive overload. It was hypothesized that with these time restraints, participants would not have the cognitive resources available to both fake and avoid detection.

Chapter 3

Results

Table 3

Alpha Coefficient Reliabilities

Scale	Subscale	Coefficient Alpha in Scale Manual	Study 1 Alpha	Study 2 Alpha
NEO-FFI	Stability (Neuroticism Reversed)	.89	.94	.81
	Extraversion	.79	.91	.78
	Openness	.76	.72	.63
	Agreeableness	.74	.91	.81
	Conscientiousness	.84	.97	.86
New Faking Response	Competitive Self-Promotion	.90	.83	.72
Strategy Scales	Social Sensitivity	.61	.90	.87
	Chameleonic Unpredictability	.89	.81	.83
	Strong-Mindedness	.76	.68	.51
	Showoff	.75	.76	.71
BIDR	Self-Deceptive Enhancement	.70	.82	.81
	Impression Management	.81	.90	.89
	Total	.83	.90	.88
MCSDS	N.A.	.88	.95	.95
Jackson's Personality Research Form	Social Desirability	.68	.92	.92
HPSI	Psychiatric Symptomatology	.74	N.A.	.94
	Social Symptomatology	.73		.94
	Depression	.84		.94
	Total Symptomatology	.83		.97

For both studies, the BIDR, MCSDS, JSDS, and HPSI were scored according to the standard scoring key. The NEO-FFI Neuroticism subscale was reflected so that all items were in the direction of positivity. Following this reflection, this subscale now measured Emotional Stability rather than Neuroticism.

Due to computer error, three items of the HPSI were omitted from data collection in Study 1. This scale was therefore omitted from further analyses for this study.

For all of the scales examined, domain scale coefficient alpha reliabilities were found to be adequately high to proceed with analysis, with the majority of the study alphas exceeding those published in the manuals of the respective scales (see Table 3).

Manipulation Check

Participants in both studies indicated that they did comply with instructions on a 9-point Likert scale ($M = 7.38$, $SD = 1.52$ for Study 1; $M = 6.60$, $SD = 1.74$ for Study 2). Participants' reported level of compliance was significantly higher in Study 1, $t(591) = 5.84$, $p < .00001$, Cohen's $d = .478$, or a moderate effect size. This significant difference may reflect the fact that time limits were imposed on the NEO-FFI items in Study 2: participants may have felt that they were less able to respond appropriately when they were also required to respond very quickly.

Traditional Measures

Discriminant Function Analysis.

For each study, I conducted four (Study 1) or five (Study 2) separate discriminant function analyses (assuming equal prior probabilities) in order to determine whether the five traditional measures of faking – the MCSDS, JSDS, HPSI, and IM and SDE subscales of the

BIDR – could detect experimentally-induced faking condition. Individually, and in both studies, each of these traditional lie detection scales predicted faking group membership significantly above the rate expected by chance (33% for three groups; see Tables 4 & 5).

Table 4

Study 1: Classification by Individual Traditional Lie Scale

	% Correctly Classified	Wilks' λ	$\chi^2(2, N = 294)$	p
MCSDS	76.1	.30	349.06	< .001
JSDS	70.0	.34	314.13	< .001
IM of BIDR	67.9	.44	236.29	< .001
SDE of BIDR	46.4	.82	58.4	< .001

Table 5

Study 2: Classification by Individual Traditional Lie Scale

	% Correctly Classified	Wilks' λ	$\chi^2(2, N = 300)$	p
MCSDS	76.7	.34	324.33	< .001
JSDS	69.7	.32	319.60	< .001
HPSI	76.3	.32	342.87	< .001
IM of BIDR	58.7	.55	175.58	< .001
SDE of BIDR	43.7	.89	33.78	< .001

Stepwise Discriminant Function Analysis.

Next, I conducted a stepwise discriminant function analysis for each set of data, in order to determine the added predictive value of each traditional measure.

For a stepwise discriminant function analysis of the Study 1 data, considering all four traditional lie scales as possible predictors, the order of entry was the MCSDS, the JSDS, and the SDE subscale of the BIDR. Entry stopped after this – the IM of the BIDR had no added value. The correct classification hit rate for these predictors was 83.7%. The overall Wilks' lambda was significant, $\lambda = .20$, $\chi^2(6, N = 294) = 469.27$, $p < .001$, indicating that overall the predictors differentiated significantly among the three faking conditions (fake good, fake bad, and honest responding).

Table 6 displays mean scores by faking condition. The BIDR subscales, MCSDS, and JSDS each differed significantly by faking condition ($p < .001$), which confirms that participants were able to fake by following military induction instructions.

Table 6

Study 1: Group Means (SDs) by Faking Condition

Scale	Subscale	Faking Good (n = 98)	Faking Bad (n = 97)	Honest (n = 98)	F (2, 290)
BIDR	SDE	5.46 (4.22)	3.11 (2.98)	1.80 (2.14)	32.37**
	IM	12.57 (4.17)	2.78 (3.10)	6.43 (3.49)	182.51**
MCSDS	N.A.	25.37 (5.43)	5.86 (5.38)	14.53 (4.93)	338.18**
Jackson's Personality Research Form	Social Desirability	14.17 (1.96)	4.28 (3.71)	11.22 (3.01)	283.34**

** $p < .001$

For a stepwise discriminant function analysis of the Study 2 data, considering all five traditional lie scales as possible predictors, the order of entry was the HPSI, the MCSDS, the

SDE subscale of the BIDR, and the JSDS. Entry stopped after this – the IM of the BIDR had no added value. The correct classification hit rate for these predictors was 79.7%. The overall Wilks' lambda was significant, $\lambda = .20$, $\chi^2(8, N = 300) = 472.68$, $p < .001$, indicating that overall the predictors differentiated significantly among the three faking conditions (fake good, fake bad, and honest responding).

Table 7 displays mean scores by faking condition. The BIDR subscales, MCSDS, JSDS, and HPSI each differed significantly by faking condition ($p < .001$), which confirms that participants were able to fake by following military induction instructions.

Table 7

Study 2: Group Means (SDs) by Faking Condition

Scale	Subscale	Faking Good (n = 101)	Faking Bad (n = 100)	Honest (n = 99)	F (2, 297)
BIDR	SDE	4.52 (4.42)	3.53 (2.53)	1.81 (2.36)	17.89**
	IM	10.98 (5.10)	2.52 (2.67)	5.85 (3.54)	119.71**
MCSDS	N.A.	24.81 (6.46)	5.54 (5.13)	14.36 (5.22)	294.07**
Jackson's Personality Research Form	Social Desirability	14.05 (2.01)	3.85 (3.23)	10.47 (3.29)	319.60**
HPSI	Total	28.01 (14.26)	95.87 (26.94)	47.84 (14.38)	322.58**

** $p < .001$

New Faking Response Strategy Scales.

I conducted another stepwise discriminant function analysis on the data from each study in order to determine whether the five new faking response strategy scale subscales could predict experimentally-induced faking condition.

For the Study 1 data, the order of entry was the Social Sensitivity subscale, the Chameleonic Unpredictability subscale, and the Competitive Self-Promotion subscale. Entry stopped after this – the Strong-Mindedness and Showoff subscales had no added value. The correct classification hit rate for these predictors was 78.1% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .26$, $\chi^2(6, N = 294) = 390.68$, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

For the Study 2 data, the order of entry was the Social Sensitivity subscale, the Chameleonic Unpredictability subscale, the Competitive Self-Promotion subscale, and the Showoff subscale. Entry stopped after this – the Strong-Mindedness subscale had no added value. The correct classification hit rate for these predictors was 74.0% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .32$, $\chi^2(8, N = 300) = 341.12$, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

Tables 8 and 9 display mean scores by faking condition. For Study 1, all of the new faking response strategy scales, except the Showoff scale, differed significantly by faking condition. For Study 2, all of the new faking response strategy scales, except the Strong-Mindedness scale, differed significantly by faking condition. These results confirm that the

Table 8

Study 1: Group Means (SDs) by Faking Condition

Scale	Subscale	Faking Good (n = 98)	Faking Bad (n = 97)	Honest (n = 97)	F(2, 289)
New Faking Response Strategy Scales	Competitive Self- Promotion	55.58 (6.65)	44.08 (10.16)	49.05 (6.86)	50.10**
	Social Sensitivity	43.54 (4.05)	28.95 (8.49)	42.39 (3.86)	186.01**
	Chameleonic	13.98 (3.53)	22.80 (4.01)	18.22 (3.36)	143.14**
	Unpredictability				
	Strong-Mindedness	15.08 (2.63)	13.64 (3.92)	13.67 (2.67)	6.77**
	Showoff	13.80 (3.76)	13.97 (5.11)	13.96 (4.41)	0.05

** $p < .001$

Table 9

Study 2: Group Means (SDs) by Faking Condition

Scale	Subscale	Faking Good (n = 101)	Faking Bad (n = 100)	Honest (n = 99)	F(2, 297)
New Faking Response Strategy Scales	Competitive Self- Promotion	53.08 (7.05)	46.38 (9.07)	48.55 (6.42)	20.35**
	Social Sensitivity	43.22 (4.20)	27.93 (7.90)	41.80 (3.99)	223.07**
	Chameleonic	15.21 (4.05)	24.07 (4.40)	18.45 (3.38)	128.29**
	Unpredictability				
	Strong-Mindedness	14.30 (2.65)	13.52 (3.25)	13.45 (2.73)	2.65
	Showoff	14.85 (3.76)	16.31 (4.82)	13.58 (4.05)	10.39**

** $p < .001$

Competitive Self-Promotion, Social Sensitivity, and Chameleonic Unpredictability scales can identify respondents who are faking.

Added Value of New Faking Response Strategy Scales.

Next, I conducted stepwise discriminant function analyses on the data from each study in order to determine whether the five new faking response strategy scale subscales could provide any added value above the traditional lie scales in predicting experimentally-induced faking condition.

For the Study 1 data, the order of entry was the MCSDS, the Social Sensitivity subscale, the JSDS, the SDE subscale of the BIDR, the Competitive Self-Promotion subscale, and the Chameleonic Unpredictability subscale. Entry stopped after this – the IM subscale of the BIDR, and the Strong-Mindedness and Showoff new faking response strategy scales had no added value. The correct classification hit rate for these predictors was 84.6% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .15$, $\chi^2(12, N = 294) = 548.19$, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

For the Study 2 data, the order of entry was the HPSI, the MCSDS, the Social Sensitivity subscale, the SDE subscale of the BIDR, the JSDS, and the Competitive Self-Promotion subscale. Entry stopped after this – IM subscale of the BIDR, the Chameleonic Unpredictability subscale, the Strong-Mindedness subscale, and the Showoff subscale had no added value. The correct classification hit rate for these predictors was 83.3% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .17$, $\chi^2(12, N = 300) =$

518.14, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

For both Study 1 and Study 2, all of the new faking response strategy scales, except the Showoff scale, differed significantly by faking condition. These results confirm that the Competitive Self-Promotion, Social Sensitivity, Chameleonic Unpredictability, and Strong-Mindedness subscales can identify respondents who are faking.

Item Order.

The cognitive overload model of faking posits that as individuals who are faking get farther through the items in a personality inventory, the valence of their responses will become increasingly extreme in the faked direction: individuals who are faking good will give increasingly favourable responses, whereas individuals who are faking bad will give increasingly unfavourable responses.

Past studies have used hybrid Rasch-latent class modeling to detect faking through response patterns (Holden & Book, 2009; Eid & Zickar, 2007; Zickar, Gibby, & Robie, 2004; Zickar & Robie, 1999). However, the use of this single method means that it is currently unclear whether significant results may be method-specific. In order to extend past findings, it is necessary that the predictive accuracy of response patterns be evaluated using new statistical methods. In order to evaluate the potential value of the cognitive overload model of faking, I examined the correlation between NEO-FFI item order (1 through 60) and item response positivity (0 through 4). We hypothesized that there would be a linear relationship between item order and item response positivity, and that it would differ by faking condition. In particular, we

expected that participants faking good would provide more favourable responses over time (positive correlations), whereas those faking bad would give less favourable responses over time (negative correlations). Participants in the honest condition were hypothesized to show a pattern of responding unrelated to item order (non-significant correlations).

Item order analyses were conducted only for the NEO-FFI items of each study. The NEO-FFI was chosen for this examination of item response pattern because it is an often-used, typical personality measure that is a modal exemplar of personality inventories. Results obtained from examination of the NEO-FFI should thus extend to other inventories, because the impact of cognitive overload on item response pattern is hypothesized to be a general phenomenon that would apply to all personality inventories. First, as previously indicated, the Neuroticism subscale of the NEO-FFI was reversed so that high scores indicated higher levels of emotional stability. Thus, higher scores on all of the NEO-FFI subscales (Stability, Openness, Extraversion, Agreeableness, and Conscientiousness) represented more favourable or positive responses. Next, I computed the correlation between the value of each item and its item order, for all items and all participants who had valid responses for at least 54 of the 60 NEO-FFI items. All 294 participants in Study 1, and 300 participants in Study 2, met this criterion and were included in these analyses.

A discriminant function analysis (assuming equal prior probabilities) was conducted for each study, using the correlation between item positivity and item order as a predictor. For Study 1, this correlation correctly classified 49.5% of individuals as either honest, faking good, or faking bad respondents. The rate of correct classification was slightly lower for the Study 2 data,

at 38.1%. The overall Wilks' lambdas for both studies were significant, $\lambda = .75$, $\chi^2(2, N = 294) = 85.46$, $p < .001$ for Study 1; $\lambda = .97$, $\chi^2(2, N = 300) = 10.69$, $p = .005$ for Study 2, indicating that this measure of response pattern detected group membership at a level significantly above chance.

We hypothesized, based on the cognitive overload model of faking, that participants in the fake good condition would give increasingly positive responses with increased item order, participants in the fake bad condition would give increasingly negative responses with increased order, and that there would be no relationship between response positivity and item order for participants in the honest condition. These predictions would be confirmed by a positive correlation between response positivity and item order for participants in the fake good condition, a negative correlation for participants in the fake bad condition, and no correlation for participants in the honest condition.

Participants in Study 1 showed the predicted correlations in both the fake good and fake bad conditions: the correlation was positive for participants faking good, and negative for participants faking bad (see Table 10). A frequency graph of the correlations for individuals in the Study 1 fake good condition shows that the majority of these participants showed this positive correlation, which went as high as 0.6 for one individual (see Figure 2). A small majority of fake bad participants in Study 1 showed the predicted negative correlation (see Figure 3), causing the slightly negative overall correlation of -0.04. Participants in the honest condition of Study 1 also showed a positive correlation, suggesting that they might have found it more difficult, as time went on, to resist a natural tendency to present themselves positively

(Heine, Lehman, Markus, & Kitayama, 1999; Mezulis, Abramson, Hyde, & Hankin, 2004). The majority of participants in the Study 1 honest condition showed this positive correlation, which went as high as 0.5 (see Figure 1).

The predicted effects were reduced for the Study 2 participants, who responded to the NEO-FFI items under strict time constraints (see Table 10). Participants in the fake good condition continued to show a positive correlation, though it was slightly less than the Study 1 correlation (0.09 for Study 2 versus 0.13 for Study 1). Once again, the majority of participants faking good showed this pattern, although the highest correlation in Study 2 was only 0.4 (see Figure 5). The correlation for Study 2 fake bad participants was slightly positive (0.03 versus -0.04 in Study 1). As for Study 1, fake bad participants in Study 2 showed a wide range of correlations (-0.4 to 0.5 in Study 1; -0.4 to 0.4 in Study 2; see Figure 3 and Figure 6). Participants in the Study 2 honest condition showed a positive correlation overall, though it was similarly reduced (0.07 versus 0.11 in Study 1). The majority of participants showed this positive correlation (see Figure 4).

Table 10

Mean Correlation Between Positivity of Response & Item Order

Experimental Group	Study 1 Correlation (<i>r</i>)	Study 2 Correlation (<i>r</i>)
1. Given standard honest instructions	0.11 (<i>n</i> = 98)	0.07 (<i>n</i> = 98)
2. Given fake good instructions	0.13 (<i>n</i> = 97)	0.09 (<i>n</i> = 101)
3. Given fake bad instructions	-0.04 (<i>n</i> = 98)	0.03 (<i>n</i> = 100)

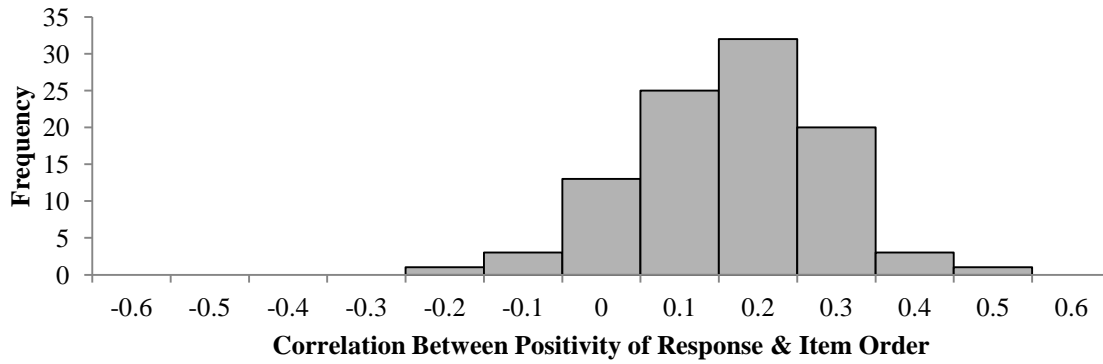


Figure 1. Frequency of correlations between positivity of response and item order for Study 1 honest condition.

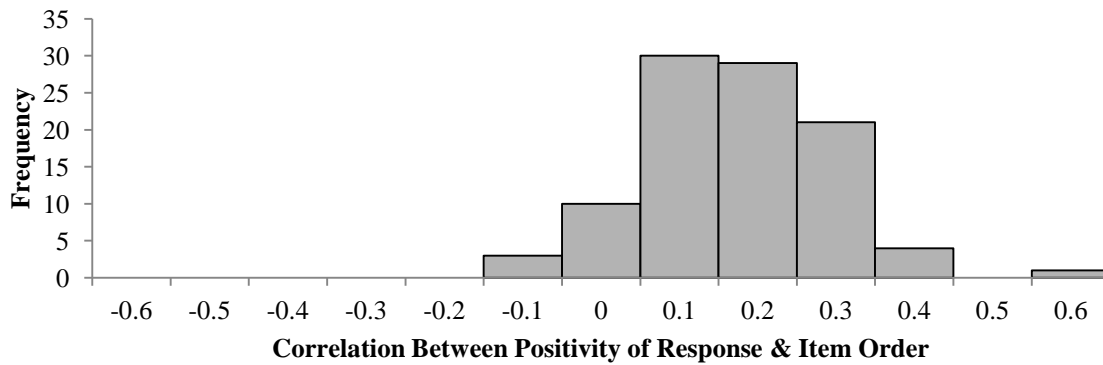


Figure 2. Frequency of correlations between positivity of response and item order for Study 1 fake good condition.

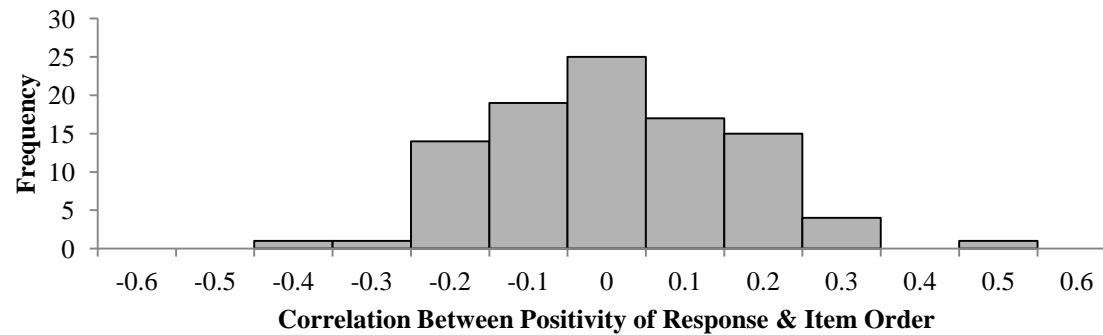


Figure 3. Frequency of correlations between positivity of response and item order for Study 1 fake bad condition.

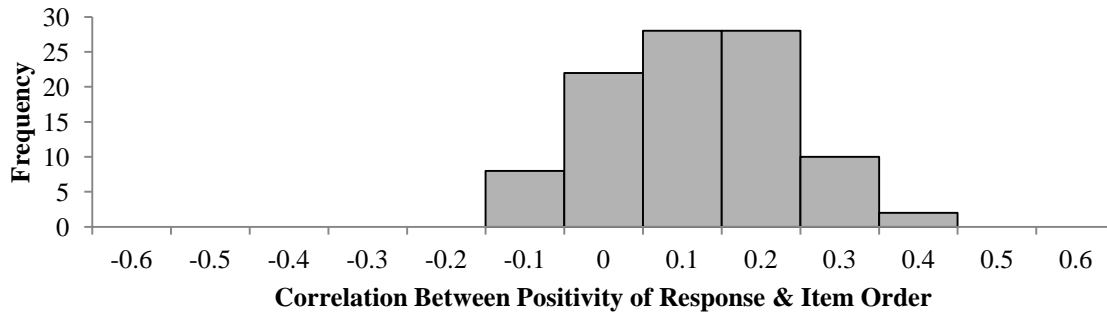


Figure 4. Frequency of correlations between positivity of response and item order for Study 2 honest condition.

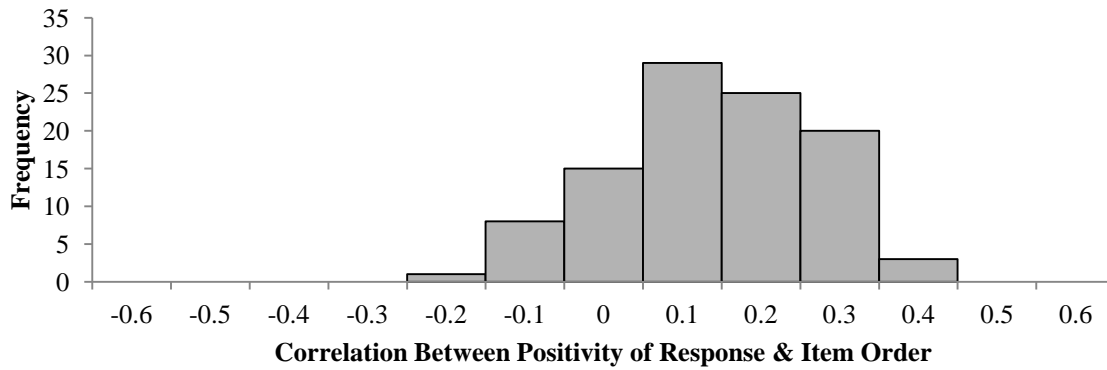


Figure 5. Frequency of correlations between positivity of response and item order for Study 2 fake good condition.

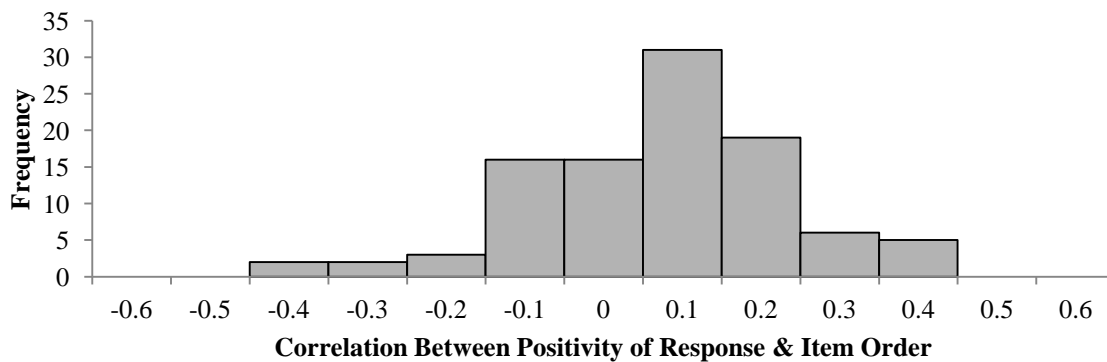


Figure 6. Frequency of correlations between positivity of response and item order for Study 2 fake bad condition.

Added Value of Item Order.

Next, I conducted stepwise discriminant function analyses on the data from each study in order to determine whether the measure of response pattern could provide any added value above the traditional measures in predicting experimentally-induced faking condition. The four traditional lie-detecting scales and the correlation between item positivity and item order were entered as predictors.

For the Study 1 data, the order of entry was the MCSDS, the JSDS, the SDE subscale of the BIDR, and the measure of response pattern. Entry stopped after this – the IM subscale of the BIDR had no added value. The correct classification hit rate for these predictors was 84.0% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .19$, $\chi^2(8, N = 294) = 480.57$, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

For the Study 2 data, the order of entry was the HPSI, the MCSDS, the SDE subscale of the BIDR, and the JSDS. Entry stopped after this – IM subscale of the BIDR, and the measure of response pattern had no added value. The correct classification hit rate for these predictors was 79.3% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .20$, $\chi^2(8, N = 300) = 473.12$, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

For both studies, all of the traditional measures and the measure of response pattern differed significantly by faking condition, which confirms that these measures can effectively identify respondents who are faking.

Restricted Response Times

Participants in Study 2 were required to read and respond to NEO-FFI items in 2 seconds each. Analyses of variance with faking condition as a between subjects variable showed a significant effect of faking condition on each of the Big Five personality dimensions for both Study 1 and Study 2, indicating that Study 2 participants were able to fake effectively despite increased time restrictions (see Table 11 and Table 12).

Table 11

Study 1: NEO-FFI Scale Scores by Faking Condition

NEO-FFI Scale	Faking Good - 2 (n = 98)	Faking Bad - 3 (n = 97)	Honest - 1 (n = 98)	F(2, 292)
Stability	37.00 (5.83)	14.16 (7.94)	26.24 (8.86)	217.43**
Extraversion	35.57 (5.13)	16.66 (7.64)	30.36 (6.44)	221.28**
Openness	28.87 (5.85)	24.25 (8.04)	30.01 (5.73)	20.68**
Agreeableness	34.51 (6.01)	18.16 (9.36)	32.56 (5.90)	147.13**
Conscientiousness	41.46 (5.40)	14.51 (9.00)	33.13 (6.97)	351.43**
** $p < .001$				

Table 12

Study 2: NEO-FFI Scale Scores by Faking Condition

NEO-FFI Scale	Faking Good - 2 (n = 101)	Faking Bad - 3 (n = 100)	Honest - 1 (n = 99)	F(2, 299)
Stability	28.18 (7.00)	22.71 (7.25)	22.72 (6.84)	20.23**
Extraversion	32.21 (4.99)	26.87 (6.99)	29.45 (6.57)	18.40**
Openness	27.56 (4.94)	25.42 (6.68)	28.23 (5.39)	6.59*
Agreeableness	31.07 (5.95)	25.60 (8.11)	29.43 (5.89)	17.46**
Conscientiousness	33.77 (5.37)	26.58 (9.34)	29.52 (5.24)	27.47**
** $p < .001$; * $p < .05$				

Response Times.

Response latencies were only analysed for the NEO-FFI and only for Study 1, where response time was not constrained. Once again, prior to any further analyses, the neuroticism subscale was reversed so that higher scales indicated higher levels of emotional stability. Thus, higher scores on all of the NEO-FFI subscales (stability, openness, extraversion, agreeableness, and conscientiousness) represented more favourable or positive responses.

Following recommended procedures (Holden, 1998; Holden & Kroner, 1992; Holden et al., 1992), raw response latencies were adjusted to reduce the effects of statistical outliers, and standardized twice, once to control for confounding person variables (such as reading speed and gender), and once to control for confounding item variables (such as length, complexity, etc.). This multi-step standardization procedure was completed as follows: first, to reduce the effect of statistical outliers, response latencies were Winsorized such that values of less than 0.5 seconds or greater than 40 seconds were set to 0.5 seconds and 40 seconds, respectively. Second, response times were standardized across items within each subject in order to control for differences between individual participants. Third, response times were standardized across subjects within each item in order to control for differences between items. Importantly, the mean and standard deviation used for this standardization were computed using only response time data from participants in the honest condition. Finally, items were re-Winsorized such that latencies of less than -3.00 or greater than 3.00 were set to -3.00 or 3.00, respectively (see Appendix D). This double-standardization procedure produces response latencies that are free of the confounding influences of individual persons, individual items, and statistical outliers.

Once items were standardized and Winsorized, mean response latencies were calculated for each of the five potential responses on the NEO-FFI scale (strongly disagree, disagree, neutral, agree, and strongly agree). Finally, a composite mean latency score was calculated by subtracting the mean response latency of the rejected items (responses of strongly disagree or disagree) from the mean response latency of the endorsed items (responses of agree or strongly agree). Items answered “neutral” were omitted. Of the 294 participants in Study 1, two were excluded due to missing values.

Next, a discriminant function analysis (assuming equal prior probabilities) was conducted for the Study 1 data, using the composite mean latency score as a predictor. This value correctly classified 51.4% of individuals as either honest, faking good, or faking bad respondents. The overall Wilks’ lambda was significant, $\lambda = .79$, $\chi^2(2, N = 292) = 69.26$, $p < .001$, indicating that the composite measure of response time detected group membership at a level significantly above chance.

Response time data conformed to Holden et al.’s (1992) congruence model of faking: for individuals faking good, unfavourable responses took longer to produce than favourable responses; conversely, for individuals faking bad, favourable responses took longer to produce than unfavourable responses. For individuals responding honestly, favourable responses took longer to produce than unfavourable responses, but the difference was not as dramatic as it was for individuals faking good (see Table 13 and Figure 7).

Table 13

Study 1: Group Means (SDs) by Faking Condition

Mean Adjusted Response Latency	Faking Good (n = 97)	Faking Bad (n = 97)	Honest (n = 98)	F (2, 289)
Endorsed Positive/Favourable Items	-0.10 (0.10)	0.16 (0.33)	-0.09 (0.12)	48.04**
Rejected Positive/Unfavourable Items	0.20 (0.49)	-0.10 (0.13)	0.03 (0.32)	17.98**
Composite Mean Latency Score	-0.30 (0.54)	0.26 (0.42)	-0.12 (0.38)	39.13**

** $p < .001$

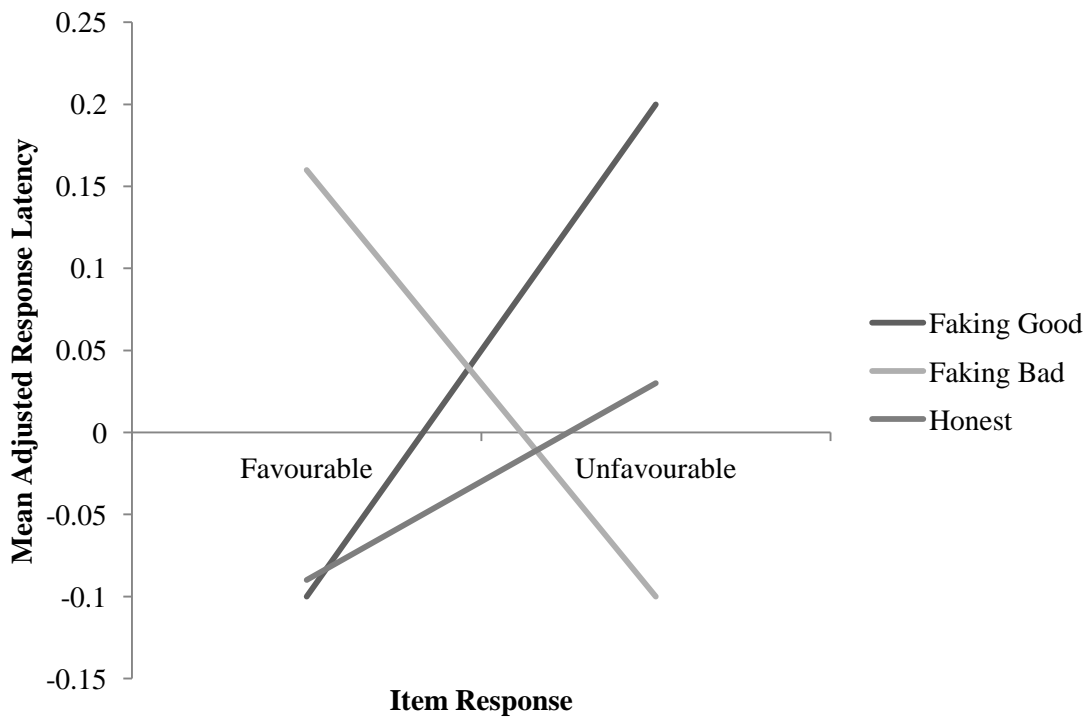


Figure 7. Mean adjusted item response latency by group and type of response.

Added Value of Response Times.

Next, I conducted stepwise discriminant function analyses on the Study 1 data in order to determine whether the composite measure of response latency could provide any added value above the traditional scales in predicting experimentally-induced faking condition.

The order of entry was the MCSDS, the JSDS, and the SDE subscale of the BIDR. Entry stopped after this – the IM subscale of the BIDR and the composite measure of response latency had no added value. The correct classification hit rate for these predictors was 83.3% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .20$, $\chi^2(6, N = 292) = 469.08$, $p < .001$, indicating that overall the predictors differentiated among the three faking conditions. For Study 1, all of the traditional lie detection scales, as well as the composite measure of response latency, differed significantly by faking condition. These results confirm the value of the traditional lie detection scales and the composite measure of response latency in detecting faking, but show that the composite measure of response time provides no additional value in detecting faking above the traditional scales.

Impression Management: An Industry Standard

The Impression Management (IM) subscale of the BIDR is presently the most widely used validity index in detecting participant faking. Currently, when researchers and clinicians are interested in determining whether an individual may be misrepresenting himself or herself on a personality inventory, they are most likely to do this by administering the IM subscale of the BIDR, along with the rest of their personality measures (Pauls & Crost, 2004; Davis, Thake, & Weekes, 2012). This is largely because, unlike other standard inventories like the MCSDS, the

BIDR uses Likert scale ratings (considered superior to the True/False scales of older measures), and provides cutoff scores for invalidity detection: participants with scores greater than 12 and 8 are probably and likely faking good, respectively, whereas participants with scores less than 1 and 2 are probably and likely faking bad, respectively (Paulhus, 1998). Because the IM subscale of the BIDR is most widely used, it is particularly interesting to examine the efficacy of alternative measures like response times and item order in detecting faking relative to this industry standard, because finding added value for the alternative measures would have meaningful practical implications. As discussed above, these alternative forms of measurement may be less susceptible to participant faking, and can be applied to the IM subscale items themselves so that practitioners could detect faking more effectively without having to administer extra inventories and extend testing time.

Impression Management vs. Item Order.

I conducted stepwise discriminant function analyses for the data from each study to examine whether the new operationalization of the impact of cognitive fatigue—the correlation between item order and response positivity—provided any added value in detecting participant faking relative to detecting participant faking using the most widely used industry standard, the IM subscale of the BIDR.

For Study 1, the order of entry was the IM subscale of the BIDR followed by the new measure of response pattern. The correct classification hit rate for these predictors was 71.0% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda =$

.39, $\chi^2(4, N = 293) = 273.79, p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

For Study 2, the IM subscale of the BIDR was the only variable entered into the analysis – the new measure of response pattern provided no added predictive power. The correct classification hit rate for these predictors was 58.7% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .55, \chi^2(2, N = 299) = 175.73, p < .001$, indicating that the IM subscale of the BIDR differentiated among the three faking conditions.

Impression Management vs. Response Times.

I conducted a stepwise discriminant function analysis for the data from Study 1 order to determine whether response times provided any added value in detecting participant faking relative to the widely used IM subscale of the BIDR. Data from Study 2 were not examined because response times were constrained. According to the congruence model of faking, response times should be shorter for responses that are congruent to a participant's faking condition (i.e., favourable responses while faking good) than for incongruent responses (i.e., unfavourable responses while faking good).

For the Study 1 data, the order of entry was the IM subscale of the BIDR followed by the composite measure of response latency. The correct classification hit rate for these predictors was 70.5% – significantly above the rate expected by chance. The overall Wilks' lambda was significant, $\lambda = .40, \chi^2(4, N = 292) = 262.00, p < .001$, indicating that overall the predictors differentiated among the three faking conditions.

Chapter 4

Summary of Findings

Traditional Measures

As expected, all of the established lie detection scales examined in this study – Paulhus’ (1998) Balanced Inventory of Desirable Responding scales of Self-Deceptive Enhancement and Impression Management, the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1960), Jackson’s (1984) Desirability Scale, and the Holden Psychological Screening Inventory (Holden, 1996) – were found to be significant predictors of instructed faking condition. When entered in a stepwise discriminant function analysis (DFA), all of the scales except the IM of the BIDR provided unique predictive power (See Table 14). This finding is particularly interesting given that the IM subscale of the BIDR is currently the most commonly used validity scale (Pauls & Crost, 2004; Davis, Thake, & Weekes, 2012). Though the IM detected faking at a rate above chance, it did so no better than other established scales. In fact, the MCSDS provided the most predictive power in Study 1, and the second most in Study 2. This finding was unexpected as the MCSDS is the oldest scale, published in 1960, and uses a True/False response scale which has been replaced in newer inventories by supposedly more sensitive Likert scales.

In Study 2, the HPSI provided the most power in predicting faking group membership, even above the MCSDS (see Table 14). The HPSI data were not available for the Study 1 data, we could not replicate this finding, but past studies have also found that the HPSI outperformed the MCSDS (Holden & Evoy, 2005; MacNeil & Holden, 2006; Holden & Book, 2009).

However, the position of the HPSI as strongest predictor in Study 2 suggests that it should also be considered as a superior alternative to the IM subscale of the BIDR in daily practice.

New Faking Response Strategy Scales

The Competitive Self-Promotion, Social Sensitivity, and Chameleonic Unpredictability subscales provided individual predictive value relative to the other subscales when entered in stepwise DFAs of the data from both Study 1 and Study 2. In contrast, the Showoff subscale had predictive value only for the Study 1 data, and the Strong-Mindedness subscale had value only for the data from Study 2 (see Table 15).

When entered into a stepwise DFA with the traditional measures of lie detection, only the Competitive Self-Promotion and Social Sensitivity subscales showed added predictive value for both studies. The Chameleonic Unpredictability subscale had added value in the first, but not in the second study. Importantly, the IM subscale of the BIDR provided no unique predictive value in the stepwise DFAs for either Study 1 or Study 2, which suggests that the new Competitive Self-Promotion and Social Sensitivity subscales are at least as effective in detecting instructed faking condition as this industry standard (see Table 16).

The MCSDS and the HPSI were again found to be the strongest predictors of instructed faking condition when entered into stepwise DFAs with all of the traditional lie scales as well as the new faking response strategy scales. When data from both inventories were present, the HPSI came out ahead. For both studies, the Social Sensitivity subscale was the next strongest predictor, demonstrating particular merit in faking-detection (see Table 16).

Table 14

Traditional Lie Scales: Individual & Stepwise DFAs

Scale	Study 1		Study 2	
	Significant Predictor? (<i>p</i> of Wilks' λ)	Entry in Stepwise DFA with Traditional Lie Detection Scales	Significant Predictor? (<i>p</i> of Wilks' λ)	Entry in Stepwise DFA with Traditional Lie Detection Scales
MCSDS	**	First	**	Second
JSDS	**	Second	**	Fourth
IM of BIDR	**	No Added Value	**	No Added Value
SDE of BIDR	**	Third	**	Third
HPSI	N.A.	N.A.	**	First

* $p < .05$, ** $p < .001$

Table 15

New Faking Response Strategy Scales: Individual DFAs

Scale		Study 1		Study 2	
		Significant Predictor? (<i>p</i> of Wilks' λ)	Entry in Stepwise DFA with Traditional Lie Detection Scales	Significant Predictor? (<i>p</i> of Wilks' λ)	Entry in Stepwise DFA with Traditional Lie Detection Scales
New Faking Response Strategy Scales	Competitive Self-Promotion	**	Third	**	Third
	Social Sensitivity	**	First	**	First
	Chameleonic Unpredictability	**	Second	**	Second
	Strong-Mindedness	**	No Added Value	n.s.	No Added Value
	Showoff	n.s.	No Added Value	**	Fourth

* $p < .05$, ** $p < .001$

Table 16

New Faking Response Strategy Scales: Stepwise DFAs

Scale		Study 1	Study 2
		Entry in Stepwise DFA with Traditional Lie Detection Scales	Entry in Stepwise DFA with Traditional Lie Detection Scales
MCSDS		1	2
JSDS		3	5
IM of BIDR		No Added Value	No Added Value
SDE of BIDR		4	4
HPSI		N.A.	1
New Faking Response Strategy Scales	Competitive Self-Promotion	5	6
	Social Sensitivity	2	3
	Chameleonic Unpredictability	6	No Added Value
	Strong-Mindedness	No Added Value	No Added Value
	Showoff	No Added Value	No Added Value

* $p < .05$, ** $p < .001$

Response Patterns/Item Order: Efficacy of the Cognitive Overload Model

This study introduced a new way to measure the effect of cognitive fatigue on participants' response patterns over time: the correlation between positivity of response to an individual item and its order. When entered as a sole predictor in a DFA assuming equal prior probabilities, this measure of item order significantly predicted instructed faking condition at a level significantly above chance (see Table 17). This finding supports the conclusion by Holden and Book (2009) that response patterns over time are a useful method of detecting participant faking, and demonstrates that their findings were not measure-specific.

Study 1 data showed the pattern of correlations predicted based on the cognitive overload model of faking: participants in the fake good condition showed a positive correlation, and participants in the fake bad condition showed a negative correlation, although it was slight. Similar to participants in the fake good condition, participants in the honest condition for both studies showed positive correlations between item order and response positivity. All effects were reduced for the Study 2 data, when participants were required to respond to the analyzed NEO-FFI items under strict time constraints. These reduced effects are opposite to the predicted effects of time constraints: we predicted that with decreased time, participants would be less able to fake effectively and thus would show more positive correlations for the Study 2 fake good condition and more negative correlations for the Study 2 fake bad condition. This unexpected finding is problematic for the cognitive overload model, because it suggests that increasing cognitive demands does not result in decreased ability to fake.

When it was included in stepwise DFAs along with traditional lie detection scales, this new measure of response pattern provided added predictive value only for the Study 1 data, where it had less individual predictive power than the MCSDS, the JSDS, and the SDE subscale of the BIDR. Still, response pattern had more predictive power than the IM subscale of the BIDR (see Table 18).

Time Limits.

Despite being required to both read and respond to each NEO-FFI item in 2 seconds, participants in Study 2 were still able to fake, as indicated by the presence of a significant main effect of faking condition for each of the Big Five personality dimensions.

Response Latencies: Efficacy of the Congruence Model

Response time data were examined only for Study 1, where response times were not constrained, and only for the NEO-FFI. A composite measure of response latency was a significant predictor of instructed faking condition when entered as a sole predictor in a DFA (see Table 17), correctly identifying 51.4% of participants. However, when entered in a stepwise DFA along with established gold-standard lie detection scales, response latencies contributed no additional predictive value (see Table 19).

Table 17

Item Order & Response Latencies: Individual DFAs

Scale	Study 1		Study 2	
	Significant Predictor? (<i>p</i> of Wilks' λ)	% Correctly Classified	Significant Predictor? (<i>p</i> of Wilks' λ)	
Correlation Between Positivity of Response & Item Order	**	49.5%	**	38.1%
Composite Response Latency	**	51.4%	N.A.	N.A.

* $p < .05$, ** $p < .001$

Table 18

Item Order: Stepwise DFAs

Scale	Study 1	Study 2
	Entry in Stepwise DFA with Traditional Lie Detection Scales	Entry in Stepwise DFA with Traditional Lie Detection Scales
MCSDS	1	2
JSDS	2	4
IM of BIDR	No Added Value	No Added Value
SDE of BIDR	3	3
HPSI	N.A.	1
Correlation Between Positivity of Response & Item Order	4	No Added Value

* $p < .05$, ** $p < .001$

Table 19

Response Latencies: Stepwise DFAs

Scale	Study 1
	Entry in Stepwise DFA with Traditional Lie Detection Scales
MCSDS	1
JSDS	2
IM of BIDR	No Added Value
SDE of BIDR	3
HPSI	N.A.
Composite Response Latency	No Added Value

* $p < .05$, ** $p < .001$

Chapter 5

Discussion

All of the self-report scales examined, both the traditional and the new faking response scales, showed some validity in detecting instructed faking condition, which supports their ongoing use. Overall, the results of this study point to the weakness of the Impression Management subscale of the BIDR in detecting instructed faking, despite the fact that it is currently considered the gold-standard in the faking-detection industry. The IM provided no unique predictive power in any of the stepwise DFAs conducted, whether it was compared to the other traditional lie detection scales, to the new faking response strategy scales, to the new measure of response pattern, or the composite measure of response latency. The added value provided by the two new non-behavioural measures examined in this study—response pattern and response latencies—strongly supports their future use alongside traditional lie detection scales.

The Marlowe-Crowne Social Desirability Scale (1960), despite commonly being considered outdated due to its age and its True/False response scale, was found to be one of the most powerful predictors of faking. Notably, the MCSDS was a more powerful detector of faking than the Impression Management and Self-Deceptive Enhancement subscales of the BIDR (1998), despite the BIDR's recent elevation to a gold-standard position in the discipline. It was also more powerful than all of the new faking response strategy scales, our new measure of response pattern, and the composite measure of response latency. The only measure that managed to outperform the MCSDS was the HPSI. This finding suggests that the MCSDS

should be reconsidered as a useful scale for everyday practice, because it captures elements of faking more effectively than more modern scales.

The Holden Psychological Screening Inventory, despite being less well-known as compared to the IM of the BIDR, was found to be the only scale that rivaled the MCSDS in its power for detecting faking. In fact, in all cases when both the MCSDS and the HPSI were entered in a stepwise discriminant function analysis, the HPSI came out as the strongest predictor. This finding suggests that the HPSI should be seriously considered for daily use in faking detection.

Of the five new faking response strategy scales (Holden, Passey, Jay, & Book, 2009), the Social Sensitivity and Competitive Self-Promotion subscales showed the most promise in detecting faking. The Social Sensitivity subscale was particularly promising, because it was consistently found to be more powerful than the JSDES, IM subscale of the BIDR, and SDE subscale of the BIDR. In contrast, the Showoff subscale showed little merit, rarely providing added value, and should likely be abandoned.

Dimensionality of Faking

In almost all of the discriminant function analyses reviewed above, multiple scales—whether traditional or new—were found to be significant predictors of faking condition. Because each scale is designed to measure a different aspect of faking, the significance of multiple scales supports past findings that faking is multidimensional, rather than being a unidimensional honesty-centred continuum ranging from negative to positive, or even bidimensional, as previously found (Paulhus, 1984; Holden & Evoy, 2005; Kitching & Paulhus, 2009).

Support for the Congruence Model

Response latency data supported the congruence model of faking (Holden, Kroner, Fekken & Popham, 1992). Discriminant function analyses using a composite measure of response time showed that participants who adopted a faking schema took relatively longer to produce schema-inconsistent responses; individuals faking good took longer to produce unfavourable responses, whereas individuals faking bad took longer to produce favourable responses. This finding is in direct contrast to that of van Hooft and Born (2012)—who found no faking condition by item desirability interaction—and supports our hypothesis that van Hooft and Born failed to replicate congruence model findings because they aggregated item response times based on whether scale *items* were negatively or positively keyed, rather than whether a participant gave a favourable or unfavourable *answer*. More thorough confirmation of this theory would require access to the original data from the study, so that they could be reanalysed after proper aggregation, but van Hooft and Born have stopped responding to e-mails.

Challenging the Cognitive Overload Model

Our new measure of response pattern, based in the cognitive overload model of faking, was a successful predictor of faking condition, but had added value only for the Study 1 data. According to the cognitive overload model of faking, participants in Study 2, who responded to the NEO-FFI under time restrictions, should have experienced more cognitive fatigue and thus greater changes in response patterns over time than participants in Study 1. In actuality, response patterns differed less for participants in Study 2 than for those in Study 1. This result conflicts with the cognitive overload model of faking. Response patterns were still found to be a useful

indicator of faking condition, but this inconsistent result suggests that cognitive overload and fatigue may not be the causal mechanism behind changing response patterns.

This inconsistent finding may support past theories that have suggest that faking is not as cognitively complex and demanding as we might assume (van Hooft & Born, 2012). It was originally proposed that faking is a cognitively complex task, resulting in higher cognitive load and longer response latencies (Zuckerman, DePaulo, & Rosenthal, 1981; Vrij, Edward, & Bull, 2001). This theory presented faking as an editing process: individuals were believed to first think of the correct answer to each question, and then to edit the response according to their adopted faking schema. An alternative theory was later proposed (Hsu, Santelli, Hsu, 1989; Holtgraves, 2004) that states that faking may be a relatively automatic, primitive skill. According to this theory, faking should take less time than responding honestly because faking requires purely semantic item interpretation, whereas honest responding requires self-referenced interpretation. From a young age, we are all taught to recognize right and wrong, and are rewarded for good behaviour and punished for bad. Thus, when required to give a favourable or unfavourable response, individuals need only consult these over-learned semantic concepts, which is much less cognitively complex than personal introspection.

Another possibility, which we tend to favour, is that participants in Study 2 began to respond to the NEO-FFI items randomly. When participants failed to respond in time, a message stating this fact appeared on the screen and participants were forced to wait before the next item was presented. Participants may have begun to respond rapidly and randomly, ignoring task demands and instead focusing on avoiding this aversive experience. Consistent with this

hypothesis, the Study 2 data show decreased differentiation in the results between conditions. The Study 1 correlations have a relatively wide spread, ranging between -0.04 for fake bad, 0.11 for honest responding, and 0.13 for fake good. In Study 2 these values are reduced to 0.03 for fake bad, 0.07 for honest responding, and 0.09 for fake good. This theory is further supported by participants' significantly reduced compliance scores in Study 2: participants may have been unaware that they had adopted a strategy of random response due to impossibly strict time constraints. Because response time data supported the congruence model of faking, it may be that this interpretation is most likely to be the correct explanation for the lack of increased efficacy of cognitive overload in detecting faking in Study 2.

Limitations

As with all experimental studies, the current study has several limitations. First, we examined instructed faking, rather than naturally-occurring faking, which many researchers believe to be a qualitatively different phenomenon because instructed faking may not replicate the high-stakes testing circumstances that are common to real-world faking scenarios. However, we attempted to overcome this limitation by giving participants a performance incentive: participants were instructed that, for every 25 participants, the individual who best followed instructions and avoided activating faking scales would receive a \$50 prize. Still, generalizability to real-world faking may be limited.

Another important limitation is the fact that the order in which items were presented is not randomized in the NEO-FFI. This means that there may be inherent effects of item order on our measure of response pattern, which would confound the effects due to cognitive fatigue

which we intended to study. A follow-up study with full item randomization would be required to rule out an inherent effect of item order; this was not done in the current study because in regular practice individuals always complete NEO-FFI items in the same order.

Review of Objectives

1. Evaluate the efficacy of the congruence model of faking (Holden, Kroner, Fekken & Popham, 1992) in detecting participant faking by using response time data.

The congruence model of faking, as measured by a composite measure of response latency, was found to be an efficacious predictor of instructed faking condition.

2. Extend the findings of Holden and Book (2009) by developing a new way to operationally measure the impact of cognitive overload on faking and to use this measure to examine the efficacy of the cognitive overload model in detecting participant faking.

The new measure of response pattern was a significant predictor of faking condition both in Study 1 and in Study 2. However, it did not have added value in Study 2.

3. Further evaluate the cognitive overload model of faking by requiring participants to respond in a shortened period of time (Study 2), inducing excessive cognitive overload that may disrupt the process of faking

The results of these studies have mixed implications for the cognitive overload model of faking.

The new measure of response pattern was a useful predictor for both the Study 1 and the Study 2 data, but had no added value for the Study 2 data. This finding is problematic for the cognitive overload model of faking because participants in Study 2 experienced time restrictions during

responding, and thus would be expected to experience more cognitive overload and increasingly extreme response patterns over time.

4. Examine the efficacy of the new faking response strategy scales developed by Holden, Passey, Jay, and Book (2009) in detecting participant faking.

Efficacious, except for the Showoff and Strong-Mindedness scales.

5. Evaluate the added value that the cognitive overload model, the congruence model, and the new faking response strategy scales can provide in detecting participant faking relative to established gold-standard detection measures.

Measures of response latencies and response patterns were found to provide added value in detecting faking, although the cognitive overload model is likely not the mechanism underlying changes in response patterns.

References

- Book, A. S., Holden, R. R., Starzyk, K. B., Wasylkiw, L., & Edwards, M. J. (2006). Psychopathic traits and experimentally induced deception in self-report assessment. *Personality and Individual Differences, 41*, 601-608.
- Butcher, J. N., Morfitt, R. C., Rouse, S. V., & Holden, R. R. (1997). Reducing MMPI-2 defensiveness: The effect of specialized instructions on retest validity in a job applicant sample. *Journal of Personality Assessment, 68*, 385-401.
- Costa, P. T. Jr., & McCrae, R. R. (1992). *NEO PI-R professional manual: Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI)*. Odessa, FL: Psychological Assessment Resources.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology, 24*, 349-354.
- Davis, C. G., Thake, J., Weekes, J. R. (2012). Impression managers: Nice guys or serious criminals? *Journal of Research in Personality, 46*, 26-31.
- Eid, M., & Zickar, M. J. (2007). Detecting response styles and faking in personality and organizational assessments by mixed Rasch models. In M. von Davier & C. H. Carstensen (Eds.), *Multivariate and mixture distribution Rasch models* (pp. 255-270). New York: Springer.
- Heine, S. J., Lehman, D. R., Markus, H. R., & Kitayama, S. (1999). Is there a universal need for positive self-regard? *Psychological Review, 106*, 766-794.

- Holden, R. R. (1996). *Holden Psychological Screening Inventory (HPSI)*. North Tonawanda, NY: Multi-Health Systems.
- Holden, R. R. (1998). Detecting fakers on a personnel test: Response latencies versus a standard validity scale. *Journal of Social Behavior and Personality, 13*, 387-398.
- Holden, R. R. (2007). Socially desirable responding does moderate scale validity both in experimental and in non-experimental contexts. *Canadian Journal of Behavioural Science, 39*, 184-201.
- Holden, R. R., & Book, A. S. (2009). Using hybrid Rasch-latent class modeling to improve the detection of fakers on a personality inventory. *Personality and Individual Differences, 47*, 185-190.
- Holden, R. R., & Book, A. S. (2012). Faking does distort self-report personality assessment. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 71-84). New York: Oxford University Press.
- Holden, R. R., Book, A. S., Edwards, M. J., Wasylkiw, L., & Starzyk, K. B. (2003). Experimental faking in self-reported psychopathology: Unidimensional or multidimensional? *Personality and Individual Differences, 23*, 1107-1117.
- Holden, R. R., & Evoy, R. A. (2005). Personality inventory faking: A four-dimensional simulation of dissimulation. *Personality and Individual Differences, 39*, 1307-1318.
- Holden, R. R., & Hibbs, N. (1995). Incremental validity of response latencies for detecting fakers on a personality test. *Journal of Research in Personality, 29*, 362-372.

- Holden, R. R., & Kroner, D. G. (1992). Relative efficacy of differential response latencies for detecting faking on a self-report measure of psychopathology. *Psychological Assessment*, 4, 170-173.
- Holden, R. R., & Troister, T. (2009). Developments in the self-report assessment of personality and psychopathology in adults. *Canadian Psychology*, 50, 120-130.
- Holden, R. R., Fekken, G. C., & Cotton, D. H. G. (1991). Assessing psychopathology using structured test-item response latencies. *Psychological Assessment: A journal of Consulting and Clinical Psychology*, 3, 111-118.
- Holden, R. R., Jay, A., & Book, A. S. (2009, August). *The dimensionality of response strategies for fakers on a personality inventory*. Poster presented at the American Psychological Association Annual Convention, Toronto, Ontario, Canada.
- Holden, R. R., Wood, L. L., & Tomashewski, L. (2001). Do response time limits counteract the effect of faking on personality inventory validity? *Journal of Personality and Social Psychology*, 81, 160-169.
- Holden, R. R., Kroner, D. G., Fekken, G. C., & Popham, S. M. (1992). A model of personality test item response dissimulation. *Journal of Personality and Social Psychology*, 63, 272-279.
- Holden, R. R., Passey, P., Jay, A. & Book, A. S. (2009). The dimensionality of response tactics for faking on a self-report personality measure. *Unpublished Manuscript*.
- Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30, 161-172.

- Hsu, L. M., Santelli, J., & Hsu, J. R. (1989). Faking detection validity and incremental validity of response latencies to MMPI subtle and obvious items. *Journal of Personality Assessment*, *53*, 278-295.
- Jackson, D. N. (1984). *Personality Research Form manual* (3rd ed.). Port Huron, MI: Sigma Assessment Systems.
- Kitching, S., & Paulhus, D. L. (2009, August). *The bi-dimensional impression management index (BIMI): A measure of agentic and communal impression management*. Poster presented at the American Psychological Association Annual Convention, Toronto, Ontario, Canada.
- Kroger, R. O., & Turnbull, W. (1975). Invalidity of validity scales: The case of the MMPI. *Journal of Consulting and Clinical Psychology*, *43*, 48-55.
- MacNeil, B. M., & Holden, R. R. (2006). Psychopathy and the detection of faking on self-report inventories of personality. *Personality and Individual Differences*, *41*, 641-651.
- Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analysis review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, *130*, 711-747.
- Paulhus, D. L. (1998). *Paulhus Deception Scales (PDS) user's manual*. North Tonawanda, NY: Multi-Health Systems.
- Pauls, C. A., & Crost, N. W. (2004). Effects of faking on self-deception and impression management scales. *Personality and Individual Differences*, *37*, 1137-1151.

- Rosse, J. G., Stetcher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology, 83*, 634-644.
- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology, 30*, 526-537.
- Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*, 197-210.
- van Hooft, E. A. J. & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology, 97*, 301-316.
- Vrij, A., Edward, K., & Bull, R. (2001). Stereotypical verbal and nonverbal responses while deceiving others. *Personality and Social Psychology Bulletin, 27*, 899-909.
- Zickar, M. J., Gibby, R. E., & Robie, C. (2004). Uncovering faking samples in applicant, incumbent, and experimental data sets: An application of mixed-model item response theory. *Organizational Research Methods, 7*, 168-190
- Zickar, M. J., & Robie, C. (1999). Modeling faking good on personality items: An item-level analysis. *Journal of Applied Psychology, 84*, 551-563.
- Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz. (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1-59). New York: NY: Academic Press.

Appendix A

Military Induction Instructions

Condition 1

Important instructions for completing this study

Please complete the remainder of the questionnaires by imagining the following scenario:

- You have been selected for possible induction into the military. This is a job you want.
- You should respond as honestly as possible.
- Do not try to make yourself look better or worse than you are as there are validity measures embedded within the questions to identify any faking on your part, so try your best to report honestly.

Click the mouse to continue...

Condition 2

Important instructions for completing this study

Please complete the remainder of the questionnaires by imagining the following scenario:

- You have been selected for possible induction into the military. This is a job you want very much. In order to maximize your chances of being inducted you must respond in a manner that exaggerates your positive attributes.
- In other words, you should fake good.

- Be aware that there are validity measures embedded within the questions to identify faking, so try your best to look good without getting caught.

Click the mouse to continue...

Condition 3

Important instructions for completing this study

Please complete the remainder of the questionnaires by imagining the following scenario:

- You have been selected for possible induction into the military. This is a job you do not want. In order to minimize your chances of being inducted you must respond in a manner that exaggerates your negative attributes.
- In other words, you should fake bad.
- Be aware that there are validity measures embedded within the questions to identify faking, so try your best to look bad without getting caught.

Click the mouse to continue...

Appendix B

Faking Strategy Scales Developed by Holden, Passey, Jay, & Book (2009)

Instructions: Please read the statements carefully and indicate whether they are True (T) or False (F).

Competitive Self-Promotion (15 items; coefficient alpha = .90)

- Losing is for others, not me.
- I can manipulate other people's opinion of me when I want to.
- I like to win.
- I always look to maximize my outcomes.
- I don't mind losing. (R)
- I tend to dwell on my bad characteristics. (R)
- I never know what others expect from me. (R)
- I can usually figure out what other people are thinking about.
- I can look anyone in the eyes and tell a lie with a straight face (if for a right end).
- I am very good at judging the thoughts of others.
- When I can to, I can make other people admire my characteristics.
- I think about my shortcomings all the time. (R)
- I find it hard to imitate the behaviour of other people. (R)
- I have trouble changing my behaviour to suit different people and different situations. (R)
- I am not particularly good at making other people like me. (R)

Social Sensitivity (11 items; coefficient alpha = .89)

- The opinions of other people matter to me.
- I care what other people think about me.
- I am an open and honest person.
- I try to live a virtuous life.
- I am a good person.
- I am a bad person. (R)
- I try to figure out what other people want from me.
- At parties and social gatherings, I do not attempt to do or say things that others will like.
(R)
- I don't care about what others want from me. (R)
- I am careful in the way I present myself.
- I want to fit in with everyone else.

Chameleonic Unpredictability (6 items; coefficient alpha = .75)

- I'm not always the person I appear to be.
- In different situations and with different people, I often act like very different persons.
- When I want to, I can make other people dislike my characteristics.
- I can be stubborn and contrary.
- I tend to behave in a random sort of way.

- My behaviour is unpredictable.

Strong-Mindedness (4 items; coefficient alpha = .61)

- I am not afraid to state my opinion.
- I would not change my opinion (or the way I do things) in order to please someone else or win their favour.
- In order to get along and be liked, I tend to be what people expect me to be rather than anything else. (R)
- I am not afraid to speak up.

Showing Off (5 items; coefficient alpha = .76)

- I would like to be a performer of some sort.
- At one time, I considered pursuing a career in the performing arts.
- I enjoy acting.
- I like to stand out in a crowd.
- I do not enjoy being the centre of attention. (R)

Note: (R) indicates a reverse-keyed item

Appendix C

Letter of Information, Informed Consent, & Debriefing Form

LETTER OF INFORMATION

This research is being conducted by Christine Lambert, who is a graduate student working on her MSc with Dr. Ronald Holden, Professor, of the Department of Psychology at Queen's University in Kingston, Ontario.

What is this study about?

The purpose of this research is to explore how university students present themselves on personality measures. We ask participants to read a brief scenario and complete a series of questionnaires concerning personality. We estimate that it takes about one hour to complete these tasks and that there are no known physical, psychological, economic, or social risks associated with them.

Is my participation voluntary?

Yes. Although it would be greatly appreciated if you would answer all material as frankly as possible, you should not feel obliged to answer any material that you find objectionable or that makes you feel uncomfortable. You may also withdraw at any time with no penalties at all.

What will happen to my responses?

We will keep your responses confidential. We will store the data in a locked room until the raw data is no longer needed. Only authorized researchers will have access to this area. To help us ensure confidentiality, please do not put your name on any of the research study answer sheets. The data being collected today may be reanalyzed in the future to answer related questions about this research area. Confidentiality will be maintained at all times. The data may also be presented in

professional psychological journals or at scientific conferences, but any such presentations will be of general findings and will never breach individual confidentiality. Should you be interested, you are entitled to a copy of the findings.

Will I be compensated for my participation?

Yes. In exchange for your help, we will grant you 1 credit towards your PSYC 100 grade. If you are not enrolled in PSYC 100, or have already received your maximum credits, we will compensate you with \$15.00 cash.

Any questions about study participation may be directed to Dr. Ronald Holden at holdenr@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at 613-533-6081 or Chair.GREB@queensu.ca

This study has been granted clearance according to the recommended principles of Canadian ethics guidelines, and Queen's policies.

Again, thank you. Your interest in participating in this research study is greatly appreciated.

Dr. Ronald Holden
Professor

Christine Lambert
MSc Student

CONSENT FORM

Name (please print clearly): _____

1. I have read the Letter of Information and have had any questions answered to my satisfaction.
2. I understand that I will be participating in the study called "Cake". I understand that this means that I will be asked to read a brief scenario and complete a series of questionnaires concerning personality so that the presentation of university students on personality measures may be better understood.
3. I understand that this session will take 1 hour and that I will receive 1 credit.
4. I understand that my participation in this study is voluntary and I may withdraw at any time without penalty.
5. I understand that every effort will be made to maintain the confidentiality of the data now and in the future. The data will be kept in a locked room that is accessible only by the researchers involved in the study. When assessing and compiling the data, all participants names will be converted to numbers and no individual data will be reported. The data collected for this study will not be used for any purpose outside of scientific research.
6. I understand that the data being collected today may be reanalyzed in the future to answer related questions about this research area. Confidentiality will be maintained at all times.

Any questions about study participation may be directed to Dr. Ronald Holden at holdenr@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at 613-533-6081 or Chair.GREB@queensu.ca

I have read the above statements and freely consent to participate in this research:

Signature: _____ Date: _____

DEBRIEFING FORM

At the beginning of the study, we told you that the purpose of the study was to explore how university students present themselves on personality measures. We only gave you a vague description of the study beforehand because we did not want to influence your responses in any way. In reality, we are interested in determining the value of several different models and measures in detecting faking.

The use of personality tests throughout Canadian society is based on the assumption that their results are valid. However, research has shown that individuals can, and do, fake their responses on personality inventories. Individuals may fake good, emphasizing their positive characteristics, or fake bad, emphasizing negative characteristics, in order to obtain a desired outcome. Recent research has provided support for a congruence model of faking, which states that schema-consistent responses are provided more quickly than schema-inconsistent responses. Faking successfully, without being detected by validity indices, requires balancing favourable and unfavourable responses, regardless of the faking schema a participant adopts. This demand results in cognitive fatigue over time, producing increasingly unbalanced response patterns. The purpose of the current study is to determine the value of the congruence and cognitive overload models, standard inventories, and recently developed Faking Response Strategy Scales in detecting fakers.

In our study, participants were randomly assigned to one of three groups. Each group read a short paragraph asking them to imagine that they were being recruited for the military. Participants were asked to complete a personality assessment and to either respond honestly, fake good to be selected, or fake bad to avoid military induction. As stated earlier, all responses to questionnaire items will be kept confidential.

In the event that you would like to see a counsellor about this study, please contact Health, Counselling and Disabilities Services at 533-2506. They are located at 146 Stuart in the St. LaSalle Bldg, (across the street from Adelaide Hall). Also, please feel free to talk to the experimenter for more information on the research you just participated in.

If you would like further information on this area of research, these are some related references that might be of interest to you:

1. Ziegler, M., MacCann, C., & Roberts, R. D. (2012). *New perspectives on faking in personality assessment*. New York: Oxford University Press.
2. van Hooft, E. & Born, M. P. (2012). Intentional response distortion on personality tests: Using eye-tracking to understand response processes when faking. *Journal of Applied Psychology*, 97(2), 301-316.

Any questions about study participation may be directed to Dr. Ronald Holden at holdenr@queensu.ca. Any ethical concerns about the study may be directed to the Chair of the General Research Ethics Board at 613-533-6081 or Chair.GREB@queensu.ca

Please do not tell other potential subjects about the purpose of this study. Again, thank you. Your interest in participating in this research study is greatly appreciated.

Dr. Ronald Holden
Professor

Christine Lambert
MSc Student

Appendix D

An Example of the Preparation and Computer of Response Latencies

Note: Data would normally comprise many more items and respondents.

Personality Items:

1. I like to be the first to apologize after an argument.
2. I get a kick out of seeing someone I dislike appear foolish in front of others.
3. If public opinion is against me, I usually decide that I am wrong.
4. I get annoyed with people who never want to go anywhere different.
5. I live from day to day without trying to fit my activities into a pattern.

Raw response time latencies (in seconds) for five respondents:

Respondent	Item					Mean	SD
	1	2	3	4	5		
1	4.12	19.82	4.50	6.42	5.06	7.98	6.67
2	9.61	10.43	6.81	6.59	5.83	7.85	2.03
3	3.95	10.66	4.51	14.23	3.40	7.35	4.84
4	4.45	9.45	0.30	3.79	4.28	4.45	3.27
5	10.77	76.31	12.14	6.29	4.50	22.00	30.52
Mean	6.58	25.33	5.65	7.46	4.61		
SD	3.33	28.80	4.32	3.95	0.90		

Step 1. Reset maximum latencies to 40 seconds and minimum latencies to 0.5 seconds (values outside this range are regarded as outliers that will unduly influence analyses).

Respondent	Item					Mean	SD
	1	2	3	4	5		
1	4.12	19.82	4.50	6.42	5.06	7.98	6.67
2	9.61	10.43	6.81	6.59	5.83	7.85	2.03
3	3.95	10.66	4.51	14.23	3.40	7.35	4.84
4	4.45	9.45	0.50	3.79	4.28	4.49	3.20
5	10.77	40.00	12.14	6.29	4.50	14.74	14.47
Mean	6.58	18.07	5.69	7.46	4.61		
SD	3.33	12.96	4.26	3.95	0.90		

Step 2. Standardize within a respondent to adjust for irrelevant person factors such as reading speed, verbal ability, motor speed, etc.

Respondent	Item					Mean	SD
	1	2	3	4	5		
1	-0.58	1.77	-0.52	-0.23	-0.44	0.00	1.00
2	0.86	1.27	-0.51	-0.62	-1.00	0.00	1.00
3	-0.70	0.68	-0.59	1.42	-0.82	0.00	1.00
4	-0.01	1.55	-1.25	-0.22	-0.07	0.00	1.00
5	-0.27	1.75	-0.18	-0.58	-0.71	0.00	1.00
Mean	-0.14	1.40	-0.61	-0.05	-0.61		
SD	0.62	0.45	0.39	0.84	0.36		

Step 3. Standardize within an item to correct for irrelevant item factors such as item length, complexity, order, etc. [Note: For experimental groups, standardizing within an item should use item means and standard deviations associated with a control or normative group]. Results are standardized times that represent latencies relative to the respondent and relative to the item.

Respondent	Item					Mean	SD
	1	2	3	4	5		
1	-0.71	0.82	0.23	-0.22	0.46	0.12	0.60
2	1.61	-0.30	0.25	-0.68	-1.08	-0.04	1.05
3	-0.90	-1.60	0.06	1.74	-0.58	-0.26	1.27
4	0.20	0.32	-1.63	-0.20	1.48	0.03	1.12
5	-0.21	0.76	1.10	-0.64	-0.28	0.15	0.75
Mean	0.00	0.00	0.00	0.00	0.00		
SD	1.00	1.00	1.00	1.00	1.00		

Step 4. These are standardized times. Reset maximum latencies to 3.00 and minimum latencies to -3.00 (values outside this range are regarded as outliers that will unduly influence analyses). [Not necessary for this example].

Step 5. Aggregate data by computing mean latencies within a respondent. Usually done separately for endorsements and rejections of a specific trait (or response style) and done separately for answering true and for answering false to true/false items.