Scoring Protocols and Ignoring Pertinent Cues in Personality-Based Structured Interviews

by

Paul-Andre S. Comeault

A Thesis presented to The University of Guelph

In partial fulfilment of requirements for the degree of Master of Arts in Psychology

Guelph, Ontario, Canada

© Paul-Andre S. Comeault, September, 2015
ABSTRACT

Scoring Protocols and Ignoring Pertinent Cues in Personality-Based Structured Interviews

Paul-Andre S. Comeault        Advisors:
University of Guelph, 2015     Professor Deborah Powell
                                Professor David Stanley

A common practice in personality-based interviews is to assess each trait using only cues solicited by a specific question, and unsolicited cues are to be ignored. Using a person perception theory framework, 247 USA microworkers (52% women, mean age 34 years) participated in an online experiment to test the effects of unsolicited cues on personality trait ratings. Interviewers presented with positive solicited cues could ignore negative unsolicited cues when instructed to do so ($d = 1.15$, 95% CI [0.45, 1.80]). The manipulation of positive unsolicited cues was ineffective and may be linked to a lack of interviewer experience or negative cues that were disproportionately stronger than positive cues. Unsolicited cues intended to be neutral seemed to impair interviewer self-efficacy. Although results suggest that interviewers can ignore unsolicited cues in some circumstances, results are interpreted with caution in light of the study’s limitations.

Keywords: Structured Employment Interview Intentional Forgetting Disregard Forget-Cued Personality Trait Assessment Person Perception Five Factor Model Impression Formation Pragmatics Question-Based Scoring Dimensional Scoring
Acknowledgements

The success of this manuscript would not have been possible without the unwavering support of my advisors, Dr. Deborah M. Powell, Dr. David J. Stanley, Dr. Peter A. Hausdorff. I would also like to thank my dear friends and colleagues Sharon Hou, Scott Cassidy and Patricia L. Baratta.

Correspondence concerning this manuscript should be addressed to Paul-Andre S. Comeault, 413-19 Four Winds Drive, Toronto, Ontario, M3J 2S9.
Email: pacomeault@gmail.com
# TABLE OF CONTENTS

TABLE OF CONTENTS.......................................................................................iv  
LIST OF TABLES..............................................................................................vi  
LIST OF FIGURES...........................................................................................vii  
INTRODUCTION.................................................................................................1  
  Interview Structure.......................................................................................2  
  Scoring Protocols and the Predictive Validity of Interviews......................3  
  Question-Based Scoring Prevailed Over Dimensional Scoring....................3  
  The Flawed Assumption of Question-Based Scoring..................................5  
  Inadmissible Information Affects Decisions in Jury Setting......................5  
  Inadmissible Information Affects Ratings of Employability......................6  
  Inadmissible Information Affects Decisions in Assessment Centers..........6  
  Inadmissible Information Affects Decisions in Interviews......................7  
  Inadmissible Information Affects Perceptions of Personality....................7  
  Scoring Personality-Based Interviews.......................................................8  
  Can Interviewers Ignore Unsolicited Cues?.............................................9  
  Two Situations with Contradictory Cues..................................................10  
  Hypotheses..................................................................................................11  
METHOD..........................................................................................................15  
  Participants..................................................................................................15  
  Apparatus and Materials...........................................................................16  
  Procedure....................................................................................................18  
  Independent Variables...............................................................................19  
  Dependent Variable...................................................................................21  
RESULTS........................................................................................................21  
  General........................................................................................................21  
  Test of Hypothesis 1A................................................................................21  
  Test of Hypothesis 2A................................................................................22  
  Test of Hypothesis 3A................................................................................22  
  Test of Hypothesis 1B................................................................................23  
  Test of Hypothesis 2B................................................................................23
Test of Hypothesis 3B ........................................................................................................24
Exploratory Analyses ........................................................................................................24
DISCUSSION ..........................................................................................................................25
Ignoring Contradictory Negative Information .................................................................26
Ignoring Contradictory Positive Information .......................................................................29
Unexpected Effects of “Neutral” Information ....................................................................31
Practical Implications ..........................................................................................................33
Limitations .............................................................................................................................34
Conclusion ..............................................................................................................................36
References ..............................................................................................................................38
Appendix A: BARS For Achievement Striving .................................................................49
Appendix B: BARS for Stress Tolerance .............................................................................50
Appendix C: Threshold Test Item 1 – High Achievement Striving ....................................51
Appendix D: Threshold Test Item 2 – Low Achievement Striving .......................................52
Appendix E: Threshold Test Item 3 – High Stress Tolerance ..............................................53
Appendix F: Threshold Test Item 4 – Low Stress Tolerance ...............................................54
Appendix G: Summary of Manipulation Checks in Post-Experimental Questionnaire ......55
Appendix H: Post-Experimental Questionnaire ....................................................................59
Appendix I: Initial Description of the Interview Like a Pro CrowdFlower Task ...............68
Appendix J: Initial Letter of Letter of Information and Consent with Deception ..............69
Appendix K: Debrief and Second Letter of Information and Consent ..............................72
Appendix L: Dimensional and Question-based Scoring Instructions .................................75
Appendix M: Q1 Unsolicited Stress Tolerance Cues .........................................................76
Appendix N: Q2 Solicited Stress Tolerance Cues ...............................................................77
LIST OF TABLES

Table 1. Study Design: Scoring protocol and Presentation of Stress Tolerance Cues ............42
Table 2. Results: Scoring protocol and Presentation of Stress Tolerance Cues ................43
LIST OF FIGURES

Figure 1. Test of H1A and H2A.................................................................44
Figure 2. Test of H3A.................................................................45
Figure 3. Test of H1B and H2B.................................................................46
Figure 4. Test of H3B.................................................................47
Figure 5. Effect of Neutral Unsolicited Cues Across Scoring Protocol Conditions..............48
Scoring Protocols and Ignoring Cues in Personality-Based Structured Interviews

Employment interviews are widely used in employee selection to assess a number of job-related constructs, including competencies, skills, and personality traits. Research in Industrial/Organizational psychology shows compelling evidence that structured interviews predict job performance better than do traditional or unstructured interviews. Interview structure improves the predictive validity of interviews by including protocols on how to ask questions and score applicants’ responses (i.e., asking specific questions, using a scoring guide). A prevalent practice, known as question-based scoring, involves assessing separate constructs with separate questions and evaluating responses to each interview question separately. This practice emerged partly from the methods (i.e., job analysis and focus groups) used to develop structured interview questions and partly from the influence of key studies discussed below. When job applicants provide cues in response to one question that pertain to a construct to be assessed with a different question, question-based scoring dictates these unsolicited cues are to be ignored. Despite this instruction, interviewers may be unable to ignore this information. Person perception theory and research suggests that instructions to disregard pertinent trait-related cues are often ineffective. This research suggests that interviewers may be unable to ignore unsolicited cues when assessing applicants’ personality traits. In this study, I hypothesize that unsolicited cues cannot be ignored in a personality-based structured interview.

When interviewers assess the focal trait targeted by a question, unsolicited cues provided in response to previous questions are to be ignored when question-based scoring is used. Consider an interview in which Question #1 targets achievement striving, a facet of conscientiousness, and Question #2 targets stress tolerance, a facet of emotional stability. Suppose that in response to Question #1 an applicant mentions an ongoing history of panic attacks – information that pertains directly to the applicant’s stress tolerance, which is to be assessed with Question #2. In the current investigation, I test if interviewers can ignore unsolicited cues when assessing a focal trait. However, whether ignoring unsolicited cues leads to more accurate predictions of future job performance (i.e., whether interviewers should ignore unsolicited cues, and whether ignoring unsolicited cues improves the predictive validity of interviews) is a complex issue to address and beyond the scope of this study. As noted, the purpose of this study is to test if interviewers can assess the focal trait of a question without
being influenced by unsolicited cues offered by an applicant in response to a previous question. Evidence that interviewers cannot ignore unsolicited cues, which is a critical assumption of question-based scoring, could cast doubt on the practical utility of this widely-used interview protocol.

In the introduction section I will review interview structure and the emergence of question-based scoring. I will then review examples from social sciences literature in which instructions to ignore information have proven ineffective and I will propose that such instructions may be particularly ineffective for assessing personality traits in an interview. Finally, I will explain how the experimental variables in this study will be combined to test my hypotheses.

**Interview Structure**

Interview structure improves the predictive validity of interviews by adding protocols to ensure that applicants are assessed on job-related constructs. Specially, structure maintains focus on job-related constructs by directing the manner in which interviewers ask and score interview questions (Campion, Palmer & Campion, 1997). Campion, Palmer and Campion (1997) identified 15 types of protocols that contribute to the degree of structure in an interview. Examples of these protocols include controlling the use of ancillary information, using scoring guides, and providing training to interviewers. Meta-analytic findings that the mean corrected (for range restriction and criterion unreliability) validity of structured interviews (.62) is double that of unstructured interviews (.31; Wiesner & Cronshaw, 1988), clearly indicate that structure helps target job-related constructs (Macan, 2009). By keeping the interview conversation focused on job-related constructs, structured interviews reduce applicants’ opportunities to discuss non job-relevant information (Blackman, 2009; Mayfield, 1964). This assertion is supported by meta-analytic findings that interviewers conducting unstructured interviews are far more influenced by non job-related information than are interviewers conducting structured interviews (Babyak, 2004; Ellis, West, Ryan & DeShon, 2002). Although structure helps target job-related constructs, very little is known about the effects of individual protocols, and scoring protocols in particular, on the predictive validity of interviews (Levashina, Hartwell, Morgeson, & Campion, 2014; Macan, 2009).
Scoring protocols and the Predictive Validity of Interviews

Among the 15 individual types of protocols, research has focused primarily on question standardization and scoring protocols used to rate applicants’ responses (Conway, Jako, & Goodman, 1995; Huffcutt & Arthur, 1994; Lievens & De Paepe, 2004; Macan, 2009). Question standardization pertains to the use of questions that target specific job-related constructs. Question standardization protocols can range from asking a series of impromptu questions (no structure) to using only standardized questions without any additional probing by the interviewer (high structure). Meta-analytic data show that the relationship between levels of question standardization and increased predictive validity is asymptotic; although low to moderate levels show significant increases, differences between moderate and high levels are non-significant (Huffcutt & Arthur, 1994; Huffcutt, Culbertson, & Weyhrauch, 2014). Because question-based scoring is paired with only the highest levels of question standardization, previous research has not determined if the asymptotic effect of increased question standardization on the predictive validity of interviews is related to question standardization, question-based scoring, or an interaction of these two variables.

Question-Based Scoring Prevailed Over Dimensional Scoring

Huffcutt and Arthur (1994) identified two scoring protocols associated with acceptable levels of reliability and validity\(^1\): dimensional and question-based (Conway, Jako, & Goodman, 1995). When dimensional scoring is applied, information provided in response to any question can be used to assess any construct. Alternatively, with question-based scoring each interview construct is assessed with a different question and unsolicited information (from another question) is to be ignored. Although it may seem perplexing that interviewers would be expected to ignore pertinent information, a review of interview structure research will help understand why question-based scoring came to dominate structured interview research.

Dimensional scoring emerged with Janz’s (1982, 1989) Patterned Behavioural Description Interview (PBDI) approach. With dimensional scoring, interview questions target

---

\(^1\) A third scoring protocol, global scoring, applies when interviewers make a single overall evaluation based on all information provided in the interview. Because the reliability and predictive validity associated with global scoring is vastly inferior to dimensional scoring and question-based scoring, the use of this scoring protocol in research has almost completely vanished over the last 20 years (Huffcutt, Culbertson, & Weyrauch, 2014).
different constructs but interviewers are allowed to consider unsolicited cues when assessing the focal trait of a question. In fact, Janz recognized the value of using unsolicited cues when rating a focal construct. The following passage from a PBDI instruction manual illustrates typical instructions for dimensional scoring:

> When you make the ratings, deal with one dimension at a time. Start with the first dimension and review your notes. Pay special attention to responses to questions that were targeted for that dimension. At the same time, be on the alert for any other responses that bear on the dimension in question. (Janz, Hellervik, & Gilmore, 1986, p 114)

Question-based scoring emerged with Latham, Saari, Pursell, & Campion’s (1980) Situational Interview (SI), in which each question presents a unique hypothetical scenario and responses are compared to a scoring guide that lists possible reactions to each scenario. Because each question illustrates a unique scenario, question-based scoring is integral to the SI; only solicited cues that describe actions within a unique hypothetical scenario can be assessed with the scoring guide. A contributing factor to the prevalence of question-based scoring is that it reflects a higher level of structure than does dimensional scoring (Campion, Palmer & Campion, 1997; Conway, Jako & Goodman, 1995). The eventual dominance of question-based scoring over dimensional scoring did not result from a direct comparison of these scoring protocols; rather, it resulted from a controversial theoretical debate surrounding whether it was better to predict performance by assessing past behaviours with Janz et al.’s (1982, 1989) PBDI or future intentions with Latham et al.’s (1980) SI.

A study by Campion, Campion and Hudson (1994) designed to compare PBDI and SI questions, but not scoring protocols, triggered an unwarranted demise of dimensional scoring. In this study, the authors contrived generic scoring guides for the PBDI questions and enforced question-based scoring across all conditions (Janz eschewed the use of scoring guides and question-based scoring for the PBDI); this allowed the authors to attribute differences across conditions to question design and it controlled for effects of scoring guides and scoring protocol.
Although the validity for the contrived PBDI questions (.51) was not significantly higher than for situational questions (.39), their results halted subsequent use of dimensional scoring and prompted researchers to exclude studies that used dimensional scoring from subsequent meta-analyses of structure (Taylor & Small, 2002). Given that Campion, Campion and Hudson’s (1994) study did not compare the merits of different scoring protocols, conclusions drawn from this study that question-based scoring is better than dimensional scoring were unsubstantiated.

Despite a marked decrease in the use of dimensional scoring, there is a lack of evidence that question-based scoring increases interview validity (Huffcutt, Culbertson, & Weyhrauch, 2014; Taylor & Small, 2002). Huffcutt, Culbertson, and Weyhrauch (2013) reported meta-analytic findings that the inter-rater reliability for question-based scoring ($\alpha = .61$) was higher than that of dimensional scoring ($\alpha = .48$). However, the authors found that interviews with question-based scoring (.70) did not correlate more highly with supervisor ratings of performance than did interviews that used dimensional scoring (.71). These results highlight the lack of evidence that question-based scoring produces better predictive validity than does dimensional scoring. Because question-based scoring relies on the assumption that unsolicited cues can be ignored, evidence that this assumption is flawed may clarify why question-based scoring fails to improve the predictive validity of interviews.

**The Flawed Assumption of Question-Based Scoring**

Instructions to ignore pertinent information may not be effective. In light of numerous examples from social sciences literature in which instructions to discount information are ineffective, the assumption that interviewers can ignore unsolicited cues may be flawed. Research in the domains of jury proceedings, employee selection, and person perception theory indicates that it is difficult to ignore pertinent information, even when there is a genuine intent to do so. The following five sections will review areas of research in which instructions to disregard information that is pertinent to a decision have proven ineffective.

**Inadmissible Information Affects Decisions in Jury Settings.** Kassin & Sommers (1997) conducted mock-trials that varied in terms of the presentation (presented/not presented) and admissibility (admissible/non-admissible) of a wiretapped confession. Relative to a control
condition in which jurors had no knowledge of a confession, jurors instructed to disregard the confession were much more likely to arrive to a guilty verdict. However, only 15 percent of participants who were instructed to disregard the confession listed it as an influential factor on their verdict. These results have two implications for the present investigation. First, disregarded information can remain influential. Second, when people try to disregard pertinent information, they may be unaware of its influence on their decision-making processes.

**Inadmissible Information Affects Ratings of Employability.** Employee equity legislation precludes employers from discriminating against job applicants based on prohibited grounds such as age, sex, and disability. Although hiring managers are expected to ignore this kind of information when making a hiring decision, research suggests that it is extremely difficult for hiring managers to ignore relevant information. Oien & Goernert (2003) presented introductory psychology students with 4 applications for a cashier-stockperson position. In some conditions, applicant #4 provided discriminatory information related to age (28 years old), disability (walks with a limp), and indicators of socio-economic status (takes the bus, has a GED, and information about alcohol consumption). Relative to participants in a control group who were not privy to any discriminatory information, participants who were exposed to the discriminatory information rated applicant #4 as significantly less employable than the other three applicants. These results suggest that, even when the law requires it, employers have difficulty ignoring prohibited information when evaluating job applications.

**Inadmissible Information Affects Decisions in Assessment Centers.** Research in employee selection indicates that hiring managers have a great deal of difficulty ignoring information about job applicants (Miron-Shatz & Ben-Shakhar, 2008; Oin & Goernert, 2003). Prior to assessing candidates’ assessment center performance, Miron-Shatz and Ben-Shakhar (2008) provided ancillary information to some human resources (HR) managers, but not to others; selected HR managers received a summary of a structured interview, personality test results, and cognitive ability test results. The ancillary information provided was artificially manipulated to reflect positively or negatively upon the applicants. Compared to a control group of HR managers who assessed the same candidates performing the same assessment center tasks without access to ancillary information, HR managers provided with ancillary information and
instructed to disregard it generated higher ratings when ancillary information was favourable and lower ratings when ancillary information was unfavourable. These results suggest that, just as HR managers have difficulty ignoring forbidden ancillary information, interviewers may have difficulty ignoring unsolicited cues.

**Inadmissible Information Affects Decisions in Interviews.** Research suggests that job irrelevant stereotypes affect interviewers’ decisions, even when they are instructed to ignore this type of information (Fiske, Bersoff, Borgida, Deaux, & Heilman, 1991; Heilman, Martell, & Simon, 1988). Segrest (2010) ran a study in which some participants attended rater error bias interview training that included techniques to reduce the effects of stereotypes on their judgments of applicants. All participants were instructed to rate the video-recorded interview performance of a job applicant who was either Hispanic or Anglo. Although training significantly reduced self-reported racio-ethnic prejudice, the Hispanic applicant’s interview performance was rated lower than that of the Anglo applicant, regardless of whether or not the raters received any training. Although participants in Segrest’s (2010) study were instructed to disregard racial stereotypes, they seemed unable to prevent this information from influencing their ratings. These results suggest that instructions and training to disregard unsolicited cues may have little effect in preventing such information from influencing interviewers’ ratings of a personality trait.

**Inadmissible Information Affects Perceptions of Personality.** Studies in person perception theory indicate that once people are aware of trait-related information, that information typically influences subsequent ratings of traits, even when people are instructed to disregard it (Wyer & Unverzagt, 1985; Golding, Fowler, Long, & Latta, 1990). In a study by Wyer and Lee Budesheim (1987), participants were presented with a series of 19 behaviours that were characteristic of a target, each written on a separate index card. In a condition that tested the effects of inconsistent inadmissible information, participants were presented with a subset of five kind cards (i.e. “Returned a wallet containing $50 to the lost and found”), a subset of five unkind cards (i.e. “Made fun of a disabled person”), and a subset of 9 neutral cards (i.e. “Exercised daily”). Participants who were instructed to disregard the kind cards prior to rating the target’s kindness rated the target as kinder than did participants in a control group that had
been shown only the unkind and neutral cards. The opposite was true for participants instructed to disregard the five unkind cards. Interestingly, these results did not hold if participants were told that the information to be disregarded was presented because of an administrative error and that it actually described a different person. These results illustrate the difficulty of disregarding inconsistent information and they suggest that interviewers may be unable to completely disregard unsolicited trait-related information. In light of numerous examples of jurors, employers, and interviewers being unable to disregard pertinent information when instructed to do so –in many cases by law, it seems unlikely that interviewers will be able to disregard unsolicited cues in a personality-based interview.

**Scoring Personality-Based Interviews**

Question-based scoring may be incompatible with personality-based interviews. Personality-based interviews may be the best method of assessing Five Factor Model (FFM) dimensions in selection settings. Meta-analytic data indicate that FFM dimensions are the most commonly assessed constructs in selection interviews (Huffcutt, Conway, Roth and Stone, 2001). Because job applicants tend to fake good on self-report measures of personality, interviewer ratings of personality traits predict job performance better than do self-report personality inventories (Alliger & Dwight, 2000; Birkeland et al., 2006; Huffcutt, Conway, Roth and Stone, 2001; Morgeson et al., 2007; Viswesvaran & Ones, 1999). Although selection interviews may mitigate the problems of using self-report measures of personality, personality-based structured interviews have unique challenges, which may include applying optimal levels of interview structure.

In personality-based interviews, interviewers may struggle with question-based scoring when personality traits must be inferred from very limited information. Research in behavioural profiling indicates that a minimum of six behavioural observations is typically required to estimate a behavioural disposition (Johnson, 1999; Shoda, Mischel, & Wright; 1994). Although self-report measures assess personality traits by averaging responses to scale items that target a variety of interpersonal circumstances, personality-based structured interviews often require interviewers to make such inferences based on an applicants’ reactions to a single, specific circumstance (Schmidt & Radar, 1999). In cases where applicants provide unsolicited trait-
related cues and question-based scoring is applied, such information is to be discarded by interviewers. Although personality traits can be accurately assessed with very limited observations (Ambady, Bernieri & Richeson; 2000), it may be difficult for interviewers to ignore unsolicited cues when very limited information is available to assess a trait. The present study will provide insight into interviewers’ decision-making in such circumstances.

**Can Interviewers Ignore Unsolicited Cues?**

This study will test the influence of unsolicited cues with a between-subjects experiment in an interview context. The dependent variable in this study will be interviewers’ assessment of a focal trait targeted by the second of two interview questions. Responses to the first question will include unsolicited cues (or not, in the neutral condition) to the focal trait and responses to the second question will include solicited cues to the focal trait.

The inclusion of two levels of scoring protocols, *dimensional scoring* and *question-based scoring*, will help determine if the effects of unsolicited cues can be eliminated with instructions given to interviewers. Participants assigned to dimensional scoring conditions will be instructed to consider unsolicited cues. Participants in question-based scoring conditions will be instructed to ignore unsolicited cues. If interviewers are unable to ignore unsolicited cues, ratings of a focal trait should not differ across scoring protocol conditions.

This study will employ the *averaging phenomenon* to test if question-based scoring can eliminate the influence of unsolicited cues. The *averaging phenomenon*, a quintessential finding in the area social cognition, is the tendency to make trait attributions that reflect an average of the valence (positive or negative) of two or more trait-related behaviours (Kashima & Kerekes, 1994). Research in person perception theory indicates that, as successive trait-related cues are observed, they are averaged into an overall trait inference that reflects an average of each individual observation (Wyer & Unverzagt, 1985) and this process has been shown to be automatic and subconscious (Tulving & Thomson, 1973). Because the averaging phenomenon can be automatic and subconscious, interviewers’ ratings of a trait in personality-based interviews may be influenced by unsolicited cues, even when question-based scoring implies that unsolicited cues are to be ignored.
Interviewers’ ability to ignore unsolicited cues can be measured by presenting some interviewers with unsolicited cues that contradict solicited cues, but not others. According to Golding et al (1990), trait-related cues can be positively or negatively valenced and weighted on a Likert-type rating scale; for example on a scale from 1 to 10 where very negative cues rate as 1 and very positive cues rate as 10. Unsolicited cues will be presented with audio-recorded responses to interview questions scripted to include positive, negative, or neutral (no) unsolicited cues. The effect of unsolicited cues will be tested by presenting interviewers with identical solicited cues and manipulating the valence of unsolicited cues.

If interviewers can ignore unsolicited cues, ratings of a focal trait made by interviewers exposed to contradictory unsolicited cues (prior to the solicited cues) should be no different from interviewers exposed to only solicited cues. However, if interviewers cannot ignore unsolicited cues, it is expected that the averaging phenomenon would cause interviewers exposed to negative unsolicited cues to give a lower rating of the focal trait compared to interviewers not exposed to such information. Interviewers exposed to negative unsolicited information would provide lower ratings because they would approximate the mean valence of the positive solicited cues and the negative unsolicited cues.

**Two Situations with Contradictory Cues.** Solicited and unsolicited cues can be contradictory in one of two ways; negative unsolicited cues may contradict positive solicited cues, or positive unsolicited cues may contradict negative solicited cues. In order to test the influence of unsolicited cues in both situations with contradictory cues, this study included three levels of unsolicited cue valence and two levels of solicited cue valence. The three levels of unsolicited cue valence will be positive, negative, and neutral. The two levels of solicited cue valence will be positive and negative. These variables allow negative unsolicited cues to be paired with positive solicited cues and for positive unsolicited cues to be paired with negative solicited cues. Table 1 shows the between subjects 2 (Scoring Method) X 3 (Valence of Unsolicited Cues) X 2 (Valence of Solicited Cues) fixed factorial design. The top half of Table 1 shows
comparisons between positive, negative, and neutral levels of unsolicited cues for each scoring protocol when solicited cues are positive. The bottom half of Table 1 shows these comparisons when solicited cues are negative. Whereas H1A, H2A, and H3A pertain to the effects of unsolicited cues when solicited cues are positive, H1B, H2B, and H3B pertain to the effects of unsolicited cues when solicited cues are negative.

The test of H1A will demonstrate the effects of unsolicited cues when positive solicited cues are assessed with dimensional scoring – interviewers in these conditions will be encouraged to consider unsolicited cues when rating a focal trait. To test H1A, some interviewers will be presented with negative unsolicited cues that contradict positive solicited cues; however, other interviewers will be presented with either positive or neutral unsolicited cues instead. Person perception research shows that ignoring forbidden information is often (but not always) a two-step process: first, an initial impression is formed using all relevant cues, and second, the initial impression is corrected (by increasing or decreasing the rating of a trait) by the degree of influence that forbidden cues are estimated to have had (Schul & Burnstein, 1985; Wyer & Lee Buddesheim, 1987). Because dimensional scoring conditions will not require participants to ignore unsolicited cues, they provide a measure of the effects of unsolicited cues before corrections. Therefore, the tests of H1A will serve primarily as a manipulation check that negative unsolicited cues have the potential to lower the ratings of a trait; if unsolicited cues do not lower ratings in dimensional scoring conditions, this would indicate insufficient salience to influence judgments.

H1A predicts that the mean of Cell 2 in Table 1 will be lower than the mean of Cell 1 and lower than the mean of Cell 3. In conditions that combine dimensional scoring and positive solicited cues, I predict that negative unsolicited cues will result in lower ratings of a focal trait:

H1A: When interviewers apply dimensional scoring to assess positive solicited cues, the ratings of the focal trait made by interviewers with negative unsolicited cues will be lower than those made by interviewers presented with positive unsolicited cues and lower than those made by interviewers presented with neutral unsolicited cues.
H2A will test the effects of unsolicited cues when positive solicited cues are assessed with question-based scoring. Interviewers in these conditions will be instructed to consider only solicited cues when rating a focal trait. In the test of H1A, is it expected that negative unsolicited cues will lower ratings of a focal trait—the purpose of H2A is to test if this effect can be eliminated with instructions to ignore unsolicited cues. Similar to the tests of H1A, the test of H2A will include a between-subjects comparison in which some interviewers will be presented with negative unsolicited cues that contradict solicited cues, whereas others will be presented with positive or neutral unsolicited cues instead. In contrast to the test of H1A, which applies dimensional scoring, H2A will test the effects of unsolicited cues when question-based scoring is applied. In H2A I predict that the mean of Cell 5 in Table 1 will be lower than the mean of Cell 4 and lower than the mean of Cell 6. In conditions that combine question-based scoring with positive solicited cues, I predict that negative unsolicited cues will lower ratings of a focal trait—despite instructions to ignore unsolicited cues:

H2A: When interviewers apply question-based scoring to assess positive solicited cues, the ratings of the focal trait made by interviewers presented with negative unsolicited cues will be lower than those made by interviewers presented with positive unsolicited cues and lower than those made by interviewers presented with neutral unsolicited cues.

H3A will compare the effects of negative unsolicited cues in dimensional scoring and question-based scoring conditions. H3A will demonstrate if instructions given to interviewers to ignore negative unsolicited cues (question-based scoring) can eliminate the influence of such cues on interviewers’ ratings. I predict that, when interviewers are presented with positive solicited cues, negative unsolicited cues will result in lower ratings of the focal trait uniformly across question-based and dimensional scoring conditions. In H3A, I predict that there will be no difference between the means of Cell 2 and Cell 5 in Table 1:

H3A: The ratings of a focal trait made by interviewers presented with negative unsolicited cues and positive solicited cues will not differ across question-based and dimensional scoring conditions.
The tests of H1B resembles the test of H1A, except that it will demonstrate the effects of unsolicited cues when negative solicited cues are assessed with dimensional scoring –as with H1A, interviewers in these conditions will be encouraged to consider unsolicited cues when rating a focal trait. To test H1B, *some* interviewers will be presented with unsolicited cues that contradict solicited cues, but not others. Similar to the test of H1A, the test of H1B serves as a manipulation check of the effects of unsolicited cues without interviewers *correcting* their ratings to eliminate the influence of such information. If positive unsolicited cues do not inflate ratings in dimensional scoring conditions, this would indicate insufficient salience to influence judgments.

H1B predicts that the mean of Cell 7 in Table 1 will be higher than the mean of Cell 8 and higher than the mean of Cell 9. In conditions that combine dimensional scoring and negative solicited cues, I predict an effect of unsolicited cue valence such that positive unsolicited cues will result in higher ratings of a focal trait:

H1B: When interviewers apply dimensional scoring to assess negative solicited cues, the ratings of the focal trait made by interviewers presented with positive unsolicited cues will be higher than those made by interviewers presented with negative unsolicited cues and higher than those made by interviewers presented with neutral unsolicited cues.

H2B will test the effects of unsolicited cues when negative solicited cues are assessed with question-based scoring. Interviewers in these conditions will be instructed to consider only solicited cues when rating a focal trait. In the test of H1B, is it expected that positive unsolicited will inflate ratings of a focal trait –the purpose of H2B is to test if this effect can be eliminated with instructions to ignore unsolicited cues. Similar to the test of H1B, the test of H2B will include a between-subjects comparison in which *some* interviewers will be presented with positive unsolicited cues that contradict solicited cues, whereas others will be presented with either negative or neutral unsolicited cues instead. In contrast to the test of H1B, which applies dimensional scoring, H2B will test the effects of unsolicited cues when question-based scoring is applied. In H2B I predict that the mean of Cell 10 in Table 1 will be higher than the mean of
Cell 11 and higher than the mean of Cell 12. In conditions that combine question-based scoring with negative solicited cues, I predict that positive unsolicited cues will inflate ratings of a focal trait—despite instructions to ignore unsolicited cues:

**H2B**: When interviewers apply question-based scoring to assess negative solicited cues, the ratings of a focal trait made by interviewers presented with positive unsolicited cues will be higher than those made by interviewers presented with negative unsolicited cues and higher than those made by interviewers presented with neutral unsolicited cues.

**H3B** will compare the effects of positive unsolicited cues across dimensional scoring and question-based scoring conditions. **H3B** will demonstrate if instructions given to interviewers to ignore positive unsolicited cues can eliminate the influence of such cues on interviewers’ ratings. I predict that, when interviewers are presented with negative solicited cues, positive unsolicited cues will inflate ratings of the focal trait uniformly across question-based and dimensional scoring conditions. In **H3B**, I predict that there will be no difference between the means of Cell 7 and Cell 10 in Table 1:

**H3B**: The ratings of a focal trait made by interviewers presented with positive unsolicited cues and negative solicited cues will not differ across question-based and dimensional scoring conditions.

**Interviewers Cannot Disregard Trait-Related Information**

The proposed study challenges the assumption that interviewers can perform a cognitive task that is implied by question-based scoring; I propose that interviewers cannot ignore unsolicited trait-related cues when rating a job applicant’s personality traits. Although personality has been linked to important organizational outcomes, the utility of self-report personality measures in selection settings has come under scrutiny. Structured interviews may be the best alternative for assessing personality in selection settings but ensuring that trait-related cues are evaluated fairly across job applicants remains a challenge; in response to a single question, some job applicants may give cues related to many personality traits whereas others
may focus only on the trait being targeted by a specific question. A common solution has been to apply question-based scoring, which is intended to preclude interviewers from considering unsolicited cues. Drawing on evidence from social psychology and person perception theory, this study aims to explore if disregarding unsolicited cues is more difficult than heretofore assumed.

This study uses a between subjects 2 (Scoring Method) X 3 (Valence of Unsolicited Cues) X 2 (Valence of Solicited Cues) design (see Table 1 for the total set of conditions). The dependent variable was the ratings of the applicant’s response to the focal trait of Question #2, which were expected to vary as a function of unsolicited and solicited cue valence but not the scoring protocol conditions.

**Method**

**Participants**

Two hundred and forty-seven participants from USA were recruited through CrowdFlower, an Internet-based system through which internet users can participate in social sciences research in exchange for a very small sum of money, typically less than $0.50 USD (Mason & Suri, 2012). Recent research by Buhrmester, Kwang and Gosling (2011) that examined the suitability of using crowd-sourced data for publishing in scientific journals showed that micro-workers are more demographically diverse than American college samples and typical Internet samples. Because participants were required to infer personality traits from complex verbal responses spoken in English, only CrowdFlower workers who had demonstrated proficiency on other online tasks completed in the English language were allowed to participate.

To be eligible to participate in the study, participants needed to be a registered CrowdFlower contributor and have access to a laptop or desktop computer with a high-speed Internet connection, sound capabilities, and Adobe Flash Player software. Sound capabilities were verified with an audio-recorded password that granted access to the study. All participants were required to demonstrate a basic understanding of the rating task in a threshold test (explained below). A total of 720 CrowdFlower workers logged onto the study; 196 aborted without providing initial consent, 76 failed the sound check, 70 aborted prior to completing the
threshold test, 126 failed the threshold test, 3 Passed the threshold test but aborted before completing the study, and 2 completed the study but withdrew consent.

Participants were compensated $0.50 USD for their participation that took approximately 17 minutes (SD = 7 minutes). The job settings of the online task and Qualtrics security features prevented participants from participating more than once. The gender of participants was split evenly between women (n = 128; 52%) and men (n = 118; 48%). The largest proportion of respondents were between 18 and 34 years of age (n = 139; 56%), followed by participants aged 35 to 55 years of age (n = 88; 36%), with the smallest group comprised of those aged 56 years and older (n = 19; 8%). With regards to participants’ education, 48% (n = 117) completed only high school, 41% (n = 102) completed a college diploma or university degree, and 11% (n = 27) completed graduate school or higher. When asked to indicate their employment status, 48% (n = 117) worked full time, 25% (n = 62) were unemployed, 16% (n = 39) were employed part time, and 11% (n = 28) were students. Out of 48 participants (19% of sample) who reported working in a managerial position, 56% (n = 27) supervised less than five subordinates and 44% (n = 21) supervised five or more subordinates. Among the 59 participants (24% of sample) who indicated that they had experience as an interviewer or hiring manager, 36% (n = 21) had interviewed between one and nine candidates, 37% (n = 21) had interviewed between ten and thirty candidates, and 27% (n = 16) had interviewed over 30 candidates.

**Apparatus and Materials**

**CrowdFlower Online Platform.** CrowdFlower is a web-based crowd-sourcing company comparable to Amazon’s Mechanical Turk (MTurk) and is based in San Francisco, California, USA. According to a 2013 survey of its online labour pool, CrowdFlower has approximately 900,000 micro-workers living in the USA; among these workers 48% are below the age of 40, 54% completed post-secondary education, 78% are Caucasian, 68% are female, 58% have a household income less than $60,000.00 per year, 47% are married or live in a domestic partnership, and 55% reported that they participate in micro-work because “it’s a great way to spend free time and get some cash” (Zukoff, 2014).
**Qualtrics Online Software Web Platform.**

Qualtrics, an online software program designed to build and manage online surveys, was used to build an online audio-tutorial comprised of four slides that explained basic principles of structured employment interviews, the use of structured interview questions, and rating scales. The audio-tutorial slides were timed so that participants could not advance without having heard the complete content of each slide. Participants could listen to all audio-recordings in the study as many times as they wished prior to advancing slides. However, they could not navigate back to previous slides nor could they go back and change their ratings of the applicant’s traits once they had been submitted.

**Rating Scales.** For each question, participants were provided with rating scales for rating achievement striving and stress tolerance. These included behavioural examples that correspond with high levels of each trait. The rating scales included a solid black horizontal line that spanned across participants’ computer screen, under which the numbers one through ten were spaced out evenly below the line. The numbers one, five, and ten were identified as “very low”, “typical”, and “very high”, respectively. Appendices A and B show the rating scales that were used to assess achievement striving and stress tolerance, respectively.

**Threshold Test.** In order to ensure that participants understood the instructions delivered in the audio tutorial, participants listened to four audio clips that illustrated high and low levels of achievement striving and stress tolerance. Each audio clip was played by a different actor. Only participants who rated responses indicative of a high level of a trait with a score of 6 out of 10 or higher and who rated responses indicative of a low level of a trait with a score of 4 out of 10 or lower were allowed to proceed to the experimental phase of the study. As a security measure to prevent cheating, the presentation of threshold test items was randomized across participants. See Appendices C, D, E, and F for threshold test items 1, 2, 3, and 4, respectively.

**Manipulation Check and Post-Experimental Questionnaire.** First year undergraduate psychology students were recruited to participate in a pilot study. The pilot study served four purposes. First it verified the proper functioning of the online platform. Second, ratings
collected from the pilot study were analyzed to ensure differences in stress tolerance cues could be detected accurately. Third, it included a post-experimental questionnaire that collected participants’ impressions of the experimental task; specifically, the post-experimental questionnaire assessed what participants could remember about the target and their impressions about the rules used to rate the applicant’s personality traits. Participants were asked if they followed the instructions given to them and if they felt they were able to follow those instructions. See Appendices G and H for a summary of the manipulation checks and the questionnaire itself. Fourth and finally, the pilot study provided evidence that the deception used in the main study would be credible; students were told their ratings would be used to make hiring decisions within the university. Short focus groups held after each experimental session revealed that the deception was very compelling – participants in the pilot study genuinely believed that they had listened to audio recordings of a real applicant applying for a real job and that the university intended to use their responses to make actual hiring decisions.

Procedure

CrowdFlower workers from USA chose to participate in this study from other CrowdFlower online tasks, which were indexed by the task duration, date of posting, and pay incentive. The heading of the task read “Interview Like a Pro Project” and the initial description of the task indicated that participants would be taught industry leading interview techniques and that they would subsequently be asked to provide ratings of real job applicants that would be used to make real crowd-sourced hiring decisions for applicants to the US federal government; this deception was used to increase participant’s motivation and sense of accountability. See Appendix I for the complete initial description of the “Interview Like a Pro Project” CrowdFlower task. CrowdFlower workers who chose to participate in this task followed a web link from the CrowdFlower portal to an independently hosted Qualtrics webpage.

Upon navigating to the Qualtrics website, CrowdFlower workers entered their CrowdFlower ID and agreed to an initial letter of information and consent. See Appendix J for the complete initial letter of information and consent with deception. This consent form provided accurate information about participants’ rights as research participants but false information related to the research organization and the true purpose of the study. In order to
conceal the researchers’ institutional affiliation temporary email accounts were created and set up to auto-forward inquiries to the principle investigator and the university director of research ethics; real telephone numbers with Ontario area codes were provided.

Participants were falsely informed that they would be randomly assigned to rate one of many candidates and that the candidates that received the highest crowd-sourced scores would be hired. Participants then read and listened to the content of the audio tutorial, which explained basic principles of structured interviews and how to use rating scales to assess achievement striving and stress tolerance. Upon completion of the tutorial, participants proceeded to the threshold test. Those who passed the threshold test were pseudo-randomly assigned to one of 12 experimental conditions; these included two levels of scoring protocols, three levels of unsolicited stress tolerance cues, and two levels of solicited stress tolerance cues. Upon completion of the study participants were presented with a second letter of information and consent and they were informed of the true nature and intent of the study and asked to provide informed consent. See Appendix K for the debrief and second letter of information and consent.

**Independent and Dependent Variables**

**Scoring protocol (Question-Based/Dimensional).** Upon completion of the threshold test, participants were randomly assigned to one of two levels of scoring protocol: question-based scoring or dimensional scoring. The question-based scoring condition instructed participants to ignore unsolicited cues when assessing a focal trait. The dimensional scoring condition presented written (presented on their computer monitor) and audio (read by the audio-tutorial narrator) instructions to participants to consider all known information about the applicant when assessing each trait; participants in this condition were encouraged to integrate unsolicited cues into their assessments of each trait. See appendix L for the instructions given in each of the scoring conditions.

**Unsolicited Cue Valence (Positive/Negative/Neutral).** Participants were presented with one of three levels of unsolicited cues that pertained to stress tolerance; these included positive, negative, and neutral unsolicited cues, which are described below. After having received scoring instructions, participants listened to an audio-recorded interview in which achievement striving
and stress tolerance were targeted with separate questions. Unsolicited cues that pertained to stress tolerance were presented in the applicant’s response to the question that targeted achievement striving. Note that in all three unsolicited cue conditions, the applicant’s response to Question #1, which targeted achievement striving, did not include information that pertained to achievement striving. Depending on the level of unsolicited cue valence to which participants were assigned (positive/negative/neutral), they would listen to one of three individually scripted responses to the question that targeted achievement striving; all three scripts were recorded separately by the same actress who was introduce to participants as Mrs. Sophia Moore. The positive unsolicited cues read as follows: “I focused on the most difficult aspects of the task first and I made a few suggestions to my boss about how we could improve our regular procedures – my boss was very appreciative and I think it brought us closer together as a team.” The negative unsolicited cues read as follows: “This was very stressful for me -I knew that I would have to really struggle to control my anxiety –even though it was apparent to everyone around me how stressed I was. Because it was taking a toll on my relationships at work, I avoided dealing with it by calling in sick.” The neutral unsolicited cues read as follows: “I knew that it was going to be a busy week for the staff of my crew. We had no choice but to try to complete the order. We did as much as we could to complete the order as quickly as possible.” See Appendix M for the complete scripts of the unsolicited cues.

**Solicited Cue Valence (Positive/Negative).** After having rated the applicant’s achievement striving, participants were pseudo-randomly (Qualtrix allocated participants to different conditions on a rotating basis) assigned to one of two levels of solicited cue valence. Solicited cues (positive/negative) were presented with separate audio-recordings of Mrs. Sophia Moore’s response to the question that targeted stress tolerance. The positive solicited cue read as follows: “Despite all of my responsibilities, I never stressed about all of my tasks, I worked diligently, and thought of ways to reduce the stress at the office”. The negative solicited cue read as follows: “Given all of my responsibilities, I was very (with vocal emphasis on the word “very”) anxious about all of my tasks and I was constantly panicked that I wouldn’t be able to manage everything”. See Appendix N for the complete scripts of solicited cues.
Dependent Variable: Ratings of the Focal Trait. After having listened to the applicant’s response to Question #2, participants were asked to rate the applicant’s stress tolerance on a 10-point scale using the rating scale for stress tolerance provided in Appendix B. It was expected that unsolicited cues would influence ratings of the focal trait. The effect of unsolicited stress tolerance cues was measured with ratings of stress tolerance.

Results

General

The effects of unsolicited cue valence on ratings of a focal trait were examined using a 2 (Scoring Protocol) X 3 (Valence of Unsolicited Cues) X 2 (Valence of Solicited Cues) design. This section will first present the effects of negative unsolicited cues (H1A, H2A, and H3A) followed by the effects of positive unsolicited cues (H1B, H2B, and H3B).

Test of Hypothesis 1A

H1A tested the effect of unsolicited cue valence when positive solicited cues were assessed with dimensional scoring; I predicted that the mean of Cell 2 in Table 2 would be lower than the mean of Cell 1 and lower than the mean of Cell 3. Results confirmed that when positive solicited cues were assessed with dimensional scoring, negative unsolicited cues resulted in lower ratings of a focal trait. As shown in the cluster on the left of Figure 1, there was an effect of unsolicited cue valence such that negative unsolicited cues ($M = 6.5$, $SD = 2.09$, $SE = .47$, $N = 20$) resulted in lower trait ratings than positive unsolicited cues ($M = 8.78$, $SD = 1.11$, $SE = .26$, $N = 18$, $t(36) = -4.25$, $p = .00$, $d = 1.34^2$, 95% CI [0.61, 2.01]), and neutral unsolicited cues ($M = 8.24$, $SD = 1.18$, $SE = .26$, $N = 21$, $t(39) = -3.26$, $p = .00$, $d = 1.03$, 95% CI [0.36, 1.66]), when positive solicited cues were assessed with dimensional scoring. These results support the hypothesis that in interview settings that apply dimensional scoring, exposure to negative unsolicited cues that contradict positive solicited cues may lower ratings of a focal trait.

---

2 Cohen (1977) suggested that $d$ values of .8, .5, and .2 expressed large, medium, and small effect sizes, respectively.
Test of Hypothesis 2A

H2A tested the effect of unsolicited cue valence when positive solicited cues were assessed with question-based scoring; I hypothesized that exposure to negative unsolicited cues would lower interviewers’ ratings of a focal trait when solicited cues were positive— even when interviewers are instructed to ignore unsolicited cues. Specifically, I predicted that the mean of Cell 5 in Table 2 would be lower than the mean of Cell 4 and lower than the mean of Cell 6. As shown in the cluster on the right of Figure 1, negative unsolicited cues ($M = 8.53, SD = 1.35, SE = .31, N = 19$) did not result in lower trait ratings than positive unsolicited cues ($M = 8.73, SD = 0.99, SE = .22, N = 20, t(37) = -.52, p = .61, d = 0.17, 95% CI [-0.46, 0.79]) or neutral unsolicited cues ($M = 8.31, SD = 1.12, SE = .22, N = 26, t(43) = -0.58, p = .57, d = 0.18, 95% CI [-0.42, 0.77]) when positive solicited cues were assessed with question-based scoring. Given that H1A was supported, these results indicate that instructions to ignore unsolicited cues completely eliminated the effects of such cues on ratings of a focal trait. These results imply that the effects of negative unsolicited cues can be eliminated with instructions to interviewers to consider only solicited cues when assessing a personality trait.

Test of Hypothesis 3A

I hypothesized that, when interviewers are presented with positive solicited cues, negative unsolicited cues would result in lower ratings of the focal trait across question-based and dimensional scoring conditions. Specifically, I predicted that there would be no difference between the means of Cell 2 and Cell 5 in Table 2. Because the effect of unsolicited cues was eliminated when positive solicited cues were assessed with question-based scoring, results did not support H3A. As shown in Figure 2, there was an effect of scoring protocol such that dimensional scoring ($M = 6.5, SD = 2.09, SE = .47, N = 20$) resulted in lower trait ratings than question-based scoring ($M = 8.53, SD = 1.35, SE = .31, N = 19$) when unsolicited cues were negative and solicited cues were positive, $t(37) = -3.62, p = .00, d = 1.15, 95\% \text{ CI } [0.45, 1.80]$. Combined with results from the test of H2A, which showed that the effect of negative unsolicited cues was eliminated in question-based scoring conditions, the rejection of H3A indicates that interviewers successfully ignored negative unsolicited cues when instructed to do so.
Test of Hypothesis 1B

H1B tested the effect of unsolicited cue valence when negative solicited cues were assessed with dimensional scoring; I predicted that the mean of Cell 7 in Table 2 would be higher than the mean of Cell 8 and higher than the mean of Cell 9. As shown in the cluster on the left of Figure 3, positive unsolicited cues ($M = 3.64$, $SD = 1.68$, $SE = .34$, $N = 25$) did not result in higher trait ratings than negative unsolicited cues ($M = 2.86$, $SD = 1.83$, $SE = .39$, $N = 22$, $t(45) = -1.51$, $p = .14$, $d = 0.44$, 95% CI [-0.14, 1.02]), or neutral unsolicited cues ($M = 4.43$, $SD = 1.65$, $SE = .44$, $N = 14$, $t(37) = 1.42$, $p = .16$, $d = 0.47$, 95% CI [-0.20, 1.12]), when negative solicited cues were assessed with dimensional scoring (the significant difference between neutral unsolicited cues and negative unsolicited cues will be discussed in the exploratory analyses section below). These results indicate that the effect of positive unsolicited cues was not sufficient to boost impressions of the interviewee’s focal trait. Results indicate that, when solicited cues were negative, positive unsolicited cues did not influence interviewers judgments sufficiently to elevate ratings of the applicant’s stress tolerance.

Test of Hypothesis 2B

H2B tested the effect of unsolicited cue valence when negative solicited cues were assessed with question-based scoring; I hypothesized that exposure to positive unsolicited cues would inflate interviewers’ ratings of a focal trait when solicited cues were negative –even when question-based scoring was applied. Specifically, I predicted that the mean of Cell 10 in Table 2 would be higher than the mean of Cell 11 and higher than the mean of Cell 12. Results did not support H2B. As shown in the cluster on the right of Figure 3, positive unsolicited cues ($M = 2.75$, $SD = 1.41$, $SE = .32$, $N = 20$) did not result in higher trait ratings than negative unsolicited cues ($M = 2.25$, $SD = 1.07$, $SE = .22$, $N = 24$, $t(42) = -1.30$, $p = .20$, $d = 0.40$, 95% CI [-0.20, 1.00]), or neutral unsolicited cues ($M = 3.67$, $SD = 1.41$, $SE = .33$, $N = 18$, $t(36) = 2.00$, $p = .05$, $d = 0.65$, 95% CI [-0.02, 1.29]), when negative solicited cues were assessed with question-based scoring. Although results did not support H2B, the results of H1B (showing that positive unsolicited cues did not inflate ratings of the focal trait in dimensional scoring conditions) indicate that the manipulations in the test of H2B failed to validly assess if positive solicited cues can be ignored. Stated otherwise, the test of H2B could/did not validly assess if participants
successfully ignored positive unsolicited cues because the test of H1B showed that interviewers essentially failed to notice them in the first place.

**Test of Hypothesis 3B**

I hypothesized that, when interviewers are presented with negative solicited cues, positive unsolicited cues would result in inflated ratings of the focal trait across dimensional and question-based scoring conditions. Specifically, I predicted that there would be no difference between the means of Cell 7 and Cell 10 in Table 2. As shown in Figure 4, the effect of scoring protocol was non-significant across dimensional ($M = 3.64$, $SD = 1.68$, $SE = .34$, $N = 25$) and question-based ($M = 2.75$, $SD = 1.41$, $SE = .32$, $N = 20$) scoring protocol conditions when unsolicited cues were positive and solicited cues were negative, $t(43) = 1.93$, $p = .06$, $d = 0.57$, 95% CI [-0.04, 1.16]. Although H3B was supported, the lack of support for H1B and H2B indicates that the test of H3B was not a valid test interviewers’ ability to ignore positive unsolicited cues—as mentioned previously, positive unsolicited cues lacked salience or were unnoticed by participants.

**Exploratory Analyses**

A 2 (Scoring Protocol) X 3 (Valence of Unsolicited Cues) fixed factorial ANOVA was used to explore the effects of neutral unsolicited cues when solicited cues were negative. In accordance with Howell (2012), this examination of the *conditional effect* of unsolicited cues in negative solicited cue conditions excluded the pooled variance for positive solicited cue conditions. Although the two-way interaction between scoring protocol and unsolicited cue valence was non-significant, $F(5, 117) = 0.92$, $p = .91$, partial $\eta^2 = .00^3$, the main effects for scoring protocol, $F(1, 117) = 7.26$, $p = .01$, partial $\eta^2 = .06$, and unsolicited cues, $F(2, 117) = 8.94$, $p = .00$, partial $\eta^2 = .13$, were significant. Levene’s test for equality of variances, $F(5, 117) = 0.58$, $p = .71$ was non-significant.

As shown in the cluster on the left of Figure 3, there was an effect of unsolicited cue valence such that neutral unsolicited cues ($M = 4.43$, $SD = 1.65$, $SE = .46$, $N = 14$) resulted in

\[^3\] Indices for partial-eta squared effect sizes: small = .01; medium = .06; large = .14 (Fritz, Morris, & Archer; 2012).
higher trait ratings than negative unsolicited cues ($M = 2.86, SD = 1.83, SE = .37, N = 22$) when negative solicited cues were assessed with dimensional scoring, $F(2, 58) = 3.57, p = .04, d = .89, 95\% \text{CI} [0.17, 1.57]$. The cluster on the right of Figure 3 shows a similar effect such that neutral unsolicited cues ($M = 3.67, SD = 1.41, SE = .30, N = 18$) resulted in higher trait ratings than positive unsolicited cues ($M = 2.75, SD = 1.41, SE = .29, N = 20, d = .65, 95\% \text{CI} [-0.02, 1.29]$) and negative unsolicited cues ($M = 2.25, SD = 1.07, SE = .26, N = 24, d = 1.15, 95\% \text{CI} [0.47, 1.79]$) when negative solicited cues were assessed with question-based scoring, $F(2, 59) = 6.25, p = .00, \eta^2 = .18$. These results suggest that when solicited cues were negative, neutral unsolicited cues had a leniency effect such that ratings of the focal trait were inflated.

Further analysis revealed that when solicited cues were negative, neutral unsolicited cues inflated ratings of the focal trait equally across dimensional and question-based scoring conditions. As shown in Figure 5, the effect of scoring protocol was non-significant across dimensional ($M = 4.43, SD = 1.65, SE = .44, N = 14$) and question-based ($M = 3.67, SD = 1.41, SE = .33, N = 18$) scoring protocol conditions when unsolicited cues were neutral and solicited cues were negative, $t(30) = 1.38, p = .18, d = 0.50, 95\% \text{CI} [-0.22, 1.2]$. These results are inconsistent with the findings of H1A; whereas question-based scoring eliminated the effect of unsolicited cues when solicited cues were positive, question-based scoring did not eliminate the effect of neutral unsolicited cues when solicited cues were negative.

**Discussion**

Evidence that interviewers can ignore unsolicited cues was mixed. H2A, which predicted that interviewers would be unable to ignore negative unsolicited cues when solicited cues were positive, was not supported. H2B, which predicted that interviewers would be unable to ignore positive unsolicited cues when solicited cues were negative, was not supported. However, post-hoc analyses indicated that when solicited cues were negative, neutral unsolicited cues inflated ratings of the focal trait across dimensional and question-based scoring conditions. Results indicated that at least one of the study’s manipulations did not operate as intended and that the ability to ignore unsolicited cues may be linked to the favourability of information.
Ignoring Contradictory Negative Information

H1A tested if interviewers presented with positive solicited cues could rate a focal trait without being influenced by negative unsolicited cues. The effect of the negative unsolicited cues was tested with a between-subjects comparison in which interviewers presented with positive solicited cues were randomly assigned to one of three unsolicited cue conditions: negative, neutral, or positive. My prediction that the mean of Cell 2 would be lower than the mean of Cell 1, and lower than the mean of Cell 3 was supported. This indicates that exposure to negative unsolicited cues may result in lower ratings of a focal trait when dimensional scoring is applied.

H2A predicted that exposure to negative unsolicited cues would result in lower ratings of a focal trait when positive solicited cues were assessed with question-based scoring; negative unsolicited cues were expected to lower ratings of the focal trait despite instructions to ignore such cues. H2A was not supported such that differences between positive, negative, and neutral unsolicited cue conditions were non-significant when positive solicited cues were assessed with question-based scoring conditions. These results indicate that the influence of unsolicited cues can be eliminated with question-based scoring when solicited cues are positive.

H3A predicted that negative unsolicited cues would lower ratings of a focal trait in positive solicited cue conditions across dimensional and question-based scoring conditions; H3A predicted that there would not be a significant difference between the means of Cell 2 and Cell 5 in Table 2. However, among interviewers presented with negative unsolicited cues and positive solicited cues, the difference in ratings between interviewers permitted to consider unsolicited cues (dimensional scoring condition; $M = 6.5, SD = 2.02$) and those instructed to ignore unsolicited cues (question-based scoring conditions; $M = 8.78, SD = 1.11$) showed a very large effect size ($d = 1.34, 95\% CI [0.62, 1.70]$). H3A was therefore not supported and results from the tests of H1A, H2A, and H3A showed that when solicited cues were positive, instructions given to interviewers in question-based scoring conditions eliminated the effects of unsolicited cues.

The results that interviewers can ignore negative unsolicited cues is inconsistent with substantial evidence from personnel selection research that instructions to disregard pertinent
information are ineffective. It is important to acknowledge that interpretation of the effects of negative unsolicited cues in this study must be tempered until further research confirms that the results of this study do not result from methodological error and threats to validity discussed in the limitations section below. However, if the effects of negative unsolicited cues are supported with further investigations, the following three paragraphs illustrate factors that may have helped interviewers ignore negative unsolicited cues.

Previous research has shown that instructions given to interviewers to disregard applicant demographics (race, gender, age, disabilities), credentials, and ancillary information (cognitive ability test scores, resumes, personality test results), and personal appearance are ineffective (Arvey, 1979; Miron-Shatz & Ben-Shakhar, 2008; Oien & Goernert, 2003; Schlenker, 1980; Tucker & Row, 1979). Research also shows that interviewers weigh negative information more heavily than positive information (Morton, 1994; Schmitt, 1976), and that this effect is magnified when first impressions are negative (Dougherty, Turban & Callender, 1994; Farr, 1973). Given the effects of negative information shown in other personnel selection research, it is surprising that interviewers in this study could ignore first impressions of an applicant that included negative unsolicited cues.

Despite limited research in personnel selection showing that interviewers can ignore pertinent information (Arvey & Campion, 1982; Schlenker, 1980), other social science literature indicates that instructions to disregard information are most effective when the information to be ignored is negative, when the information to be ignored is not emotionally charged, and when instructions are given sooner rather than later. In the aforementioned study by Wyer and Budesheim (1987) in which instructions to ignore trait-related cues were ineffective, participants were better able to ignore negative cues than positive cues; although instructions did not eliminate the influence of positive and negative cues, participants’ ratings of a trait were less influenced by negative cues than by positive cues. The authors attributed this effect to participants’ aversion to give a negative evaluation. Results from this study showing that participants could ignore negative unsolicited cues may have been facilitated by participants’ aversion to giving a negative evaluation.
Participants’ ability to ignore negative unsolicited cues more easily than other cues contrasts with previous findings that emotionally charged information is tougher to ignore. Research in trial proceedings indicates that emotionally charged information is harder to ignore (Kassin & Sommers, 1997). Edwards and Bryan (1997) conducted identical mock trials in which a witness’ testimony was manipulated to reflect either graphic phrases (i.e., “hacked up”) or neutral language (i.e., “assault with a weapon”). Instructions to disregard emotionally charged testimony showed a paradoxical effect; compared to jurors in a control condition who were permitted to consider the graphic testimony, the influence of emotionally-charged testimony was magnified among jurors instructed to disregard it. If emotionally charged information is more difficult to ignore, participants’ ability to ignore disclosures of absenteeism and emotional distress presented in the negative solicited cues (i.e., “Because [my anxiety] was taking a toll on my relationships at work, I avoided dealing with it by calling in sick.” See Appendix J for full responses.) is surprising.

Participants’ ability to ignore negative unsolicited cues may have been facilitated by the way in which instructions to ignore unsolicited cues were presented. Research in intentional forgetting shows that instructions to ignore pertinent information are most effective when given at the time of encoding (Johnson, 1994). Whereas previous person perception studies instructed participants to ignore trait-related cues after they have been presented, the instructions in this study were given before the presentation of trait-related cues such that participants knew to ignore unsolicited cues at the time of encoding. Therefore, participants’ ability to ignore trait-related cues may have been more effective in this study than in previous person perception studies because instructions to ignore such information were given to participants before, rather than after, such cues were presented. However, other research has shown interviewers to be swayed by prohibited grounds of discrimination despite foreknowledge that such information is to be ignored (Pingitore, Dugoni, Tindale, & Spring, 1994; Segrest, 2010). Therefore, it is unlikely that the timing of instructions to apply question-based scoring in this study accounts for the full success of ignoring unsolicited cues.
Ignoring Contradictory Positive Information

The results of H1B, H2B, and H3B indicated that the manipulations of unsolicited cues may have been flawed and that neutral unsolicited cues unexpectedly increased ratings of the focal trait. H1B, H2B, and H3B predicted that exposure to positive unsolicited cues would increase ratings of a focal trait when solicited cues were negative. H1B predicted that positive unsolicited cues would inflate ratings of a focal trait when negative solicited cues were assessed with dimensional scoring—it was expected that the mean of Cell 7 in Table 2 would be higher than the mean of Cell 8, and higher than the mean of Cell 9. H1B was not supported because positive unsolicited cues did not result in ratings that were higher ($M = 3.64$, $SD = 1.68$) than those obtained in neutral ($M = 4.43$, $SD = 1.65$) and negative ($M = 2.86$, $SD = 1.83$) unsolicited cue conditions when negative solicited cues were assessed with dimensional scoring (see cluster on the left of Figure 3). The rejection of H1B indicates that positive unsolicited cues lacked sufficient salience to inflate ratings of the focal trait.

H2B and H3B predicted that the ratings of interviewers in question-based scoring conditions would be inflated by positive unsolicited cues, despite instructions to ignore unsolicited cues. The test of H2B was a comparison between the ratings of interviewers presented with positive unsolicited cues and interviewers not presented with such cues. The test of H3B was a comparison of the effects of positive unsolicited cues between the ratings of interviews permitted to use such cues (dimensional scoring condition) and interviewers not permitted to consider them (question-based scoring condition). Unfortunately, because the test of H1B indicated that positive unsolicited cues were essentially unnoticed by interviewers, the tests of H2B and H3B are not valid tests of whether or not interviewers can ignore positive unsolicited cues. Because participants did not perceive positive unsolicited cues as pertinent, any attempts to test if participants could ignore these cues were futile—knowing about pertinent information and attempting to ignore it is not the same as failing to recognize its pertinence to a decision (or, in the case of an experiment, failing to develop salient and compelling experimental manipulations).
Participants’ failure to recognize the relevance of positive unsolicited cues may have resulted from a) insufficient weighting on positive cues, b) inexperience assessing sub-facets of emotional stability, or c) difficulty differentiating moderately from highly positive information. A substantial body of research has shown that negative information influences interviewers’ decisions more heavily than does positive information (Knouse, 1983; Schmitt, 1976; Springbett, 1958; Webster, 1964). Hollmann (1972) tested if the disproportionate weighting of negative information on interviewers’ decisions was linked to interviewers’ ability to recognize and evaluate positive and negative information. Results showed that interviewers recognized and evaluated positive and negative information equally well, however, they tended to overlook the importance of positive information when assessing interview constructs. These results suggest that, even if participants in this study recognized and accurately evaluated positive unsolicited cues, they may have failed to integrate this information when rating the focal trait.

Participants’ inexperience rating cues indicative of high stress tolerance may have impaired their ability to recognize that positive unsolicited stress tolerance cues were pertinent to stress tolerance, the focal trait assessed with Question #2. Emotional stability is the over-arching FFM trait for stress tolerance. Although emotional stability is a valid predictor of performance, it has been shown to be the most difficult FFM trait to assess in an interview (Oh, Wang, & Mount, 2011). Although negative stress tolerance cues are probably easier to recognize, the difference between moderately and highly positive stress tolerance cues is likely context-dependent. For example, the positive stress tolerance cues included the following behaviours that could be considered normal requirements of any job: “I focused on the most difficult aspects first and I made a few suggestions […] about how we could improve…”. This exact quotation could represent average or above-average stress tolerance, depending on the level of adversity faced by the employee and the employee’s specific actions and suggestions. Therefore, I suggest that positive unsolicited stress tolerance cues may not have shown the desired effects due to participant’s lack of experience differentiating moderate and high levels of stress tolerance.

Based on previous findings pertaining to the effects of positive information in interview settings, I suspect that positive unsolicited cues failed to influence judgments because of a disparity in the scale of positive and negative cues and due to a lack of interviewer experience.
According to Johnson (1994), personality cues can be classified according to valence and scale. The valence is either positive or negative and it determines whether a cue will increase or decrease the rating of a trait. In addition to having a valence, a cue can be ranked on scale according to the magnitude of its influence on the rating of a trait.

I propose that, relative to other cues, the scale of positive unsolicited cues may have been disproportionately smaller than the scale of negative solicited cues. Imagine that an interviewer must assess a trait with knowledge of only two personality cues, one positive and one negative. If valence of positive and negative cues was +1 and -20, respectively, it is probable that the positive cue would have very little influence on the interviewers rating of the trait. Whereas the positive unsolicited cues could be interpreted as meeting -but not exceeding a standard of performance (i.e., “I focused on the most difficult aspects first”), the scale of negative solicited cues arguably depicted extremely low levels of emotional stability (i.e., “constantly panicked”). Therefore, the disproportionate scale of negative solicited cues relative to positive unsolicited cues may have caused the effect of negative solicited cues to completely overshadow the effect of positive unsolicited cues.

**Unexpected Effects of “Neutral” Information**

When solicited cues were negative, there are three factors that may have caused neutral unsolicited cues to increase ratings of a focal trait. First, research in tribunal proceedings has shown that weak evidence is sometimes harder to ignore than strong evidence; therefore, participants may have had more difficulty ignoring neutral unsolicited cues if these were perceived to be mildly positive. Second, these cues may have made novice interviewers averse to giving negatively critical ratings. Each of these two factors will be explained below.

Post-hoc analyses indicated that, despite instructions to ignore unsolicited cues (given in question-based scoring conditions), neutral unsolicited cues resulted in higher trait ratings \((M = 3.67, SD = 1.41)\) than did positive unsolicited cues \((M = 2.75, SD = 1.41)\) and negative unsolicited cues \((M = 2.25, SD = 1.07)\); this effect was moderate for positive unsolicited cues \((d = .65, 95\% \text{ CI } [-0.01, 1.30])\) and very large for negative unsolicited cues \((d = 1.15, 95\% \text{ CI } [0.48, 1.80])\). A post-hoc comparison of trait ratings between dimensional \((M = 4.43, SD = 1.65)\) and
question-based \((M = 3.67, \ SD = 1.41)\) scoring protocol conditions indicated that instructions to ignore neutral unsolicited cues were ineffective when solicited cues were negative.

The first potential explanation for the effect of neutral unsolicited cues is that they provided weak evidence, which is sometimes more difficult to ignore. In a study by Schul and Goren (1997), mock jurors were presented with testimony that was strongly or weakly incriminating and subsequently instructed by the judge to ignore it; whereas the ratings of guilt made by jurors presented with strong testimony resembled those of jurors in a control condition not presented with inadmissible testimony, the ratings of jurors presented with weak evidence and instructed to ignore it resembled those of jurors presented with strong evidence and allowed to consider it. Relative to jurors presented with strong inadmissible evidence, jurors presented with weak inadmissible evidence were unaware of its influence on their ratings of guilt. The authors concluded that the influence of weak evidence on judgments was enhanced because jurors failed to realize the need to correct for its influence. If neutral unsolicited cues were weak cues, participants’ reduced awareness of their influence on judgments may have impaired their ability to ignore them.

The second potential explanation for the positive effect of neutral unsolicited cues is that they may have undermined novice interviewers’ task-related self-efficacy, making them reluctant to give negatively critical ratings. Neutral unsolicited cues were scripted to be devoid of specific behaviours and described a collective effort to achieve a goal without appropriate benchmarks of success (i.e., “We did as much as we could to complete the order as quickly as possible”). In contrast to the four behaviourally-rich practice exercises, the neutral unsolicited cues were not ratable in accordance with the tutorial, which instructed participants to assess past behaviours. The contrasting difficulty between assessing practice exercises and a “real” applicant may have shaken participants’ self-efficacy at rating traits. Given that lowered rater self-efficacy has been shown to increase rater leniency (Bernardin & Villanova, 2005), I proposed that neutral unsolicited cues diminished rater self-efficacy and caused participants to inflate trait ratings.
Practical Implications

Results indicated that unsolicited cues can be ignored in some circumstances. Although further evidence is required, findings that interviewers can ignore unsolicited cues could have three primary implications for HR managers and I/O psychologists. First, training must teach interviewers to discern and address responses to questions that lack sufficient detail to be assessed. Second, question-based and dimensional scoring protocols may result in different rank-ordering of candidates. Third, overly rigid adherence to question-based scoring is likely inadvisable.

All valid a priori tests of the hypotheses in this study indicated that unsolicited cues can be ignored; in question-based scoring conditions positive unsolicited cues had no effect on trait ratings when solicited cues were negative and negative unsolicited cues had no effect on trait ratings when solicited cues were positive. However, the ambiguous statements provided in the neutral unsolicited cues seem to have flustered participants’ rater self-efficacy, resulting in more lenient ratings of the focal trait. Although previous research indicates that interviewer leniency declines with experience (Furnham & Burbeck, 1989; Rowe, 1963; Wiener and Schneiderman, 1974), training that would help novice interviewers to recognize and address causes of rating leniency, including difficult-to-assess responses, may accelerate the development of interviewers’ rating accuracy.

Given that unsolicited cues can be ignored when applying question-based scoring, interview practitioners must be cognizant that scoring protocols may affect the rank-ordering of applicants. The difference in ratings between question-based ($M = 8.53$) and dimensional ($M = 6.5$) scoring protocols when the applicant provided negative unsolicited cues and positive solicited cues illustrates how scoring protocols could change a candidate’s fate in a selection process; in this scenario dimensional scoring could plausibly lower the applicant’s ranking such that she would be eliminated from the hiring process. Although dimensional scoring may facilitate the differentiation of applicants because it would allow interviewers to consider applicants’ behaviour in different contexts when rating a personality trait, Conway, Jako, and Goodman (1995) warn that the consideration of unsolicited cues allowed in dimensional scoring may cause interviewers to consider irrelevant information:
When multiple-dimension ratings are being made, the relevance of specific answers to each rating is likely to be much less clear. This leaves more opportunity for a rater to fall prey to cognitive biases, such as reliance on an overall impression. (p. 568)

Given that properly developed structured interview questions are specifically designed and tested to illicit responses that pertain to a specific construct and that interviewers can effectively ignore unsolicited cues, it is recommend that interview practitioners use question-based scoring as a default. However, unsolicited cues should not be ignored in all cases. For example, in the negative unsolicited cue condition, the applicant disclosed severe anxiety and extremely poor coping strategies in the workplace. Although Figure 1 shows that interviewers were able to completely overlook these cues in the question-based scoring condition, doing so would likely result in a poor hiring decision with unfavourable outcomes for the organization. Therefore, depending on the training, experience, and skills of interviewers within an organization, it is recommended that dimensional scoring be allowed in cases that unsolicited information clearly pertains to a construct that is to be assessed with a different interview question.

Limitations

The greatest limitation of this study is that some conditions failed to test the effects of unsolicited cues on ratings of a focal trait when solicited cues were negative. Specifically, when solicited cues were negative, neutral unsolicited cues resulted in an unexplained leniency effect and participants failed to perceive the relevance of positive unsolicited cues to the focal trait. A further limitation of these results is that seven out of 12 conditions had 20 or fewer participants (including both conditions that combined neutral unsolicited cues and negative solicited cues); although outliers did not skew any of the cases, more participants may have provided more reliable results.

Participants’ ability to ignore unsolicited cues may have been enhanced by experimental factors that enhanced memory and recall of the applicant’s responses to interview questions. The interview conditions in this study were far removed from real interviews in several ways. First,
most real interviews include face-to-face exposure to non-verbal cues and two-way dialogue, which arguably increase the cognitive load of the interviewer. Second, responses to interview questions in this study, which were limited to three to five sentences, were probably much shorter than real responses to interview questions. Third, whereas the unsolicited cues in a real interview would be intertwined with solicited cues, the applicant’s responses to Question #1 contained no solicited cues for achievement striving. Fourth, participants were able to listen to recordings of responses to interview questions as many times as they wished prior to rating them. Given that it is easier to ignore information that can be recalled into memory (Bjork & Bjork, 2003; Broeder, 1959; Johnson, 1994; Storm, Bjork, & Bjork, 2005) and harder to ignore information when cognitive load is increased (Kulik, Perry, & Bourhis, 2000), participants’ ability to ignore unsolicited cues may have been enhanced by multiple factors that enhanced participants’ ability to recall the applicant’s responses.

The generalizability of this study is limited by the absence of a high-stakes situation. Although the participant sample closely matched US demographics, these participants were not professional interviewers assessing a real candidate in a high-stakes situation. Rather, these were crowd-sourced workers, the majority of which participated in this task because “it’s a great way to spend free time and get some cash”. Although I tried to replicate the sense of accountability felt by a hiring manager with a deception that falsely informed candidates that their ratings would be used to hire US government employees, adding that participants’ ratings would be averaged with those of other micro-workers was necessary to make the deception plausible. Unfortunately this inevitably diffused participants’ individual responsibility down to a negligible fraction of that felt by a hiring manager in a real selection setting.

The potentially largest limitation to the external validity of this study is that novice interviewers and mock-interviewers assess applicants more leniently than do professional interviewers (Barr and Hitt; 1986, Dipboye, Fromkin, & Wiback, 1975). This assertion is anecdotally supported by comments from focus groups conducted after pilot study sessions in which participants indicated they felt bad rating the applicant unfavourably. Given that negative information tends to affect the decisions of professional interviewers more than those of novice interviewers and mock-interviewers, (Landy & Bates, 1973, Hakel, Ohnesorge, & Dunnette,
1970; Bernstein, Hakel, & Harlan, 1975), the results of this study, including the inflation of ratings caused by neutral unsolicited cues, may not have occurred with a sample of professional interviewers.

**Conclusion**

This study revealed two primary findings. First, it was found that when solicited cues were positive, the influence of negative information could be eliminated with the use of question-based scoring. Second, when solicited cues were negative, neutral unsolicited cues caused a leniency effect of ratings of the focal trait. Although the tests of H1A, H2A, and H3A indicated that unsolicited cues could be ignored when solicited cues were positive, the unexplained effect of neutral unsolicited cues suggests that unsolicited cues cannot be ignored in all cases.

Although methodological corrections could help clarify the conditions in which unsolicited cues influence trait ratings, this line of study should be pursued *after* an examination of whether ignoring unsolicited cues, as prescribed by question-based scoring, leads to greater predictive validity of trait ratings. Although unsolicited cues may offer a rich source of information that could enhance the accuracy of FFM trait ratings, it may be best to limit the amount of information interviewers process by using question-based scoring protocols. Although dimensional scoring may allow more opportunities to gather trait-related information, the predictive validity of trait ratings may plummet if interviewers mistakenly incorporate unsolicited cues that are irrelevant to a focal trait into their evaluations. Future research using real interviewers and real applicants applying for a real job should compare the predictive validity of question-based scoring versus dimensional scoring.

Depending on the effects of question-based and dimensional scoring on the predictive validity of personality trait ratings, it is recommended that subsequent research explore motivational factors on interviewers’ ability to ignore trait-related information. Given that numerous studies have demonstrated interviewers’ reluctance to use higher levels of interview structure (Jelf, 1999; Lievens & De Paepe, 2004), interviewers’ ability to ignore or incorporate unsolicited cues into their decision making may be irrelevant if they lack the motivation to adhere either question-based or dimensional scoring protocols. Although results from this study
suggest that question-based scoring can eliminate the effects of unsolicited cues, future research should investigate if consideration of unsolicited helps or hinders the predictive validity of structured employment interviews.
References


Tucker, I. H. & Rowe, P. M. Relationship between expectancy, causal attributions, and final hiring decisions in the employment interview. *Journal of Applied Psychology, 1979, 64*, 27-34.


### Table 1

*Study Design: Scoring protocol and Presentation of Stress Tolerance Cues*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensional</td>
<td>Positive: Cell 1</td>
<td>Negative: <strong>Cell 2</strong></td>
<td>Neutral: Cell 3</td>
</tr>
<tr>
<td>Question-Based</td>
<td>Positive: Cell 4</td>
<td>Negative: <strong>Cell 5</strong></td>
<td>Neutral: Cell 6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scoring Protocol</th>
<th>Unsolicited Cues: Negative Valence</th>
<th>Unsolicited Cues: Neutral Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensional</td>
<td>Positive: <strong>Cell 7</strong></td>
<td>Negative: Cell 8</td>
</tr>
<tr>
<td>Question-Based</td>
<td>Positive: <strong>Cell 10</strong></td>
<td>Negative: Cell 11</td>
</tr>
</tbody>
</table>

*Note:* Conditions written in bold font presented contradictory cues. This table illustrates a between subjects 2 (scoring protocol: question-based, dimensional) x 3 (unsolicited cues: positive, negative, neutral) x 2 (solicited cues: positive, negative) fixed factorial design.
Table 2

Results: Scoring protocol and Presentation of Stress Tolerance Cues

<table>
<thead>
<tr>
<th>Solicited Cues: Positive Valence</th>
<th>Unsolicited Cue Valence</th>
<th>Solicited Cues: Negative Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>Cell 1</strong></td>
<td>(M = 8.78, SD = 1.11, N = 18)</td>
<td><strong>Cell 2</strong></td>
</tr>
<tr>
<td><strong>Cell 4</strong></td>
<td>(M = 8.73, SD = 0.99, N = 20)</td>
<td><strong>Cell 5</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Solicited Cues: Negative Valence</th>
<th>Unsolicited Cue Valence</th>
<th>Solicited Cues: Negative Valence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Positive</td>
<td>Negative</td>
</tr>
<tr>
<td><strong>Cell 7</strong></td>
<td>(M = 3.64, SD = 1.68, N = 25)</td>
<td><strong>Cell 8</strong></td>
</tr>
<tr>
<td><strong>Cell 10</strong></td>
<td>(M = 2.75, SD = 1.41, N = 20)</td>
<td><strong>Cell 11</strong></td>
</tr>
</tbody>
</table>

*Conditions written in bold font presented contradictory cues. This table illustrates a between subjects 2 (scoring protocol: question-based, dimensional) x 3 (unsolicited cues: positive, negative, neutral) x 2 (solicited cues: positive, negative) fixed factorial design.*
**Scoring Protocol**

*Figure 1.* Test of H1A and H2A: Effect of Unsolicited Cues When Solicited Cues Are Positive. UC = Unsolicited Cues; SC = Solicited Cues.
Figure 2. Test of H3A: Scoring Protocol Negative With Positive Solicited Cues.

Order of Presentation of Contradictory Cues
Figure 3. Test of H1B and H2B: Effect of Unsolicited Cues When Solicited Cues Are Negative.

UC = Unsolicited Cues; SC = Solicited Cues.

Scoring Protocol
Figure 4. Test of H3B: Scoring Protocol With Negative Solicited Cues.
Figure 5. Effect of Neutral Unsolicited Cues Across Scoring Protocol Conditions.
Appendix A: Rating Scale For Achievement Striving

Interview Like a Pro PART ONE

SOPHIA MOORE:

ACHIEVEMENT STRIVING

A person high in achievement striving:

Gives effort to perform well.
Approaches work with persistence, consistency, and punctuality.
Strives for excellence.
Recognized as successful.
Stays on track to achieve goals.

Please rate the applicant's ACHIEVEMENT STRIVING from 1 to 10 where:

<table>
<thead>
<tr>
<th>Very Low</th>
<th>Typical</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B: Rating Scale for Stress Tolerance

Interview Like a Pro PART ONE

SOPHIA MOORE:

STRESS TOLERANCE

A person high in stress tolerance:

Shows a calm, relaxed approach to situations, events, or people.
Shows emotionally controlled response to change.
Remains tolerant and optimistic under stress.
Deals with an issue as a challenge rather than crisis.
Maintains stable performance under pressure or opposition.
Handles stress in a manner that is acceptable to others and to the organization.

Please rate the applicant's STRESS TOLERANCE from 1 to 10 where:

<table>
<thead>
<tr>
<th>Very Low</th>
<th>Typical</th>
<th>Very High</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>


Appendix C: Threshold Test Item 1 – High Achievement Striving

This is the script for the Practice Exercise 1 – High Achievement Striving. The applicant’s response to Question #1 included cues that reflected a high level of achievement striving which are written in bold and underlined font.

QUESTION #1

INTERVIEWER: “GIVE ME A SPECIFIC EXAMPLE OF A TIME WHERE YOU DID NOT MEET A DEADLINE. HOW DID YOU HANDLE IT?”

APPLICANT: “A few months ago I was working as a construction project manager to build a house for a new client –it was our biggest contract. Unfortunately there was a delay getting building permits –which you need to have before you start construction. In order to get the project back on track, I had to work much harder and put in longer hours. I scheduled and inspected everything very carefully –this helped us to catch up and it ensured that we delivered a much higher quality product on time. The client was extremely grateful for the extra effort we put in to have their home ready in a timely manner.”
Appendix D: Threshold Test Item 2 – Low Achievement Striving

This is the script for the Practice Exercise 2 – Low Achievement Striving. The applicant’s response to Question #1 included cues that reflected a low level of achievement striving which are written in **bold and underlined font**.

QUESTION #1

INTERVIEWER: “GIVE ME A SPECIFIC EXAMPLE OF A TIME WHERE YOU DID NOT MEET A DEADLINE. HOW DID YOU HANDLE IT?”

APPLICANT: “A few months ago my boss asked me to present the summary of our department’s expenditures to the executive team. I thought it seemed like I had plenty of time to get the job done so I didn’t REALLY start working on it until much later. When I finally got around to it, I realized that it was a big mistake to leave everything until the last minute. Because I failed to start on the budget report sooner, my final report was missing a lot of the necessary data. My boss warned me to clean up my act or else he’d fire me.”
Appendix E: Threshold Test Item 3 – High Stress Tolerance

This is the script for the Practice Exercise 3 – High Stress Tolerance. The applicant’s response to Question #2 included cues that reflected a high level of stress tolerance which are written in **bold and underlined font**.

**QUESTION #2**

**INTERVIEWER:** “TELL ME ABOUT A TIME WHEN YOU WERE SO BUSY THAT YOU WERE PUSHED YOU TO YOUR LIMITS.”

**APPLICANT:** “A few months ago, two of my co-workers went on stress leave which left me responsible for doing the work of three people. My workdays were non-stop but I didn’t feel stressed. The extra workload was just about as much as I could manage but I dealt with everything in stride. I focused only on the things that I had control over and I found the best strategy was to keep a cool head. My boss was very pleased with my ability to keep up with the increased workload and manage my stress level professionally.”
Appendix F: Threshold Test Item 4 – Low Stress Tolerance

This is the script for the Practice Exercise 3 – Low Stress Tolerance. The applicant’s response to Question #2 included cues that reflect a low level of stress tolerance which are written in **bold and underlined font**.

QUESTION #2

INTERVIEWER:  “TELL ME ABOUT A TIME WHEN YOU WERE SO BUSY THAT YOU WERE PUSHED YOU TO YOUR LIMITS.”

APPLICANT: “Over the summer I had a retail job. My manager asked me to work longer hours and I felt pressured to work more quickly -which was REALLY stressful for me, because I’m not someone who performs well under pressure. He also told me that the store might be moving across the street –which meant even MORE work; I was TOTALLY overwhelmed. My manager said that he could tell that I was stressed and he tried to help me out but he ended up hiring extra employees and I was let go a few weeks later –I guess he felt the pressure was just too much for me and I couldn’t hack it.”
### Appendix G: Summary of Manipulation Checks in Post-Experimental Questionnaire

<table>
<thead>
<tr>
<th>Manipulation Check</th>
<th>Question</th>
<th>Response Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likeability of the job applicant</td>
<td>To what extent do you think the applicant is likeable or dislikeable?</td>
<td>Likert Scale: 1-5</td>
</tr>
<tr>
<td>Rating of Suitability</td>
<td>To what extent do you feel the applicant would be suitable for a job?</td>
<td>Likert Scale: 1-5</td>
</tr>
<tr>
<td></td>
<td>To what extent do you feel the applicant would be suitable for a job that requires high achievement striving and high stress tolerance? (last question of survey)</td>
<td>Likert Scale: 1-5</td>
</tr>
<tr>
<td>Employability</td>
<td>Would you hire this applicant?</td>
<td>Yes/No</td>
</tr>
<tr>
<td></td>
<td>Please list in point form all information about the applicant that pertains to the applicant’s achievement striving. Please consider all information that you know about the applicant when listing your points. The information you provide may be from the applicant’s behaviours or comments observed throughout the entire interview.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please list in point form all information about the applicant that pertains to the applicant’s stress tolerance. Please consider all information that you know about the applicant when listing your points. The information you provide may be from the applicant’s behaviours or comments observed throughout the entire interview.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please type a dollar sign ($) beside any information that you considered when rating the applicant’s achievement striving.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please type a dollar sign ($) beside any information that you considered when rating the applicant’s stress tolerance.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please type a hashtag or number sign (#) beside all information that you were instructed to consider when rating the applicant’s achievement striving.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please type a hashtag or number sign (#) beside all information that you were instructed to consider when rating the applicant’s stress tolerance.</td>
<td>Free recall – typed</td>
</tr>
</tbody>
</table>
## Appendix G: Summary of Manipulation Checks in Post-Experimental Questionnaire

<table>
<thead>
<tr>
<th>Manipulation Check</th>
<th>Question</th>
<th>Response Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Willful adherence to scoring protocol</td>
<td>Please type a capital X beside all information you feel that you should have been allowed to consider when rating the applicant’s achievement striving.</td>
<td>Free recall</td>
</tr>
<tr>
<td></td>
<td>Please type a capital X beside all information you feel that you should have been allowed to consider when rating the applicant’s stress tolerance.</td>
<td>- typed</td>
</tr>
<tr>
<td></td>
<td>Please type the @ symbol beside all information you feel influenced your rating of the applicant’s achievement striving.</td>
<td>Free recall</td>
</tr>
<tr>
<td></td>
<td>Please type the @ symbol beside all information you feel influenced your rating of the applicant’s stress tolerance.</td>
<td>- typed</td>
</tr>
<tr>
<td></td>
<td>In point form, what information had the greatest influence on your rating of the applicant’s achievement striving?</td>
<td>Free recall</td>
</tr>
<tr>
<td></td>
<td>In point form, what information had the greatest influence on your rating of the applicant’s stress tolerance?</td>
<td>- typed</td>
</tr>
<tr>
<td>Influence of scoring protocol on decision-making</td>
<td>Briefly describe your thoughts or feelings about rating the applicant’s achievement striving.</td>
<td>Free recall</td>
</tr>
<tr>
<td></td>
<td>Briefly describe your thoughts or feelings about rating the applicant’s stress tolerance.</td>
<td>- typed</td>
</tr>
<tr>
<td></td>
<td>Think back to the moment that you were assigning a number to rate the applicant’s achievement striving. Was this a simple task? If yes, why? If not, why not?</td>
<td>Free recall</td>
</tr>
<tr>
<td></td>
<td>Think back to the moment that you were assigning a number to rate the applicant’s stress tolerance. Was this a simple task? If yes, why? If not, why not?</td>
<td>- typed</td>
</tr>
<tr>
<td></td>
<td>Thinking back to the moment that you were assigning a number to rate the applicant’s achievement striving. Was there anything about coming up with this number that you didn’t expect? If yes, what?</td>
<td>Free recall</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- typed</td>
</tr>
<tr>
<td>Manipulation Check</td>
<td>Question</td>
<td>Response Format</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td><strong>Influence of scoring protocol on decision-making</strong></td>
<td>Thinking back to the moment that you were assigning a number to rate the applicant’s stress tolerance. Was there anything about coming up with this number that you didn’t expect? If yes, what?</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Briefly, describe your decision making process when rating the applicant’s stress tolerance.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td><strong>Resistance to scoring protocol</strong></td>
<td>A person who is suitable for a job has what it takes to be good in that job. Do you think that your rating of the applicant’s achievement striving is a good reflection of how suitable he or she is for the job? If yes, why? If not, why not?</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>A person who is suitable for a job has what it takes to be good in that job. Do you think that your rating of the applicant’s stress tolerance is a good reflection of how suitable he or she is for the job? If yes, why? If not, why not?</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please mark an asterisk or star (*) beside all information that you would have considered when rating the applicant’s achievement striving if it was up to you.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Please mark an asterisk or star (*) beside all information that you would have considered when rating the applicant’s stress tolerance if it was up to you.</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Briefly, what are the benefits and disadvantages of using all available information when rating a personality trait?</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td></td>
<td>Did you disregard any of the instructions provided in the training you received when rating any of the applicant’s personality traits?</td>
<td>Yes/No</td>
</tr>
<tr>
<td></td>
<td>Were you tempted to disregard any of the instructions provided in the training you received when rating any of the applicant’s personality traits?</td>
<td>Yes/No</td>
</tr>
</tbody>
</table>
### Appendix G: Summary of Manipulation Checks in Post-Experimental Questionnaire

<table>
<thead>
<tr>
<th>Manipulation Check</th>
<th>Question</th>
<th>Response Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resistance to scoring protocol</td>
<td>Briefly, what are the benefits and disadvantages of using only information provided by the applicant in direct response to a specific question when rating the personality trait?</td>
<td>Free recall – typed</td>
</tr>
<tr>
<td>Scoring Method Preference</td>
<td>Please select the answer that best describes your opinion: a) It is better to consider only information provided by the applicant in direct response to a specific question when rating a personality trait. b) It is better to use all available information provided by the applicant when rating a personality trait.</td>
<td>Multiple choice</td>
</tr>
</tbody>
</table>
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

Please select the answer that best describes your opinion:

1.1) To what extent do you think the applicant is likeable or dislikeable?

<table>
<thead>
<tr>
<th>Scale</th>
<th>Very dislikeable</th>
<th>Neither likeable nor dislikeable</th>
<th>Very likeable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

1.2) To what extent do you feel the applicant would be suitable for a job?

<table>
<thead>
<tr>
<th>Scale</th>
<th>Very unsuitable</th>
<th>Neither suitable nor unsuitable</th>
<th>Very suitable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

1.3) Would you hire this applicant? **YES / NO**

Once you have completed all questions on this page, you will not be permitted to return and/or change your answers.
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

2.1) Please list in point form all information about the applicant that pertains to the applicant’s achievement striving. Please consider all information that you know about the applicant when listing your points. The information you provide may be from the applicant’s behaviours or comments observed throughout the entire interview.

2.2) Please type a dollar sign ($) beside any information that you considered when rating the applicant’s achievement striving.

2.3) Please type a hashtag or number sign (#) beside all information that you were instructed to consider when rating the applicant’s achievement striving.

2.4) Please type an X beside all information you feel that you should have been allowed to consider when rating the applicant’s achievement striving.

2.5) Please mark an asterisk/star (*) beside all information that you would have considered when rating the applicant’s achievement striving if it was up to you.

2.6) Please type the @ symbol beside all information you feel influenced your rating of the applicant’s achievement striving.

Once you have completed all questions on this page, you will not be permitted to return and/or change your answers.
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

3.1) In point form, what information had the greatest influence on your rating of the applicant’s achievement striving?

3.2) Briefly describe your thoughts or feelings about rating the applicant’s achievement striving.

3.3) Think back to the moment that you were assigning a number to rate the applicant’s achievement striving. Was this a simple task? If yes, why? If not, why not?

3.4) Thinking back to the moment that you were assigning a number to rate the applicant’s achievement striving. Was there anything about coming up with this number that you didn’t expect? If yes, what?

3.5) Briefly, describe your decision making process when rating the applicant’s achievement striving.

Once you have completed all questions on this page, you will not be permitted to return and/or change your answers.
A person who is suitable for a job has what it takes to be good in that job. Do you think that your rating of the applicant’s achievement striving is a good reflection of how suitable he or she is for the job? If yes, why? If not, why not?
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

4.1) Please list in point form all information about the applicant that pertains to the applicant’s stress tolerance. Please consider all information that you know about the applicant when listing your points. The information you provide may be from the applicant’s behaviours or comments observed throughout the entire interview.

4.2) Please type a dollar sign ($) beside any information that you considered when rating the applicant’s stress tolerance.

4.3) Please type a hashtag or number sign (#) beside all information that you were instructed to consider when rating the applicant’s stress tolerance.

4.4) Please type a capital X beside all information you feel that you should have been allowed to consider when rating the applicant’s stress tolerance.

4.5) Please mark an asterisk or star (*) beside all information that you would have considered when rating the applicant’s stress tolerance if it was up to you.

4.6) Please type the @ symbol beside all information you feel influenced your rating of the applicant’s stress tolerance.

Once you have completed all questions on this page, you will not be permitted to return and/or change your answers.
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

5.1) In point form, what information had the greatest influence on your rating of the applicant’s stress tolerance?

5.2) Briefly describe your thoughts or feelings about rating the applicant’s stress tolerance.

5.3) Think back to the moment that you were assigning a number to rate the applicant’s stress tolerance. Was this a simple task? If yes, why? If not, why not?

5.4) Thinking back to the moment that you were assigning a number to rate the applicant’s stress tolerance. Was there anything about coming up with this number that you didn’t expect? If yes, what?

Once you have completed all questions on this page, you will not be permitted to return and/or change your answers.
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

6.1) Briefly, describe your decision making process when rating the applicant’s stress tolerance.

6.2) A person who is suitable for a job has what it takes to be good in that job. Do you think that your rating of the applicant’s stress tolerance is a good reflection of how suitable he or she is for the job? If yes, why? If not, why not?

Once you have completed all questions on this page, you will not be permitted to return and/or change your answers.
Appendix H: Post-Experimental Questionnaire

PLEASE RESPOND HONESTLY TO ALL OF THE FOLLOWING QUESTIONS IN THE ORDER THAT THEY ARE PRESENTED. AS YOU ANSWER EACH QUESTION ONE AT A TIME, PLEASE DO NOT GO BACK TO ANY PREVIOUS QUESTIONS. IT IS IMPORTANT THAT YOU DO NOT SKIP AHEAD OR READ ANY OF THE LATER QUESTIONS BEFORE FULLY COMPLETING THE PREVIOUS ONES.

Please respond to the following questions by clicking YES or NO or by filling in the blank:

7.1) Briefly, what are the benefits and disadvantages of using only information provided by the applicant in direct response to a specific question when rating the personality trait?

7.2) Briefly, what are the benefits and disadvantages of using all available information when rating a personality trait?

7.3) Please select the answer that best describes your opinion:
   a) It is better to consider only information provided by the applicant in direct response to a specific question when rating a personality trait.
   b) It is better to use all available information provided by the applicant when rating a personality trait.

7.4) Did you disregard any of the instructions provided in the training you received when rating any of the applicant’s personality traits? YES / NO

7.5) If Yes, which instructions?
Appendix H: Post-Experimental Questionnaire

8.1) Were you tempted to disregard any of the instructions provided in the training you received when rating any of the applicant’s personality traits? **YES / NO**

8.2) If Yes, which instructions?

8.3) To what extent do you feel the applicant would be suitable for a job that requires high achievement striving and high stress tolerance?
Appendix I: Initial Description of the Interview Like a Pro CrowdFlower Task

$50 Usd Interview Like A Pro

Instructions

Interview Like a Pro Part ONE
You will earn $50 USD in approximately 12 minutes:
- 7 minutes to view a audio-visual tutorial
- 4 minutes to assess an actual job applicant
- 1 minute to redeem payment

This job MUST be completed ALL AT ONCE WITHOUT INTERRUPTION.

Your responses will help assess applicants who are applying to work within the United States government.

This job must be completed from a DESKTOP or LAPTOP computer - it cannot be completed using a smartphone.

To complete this task, you will require the following:
1- A HIGH-SPEED Internet connection
2- SOUND capabilities on your computer
3- ADOBE FLASH PLAYER on your computer
4 - COOKIES MUST BE ENABLED on your computer

In order to complete this task, you must go to the URL below and complete the following steps:

1) Watch and listen to an audio-tutorial and learn how to assess personality traits in a job interview.
2) Practice your skills on four applicants.
3) If you demonstrate the ability to accurately assess personality traits you will assess actual job applicants in the Interview Like a Pro Part 2 task.

Click Here to go to the Interview Like a Pro website

Text

Enter here...

Enter Survey code in this field after completing
Appendix J: Initial Letter of Information and Consent with Deception

Information Letter

Interview Like a Pro

You are asked to participate in the Interview Like a Pro hiring project of employees in the U.S. federal government.

If you have any questions or concerns about the project, please feel free to contact Paul Comeault 226-821-3635 (pacomeault@gmail.com) or Dr. Deborah Powell 519 824 4120 x 52167 (d.powell.interviewlikeapro@gmail.com).

PURPOSE OF THE PROJECT

The purpose of this project is to determine if ratings of an applicant’s ability to respond to job interview questions obtained using a web-based crowd sourcing technique can accurately predict job performance. Crowdsourcing is the process of soliciting judgements from a large number of internet users. Specifically, we will compare the performance of job applicants selected through crowd-sourcing to the performance of applicants hired using traditional one-on-one job interviews.

PROCEDURES

If you volunteer to participate in this project, we would ask you to do the following things:

1. We will ask you to check the box indicating that you consent to participate in this project.

2. We will ask you to listen to an audio tutorial in which you will be instructed how to apply a professional interview method known as a structured interview.

3. You will then listen to four applicant’s responses to interview questions and practice applying the rating system taught to you.

4. You will then listen to audio-recorded responses to interview questions made by a job applicant. You will then score that applicant’s responses using the rating system taught to you.

POTENTIAL RISKS AND DISCOMFORTS

Although we will take extensive precautions to ensure your confidentiality is maintained, there is a minor risk that your confidentiality could be violated. For an outline of measure that will be taken to ensure that your confidentiality is maintained, please see subsection “Confidentiality”.

69
Appendix J: Initial Letter of Information and Consent with Deception

POTENTIAL BENEFITS TO PARTICIPANTS AND/OR TO SOCIETY

Benefits to participants: Participants will learn about the following principles of professional job interviews:

1) Personality traits predict job performance.
2) A personality trait can be good to have in one job but very bad to have in a different job.
3) Professional interviews focus on past behaviour.
4) Scores are more accurate when a job applicant is assessed by a greater number of interviewers.

Benefits to the scientific community and society: this project will help practitioners in Human Resource Management to better understand the merits of modernized interview techniques.

CREDIT FOR PARTICIPATION

You will be paid $0.50 USD for your participation in the first phase of this project and an additional $0.50 USD for your participation in the second phase of this project.

CONFIDENTIALITY

Every effort will be made to ensure confidentiality of any identifying information that is obtained in connection with this project.

The ratings you will be asked to provide will be linked to the task authentication code given to you by Crowdflower and the task completion code given to you at the end of this task. Only Crowdflower will have the ability to match these codes with your Crowdflower worker ID. Crowdflower will not have access to the responses that you make throughout your participation in this project but they will be informed about the quality of the work that you complete. Members of the Interview Like a Pro project team will have no way of linking your responses to your Crowdflower worker ID or any other personal identifying information. Please note that confidentiality cannot be guaranteed while data are in transit over the internet. However, your responses will be encrypted using the same encryption type (SSL) that is used by online banking sites to transmit secure information.

PARTICIPATION AND WITHDRAWAL

You can choose whether to be in this project or not. If you volunteer to be in this project, you may withdraw at any time without consequences of any kind. You may exercise the option of removing your data from the project. You may also refuse to answer any questions you don’t want to answer and still remain in the project. The investigator may withdraw you from this project if circumstances arise that warrant doing so.
Appendix J: Initial Letter of Information and Consent with Deception

RIGHTS OF PROJECT PARTICIPANTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this project. This project has been reviewed and received clearance through a Project Ethics Board that adheres to the American Psychological Association guidelines for ethical project with human participants. If you have questions regarding your rights as a project participant, contact: Director, Project Ethics: (519) 824-4120, ext. 56606; directorprojectethics@gmail.com

CONSENT TO PARTICIPATE IN THE PROJECT

Please check one of the following boxes:

☐ I have read this information letter, had any questions answered, and I consent to participating in this project as described. I further understand that I may withdraw my participation in this project at any time and my responses will be excluded from the final analyses of this project.
Appendix K: Debrief and Second Letter of Information and Consent

Interview Like a Pro

THIS PROJECT WAS IN NO WAY AFFILIATED WITH THE UNITED STATES GOVERNMENT.

You have been asked to participate in a research study conducted by Paul Comeault from the Department of Psychology at the University of Guelph. The results of this study will be used, in the form of a thesis authored by Paul Comeault towards completion of a Master’s degree in Industrial/Organizational Psychology.

If you have any questions or concerns about the research, please feel free to contact Paul Comeault 226-821-3635 (pcomeaul@uoguelph.ca) or Dr. Deborah Powell 519 824 4120 x 52167 (dpowell@uoguelph.ca).

READ CAREFULLY: REAL PURPOSE OF THE STUDY

You were informed that your participation in this project was to determine if ratings of an applicant’s interview performance obtained using a web-based crowd sourcing technique could be used to accurately predict job performance. In fact, the purpose of this project was not to see if applicants hired using a crowd sourcing technique would perform better than applicant’s hired using traditional methods. Rather, the online task that you completed was part of a study conducted as part of Paul Comeault’s Masters thesis from the Department of Industrial/Organizational Psychology at the University of Guelph in Ontario, Canada. All audio recordings of responses to interview questions were scripted responses read by trained actors; the applicants that you heard were not in fact applying for a real government job. This study will explore if interviewers are able to disregard information given to them in an interview when rating an applicant’s responses to interview questions. In actuality, the purpose of the Interview Like a Pro project was to test if statements made in response to the first question would affect the ratings of statements made in response to the second question. Although the applicants and their responses to the interview question that you assessed were fictional, the techniques that you applied were legitimately those that would be used in a professional interview.

Depending on the experimental group to which you were assigned, you may have been instructed to apply across-question scoring or within-question scoring. If you were assigned to one of the across question scoring conditions, you would have been instructed to consider the applicant’s responses to both interview questions when rating each construct. If you were assigned to one of the within-question scoring conditions, you would have been instructed to consider only information that was provided in response to the question that targets a specific construct.
In the field of Industrial/Organizational psychology, within-question scoring has become the industry standard but it relies on the assumption that interviewers can ignore potentially pertinent information when it is not provided in direct response to a specific interview question—an assumption that until now has never been tested. Evidence from other social sciences literature indicates that it is hard to disregard or to “unknow” a fact that is pertinent to a decision. For example, studies in which jurors are asked to ignore incriminating information because it is found to be inadmissible based on a legal technicality show that jurors are more likely to reach a guilty verdict despite strict instructions from the presiding judge to ignore such evidence. My hypothesis was that participants who were instructed to apply the within-question scoring method and who were exposed to information pertaining to stress tolerance in response to question #1 would be influenced by that information when rating the applicant’s response to question #2.

We apologize that it was necessary to use deception to obtain accurate results for this project. Please understand that a certain degree of deception was necessary so that you would experience a sense of responsibility for your responses similar to that experienced by a hiring manager.

Because you were given inaccurate information about the nature of this study, the consent that you provided at the beginning of today’s session is void. In order to continue this research project ethically and in good faith, we wish to provide you with an opportunity to provide informed consent now, by giving permission to use your responses.

POTENTIAL RISKS AND DISCOMFORTS

Your personal information has not been recorded. For an outline of measures that will be taken by researchers to ensure that your anonymity is maintained, please see subsection “Confidentiality”.

PAYMENT FOR PARTICIPATION

You will be paid $0.50 USD for your participation in the first phase of this study and an additional $0.50 USD for your participation in the second phase of this study.

CONFIDENTIALITY

Every effort will be made to ensure confidentiality of any identifying information that is obtained in connection with this study. The ratings you will be asked to provide will be linked to the task authentication code given to you by Crowdflower and the task completion code given to you at the end of this task. Only Crowdflower will have the ability to match these codes with your Crowdflower worker ID. Crowdflower will not have access to the responses that you make throughout your participation in this project but they will be informed about the quality of the work that you complete. Members of the Interview Like a Pro research team will have no way of linking your responses to your Crowdflower worker ID or any other personal identifying information.
Appendix K: Debrief and Second Letter of Information and Consent

Please note that confidentiality cannot be guaranteed while data are in transit over the internet. However, your responses will be encrypted using the same encryption type (SSL) that is used by online banking sites to transmit secure information.

As a participant, you have the right to review the data that you provided. If you wish to review the data you provided, please record your task completion code and contact Paul Comeault at 226-821-3635 or pcomeaul@uoguelph.ca.

PARTICIPATION AND WITHDRAWAL

You may exercise the option of removing your data from the study. The investigator may withdraw you from this research if circumstances arise that warrant doing so.

RIGHTS OF RESEARCH PARTICIPANTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. This study has been reviewed and received ethics clearance through the University of Guelph Research Ethics Board. If you have questions regarding your rights as a research participant, contact: Director, Research Ethics; (519) 824-4120, ext. 56606; sauld@uoguelph.ca

CONSENT TO INCLUSION OF RESPONSES IN THE STUDY

Please check one of the following boxes:

☐ I consent to having my responses included in the final analyses of this study.

☐ I wish to exercise my option of having my data removed from the final analyses of this study.

CLICK HERE TO PRINT A COPY OF THIS LETTER OF INFORMATION AND CONSENT
Appendix L: Dimensional and Question-based Scoring Instructions

| Instructions Common to Both Dimensional and Question-Based Scoring Conditions |
|---|---|
| “You are about to hear a candidate interviewing for an administrative job within the United States Government. The applicant will respond to the interview questions that target achievement striving and stress tolerance” |

<table>
<thead>
<tr>
<th>Question-based Scoring</th>
<th>Dimensional Scoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Studies show that rating applicant’s responses to each interview question separately leads to more accurate ratings of personality trait. Please assess each question separately. Information provided in response to Question #1 must not be considered when rating the applicant’s response to Question #2.”</td>
<td>“Studies show that using all available information about an applicant leads to more accurate ratings of personality traits. When rating a personality trait, you may use information provided by the applicant at anytime during the interview. Information provided in response to Question #1 may be considered when rating the applicant’s response to Question #2.”</td>
</tr>
</tbody>
</table>
Appendix M: Q1 Unsolicited Stress Tolerance Cues

This appendix illustrates how the applicant’s response to Question #1 will be manipulated to reflect three levels unsolicited cues pertaining to the applicant’s stress tolerance (positive, negative, and neutral). The script that is common to all three levels of unsolicited cues is written in italic font without boldfaced font. The portions of the script that will be manipulated to reflect each level of unsolicited stress tolerance cues are written with bold font in separately labeled columns.

QUESTION 1#

INTERVIEWER:  “GIVE ME AN EXAMPLE OF A TIME AT WORK WHEN YOU WERE REALLY CONCERNED THAT YOU WOULD MISS A DEADLINE”

APPLICANT: “In my previous job I worked as a manager at a company that shipped supplies up North. Last summer, during our busiest period my supervisor gave me a large order to fill at the last minute -it was a tall order to fill...”

<table>
<thead>
<tr>
<th>Positive Unsolicited Stress Tolerance Cues</th>
<th>Negative Unsolicited Stress Tolerance Cues</th>
<th>Neutral Unsolicited Stress Tolerance Cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPLICANT: “...but I focused on the most difficult aspects of the task first and I made a few suggestions to my boss about how we could improve our regular procedures –my boss was very appreciative and I think it brought us closer together as a team.”</td>
<td>APPLICANT: “...and this was very stressful for me -I knew that I would have to really struggle to control my anxiety –even though it was apparent to everyone around me how stressed I was. Because it was taking a toll on my relationships at work, I avoided dealing with it by calling in sick.”</td>
<td>APPLICANT: “... and I knew that it was going to be a busy week for the staff of my crew. We had no choice but to try to complete the order. We did as much as we could to complete the order as quickly as possible.”</td>
</tr>
</tbody>
</table>
Appendix N: Q2 Solicited Stress Tolerance Cues

This appendix illustrates how the applicant’s response to Question #2 will be manipulated to reflect two levels solicited information pertaining to the applicant’s stress tolerance (positive and negative). The script that is common to both levels of solicited information is written in italic font without boldfaced font. The portions of the script that will be manipulated to reflect each level of solicited stress tolerance cues are written with bold font in separately labeled columns.

QUESTION 2

INTERVIEWER: “TELL ME ABOUT A TIME WHEN YOU WERE SO BUSY THAT YOU WERE PUSHED YOU TO YOUR LIMITS”

APPLICANT: “I guess that would have been when I was still the shipping manager at AllStar Freight around Christmas time. That’s an extremely busy time of year. So much rides on the Christmas season that my boss usually monitors everybody’s work more closely...”

<table>
<thead>
<tr>
<th>Positive solicited stress tolerance cues</th>
<th>Negative Solicited stress tolerance cues</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPLICANT: “…Despite all of my responsibilities, I never stressed about all of my tasks, I worked diligently, and thought of ways to reduce the stress at the office.”</td>
<td>APPLICANT: “…“...Given all of my responsibilities, I was VERY (vocal emphasis on the word “very”) anxious about all of my tasks and I was constantly panicked that I wouldn’t be able to manage everything.”</td>
</tr>
</tbody>
</table>

APPLICANT: “Even though there were a couple instances when it looked like we were going to miss a deadline, my team seemed to pull through at the very last minute.”