Interviews Instead of Self-Reports? Investigating Cues and Questions for the Detection of Honesty-Humility in Employment Interviews

by

Melissa Pike

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

© Melissa Pike, August, 2019
ABSTRACT

INTERVIEWS INSTEAD OF SELF-REPORTS? INVESTIGATING CUES AND QUESTIONS FOR THE DETECTION OF HONESTY-HUMILITY IN EMPLOYMENT INTERVIEWS

Melissa Pike Advisor: Dr. Deborah Powell
University of Guelph, 2019

Honesty-Humility is a valuable predictor in selection because of its organizational and job consequences. This research examines the interview as a selection method for Honesty-Humility and how to improve its detection. Study One investigated interviewers’ ability to detect this trait, as well as cues of this trait in the interview. Subject matter experts watched video-taped interviews and rated the candidates. Results demonstrated that experts are reliable at detecting Honesty-Humility, however their accuracy is low. Study Two then examined interview questions to detect Honesty-Humility. It was hypothesized that general personality-tailored questions help increase Honesty-Humility detection more than questions tapping into Honesty-Humility. It was also hypothesized that using probes would increase Honesty-Humility detection. Nine hundred and thirty-three M-Turk workers watched and rated five interviews. Results found that general questions and probes, and specific questions without probes improve Honesty-Humility detection. These findings support the Realistic Accuracy Model and provide implications for HH-based interviews.
DEDICATION

I want to dedicate my thesis to my loving parents, Betty and Frazer. Thank you for your unconditional love that has helped get me through the move to the University of Guelph and through all the obstacles I faced on the journey to obtaining my degree. You have done everything and anything to make this experience as easy as possible. I am eternally grateful that I have such amazing and supportive parents. I would not be where I am or who I am today without you.
ACKNOWLEDGEMENTS

To my advisor, Dr. Deborah Powell, I want to give a colossal thank you. Thank you for all of your support throughout my Masters’ journey. I could not have asked for a better advisor to help me attain my degree and finish this thesis. You have always gone above and beyond to help me achieve all my goals, and I would not have been able to accomplish so much throughout this degree if it wasn’t for your kind and generous assistance. You have pushed me to become the student I am today, and I can never thank you enough for all that you have done to get me to this point. I am so happy that I will be continuing this journey with you for my PhD in the fall and I cannot wait to see what is yet to come.

To my committee members, Dr. Joshua Bourdage and Dr. Jeff Spence, thank you for pushing me and helping me to gain new and unique perspectives on my project. You have helped me to see things differently and encouraged me to produce great work and a well-rounded thesis. Thank you for always being there to provide support whenever necessary and providing me with so much helpful and constructive feedback and ideas. You have thoroughly aided in my completion of a research project that I am proud of. To my external committee member, Dr. Peter Hausdorf, thank you for your interest in my project and for your thought-provoking questions at my defense. I really enjoyed our conversation about my thesis during my defense and you continue to help me think critically about my project in ways that I may not have without your assistance.

To my cohort members, Alexandria Elms, Craig Leonard, Katherine Gibbard and Nouran Sakr, I am so grateful to have people as amazing as you throughout this journey. You have provided me with more love and support than I could have ever asked for. Your friendship and guidance have helped me through this degree in ways for which I could never thank you enough.
Whether it is a stats assignment that we are struggling through, or thesis deadlines and roadblocks, I know that I have you to lean on and help to push me to achieve. I could not have asked for a better group for my cohort. I am overjoyed that we are all continuing this journey together to the PhD and I want to thank you in advance for being my family and for the help and support I know you will continue to provide.

To my colleague and friend, Marian Pitel, thank you for your love and support throughout my degree. I have found not only unconditional support through you, but also someone I would consider my family. I am eternally grateful for the many times you were there for me, whether it was just to talk or to provide me with any resource possible when I was struggling with something. I feel blessed to have such a strong and intelligent woman in my support system.

To the cohort above mine, Jordan Ho, Parco Sin, and Brooke Charbonneau, thank you for your support and assistance with scripts and tools for my thesis and for providing me with advice when I needed it.

To the IO program as a whole, thank you for all of your encouragement and for creating a culture in which I feel so supported and able to thrive. Each one of you has helped to contribute to my success in some way, whether it was providing me questions and feedback at brownbags or just being a helpful listening ear when I was having a problem. Thank you for helping me grow into the student and researcher that I am today.

To my friends at home in Calgary, thank you for creating a fun and loving space to help me unwind from any kind of academic stress I am having. Thank you especially to Jenna
Seymour, Briena Vanderwal and Haley Shipley for being such loving and amazing friends. You are family to me, and my journey would not have been the same without you.

Lastly, to everyone and anyone who has helped me throughout the past two years, those who have picked me up when I was down, those who helped me unwind, those who were there to celebrate any of my small or big achievements, those who provided me any tools that helped me to succeed, those who were a shoulder to cry on, an ear to listen and there to provide me advice, those who helped model the student and career woman I can only one day hope to be, I wholeheartedly thank you.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii

Dedication ....................................................................................................................................... iii

Acknowledgements ........................................................................................................................ iv

Table of Contents .......................................................................................................................... vii

List of Tables .................................................................................................................................... xii

List of Figures ................................................................................................................................... xiii

Interviews Instead of Self-Reports? Investigating Cues and Questions for the Detection of Honesty-Humility in Employment Interviews .............................................................. 1

The HEXACO Model of Personality ............................................................................................... 4

Funder’s Realistic Accuracy Model .................................................................................................. 4

Accuracy of Observer Personality Ratings ..................................................................................... 7

True Scores of Personality ............................................................................................................. 8

Detecting the Estimated True Score of HH in the Job Interview ....................................................... 9

Study One ......................................................................................................................................... 10

Interviewers Judgements of Honesty-Humility ............................................................................. 10

Between-Rater Accuracy ................................................................................................................. 11

Generalizability Theory ................................................................................................................... 11

Using Interviewer Accuracy and Cues for Detecting Honesty-Humility ......................................... 12
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cues for Detecting Honesty-Humility</td>
<td>12</td>
</tr>
<tr>
<td>Contributions</td>
<td>14</td>
</tr>
<tr>
<td>Methods</td>
<td>14</td>
</tr>
<tr>
<td>Participants</td>
<td>14</td>
</tr>
<tr>
<td>Interviewers</td>
<td>14</td>
</tr>
<tr>
<td>Subject Matter Experts</td>
<td>14</td>
</tr>
<tr>
<td>Procedure</td>
<td>15</td>
</tr>
<tr>
<td>Materials</td>
<td>15</td>
</tr>
<tr>
<td>Training</td>
<td>15</td>
</tr>
<tr>
<td>Relative Percentile Method (RPM) Scale</td>
<td>16</td>
</tr>
<tr>
<td>Personality</td>
<td>16</td>
</tr>
<tr>
<td>Results</td>
<td>16</td>
</tr>
<tr>
<td>SME Accuracy</td>
<td>16</td>
</tr>
<tr>
<td>Reliability</td>
<td>17</td>
</tr>
<tr>
<td>Cues</td>
<td>19</td>
</tr>
<tr>
<td>Discussion</td>
<td>19</td>
</tr>
<tr>
<td>Study Two</td>
<td>20</td>
</tr>
<tr>
<td>Situation Strength</td>
<td>21</td>
</tr>
<tr>
<td>Structured Interviews vs Unstructured Interviews</td>
<td>22</td>
</tr>
<tr>
<td>Structured Interview Questions</td>
<td>23</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>General vs Specific Personality Questions</td>
<td>24</td>
</tr>
<tr>
<td>Probing</td>
<td>25</td>
</tr>
<tr>
<td>Contributions</td>
<td>27</td>
</tr>
<tr>
<td>Methods</td>
<td>27</td>
</tr>
<tr>
<td>Participants</td>
<td>27</td>
</tr>
<tr>
<td>Procedure</td>
<td>28</td>
</tr>
<tr>
<td>Interview Videos</td>
<td>30</td>
</tr>
<tr>
<td>Measures</td>
<td>30</td>
</tr>
<tr>
<td>Relative Percentile Method (RPM) Scale</td>
<td>30</td>
</tr>
<tr>
<td>Demographics</td>
<td>31</td>
</tr>
<tr>
<td>Analytic Approach</td>
<td>31</td>
</tr>
<tr>
<td>Participant Accuracy</td>
<td>31</td>
</tr>
<tr>
<td>Condition Comparison</td>
<td>32</td>
</tr>
<tr>
<td>SME Accuracy</td>
<td>33</td>
</tr>
<tr>
<td>Cues</td>
<td>33</td>
</tr>
<tr>
<td>Results</td>
<td>33</td>
</tr>
<tr>
<td>Condition Comparison</td>
<td>34</td>
</tr>
<tr>
<td>Question Type</td>
<td>34</td>
</tr>
<tr>
<td>Probes vs No Probes</td>
<td>35</td>
</tr>
<tr>
<td>Condition Effectiveness</td>
<td>36</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Exploratory Analyses</td>
<td>37</td>
</tr>
<tr>
<td>Question Type</td>
<td>38</td>
</tr>
<tr>
<td>Probes vs No Probes</td>
<td>39</td>
</tr>
<tr>
<td>Condition Effectiveness</td>
<td>40</td>
</tr>
<tr>
<td>Cues</td>
<td>41</td>
</tr>
<tr>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>Exploratory Analyses</td>
<td>43</td>
</tr>
<tr>
<td>General Discussion</td>
<td>43</td>
</tr>
<tr>
<td>Limitations and Future Research</td>
<td>50</td>
</tr>
<tr>
<td>Study One</td>
<td>50</td>
</tr>
<tr>
<td>Study Two</td>
<td>51</td>
</tr>
<tr>
<td>Implications</td>
<td>53</td>
</tr>
<tr>
<td>References</td>
<td>55</td>
</tr>
<tr>
<td>Appendix A: Study One Research Assistantship Description</td>
<td>78</td>
</tr>
<tr>
<td>Appendix B: Study One Interview Questions and Rating Scale</td>
<td>79</td>
</tr>
<tr>
<td>Appendix C: Honesty-Humility Subject Matter Expert Training</td>
<td>83</td>
</tr>
<tr>
<td>Appendix D: Honesty-Humility (HH) SME Rating Sheet</td>
<td>93</td>
</tr>
<tr>
<td>Appendix E: M-turk HIT Description</td>
<td>97</td>
</tr>
<tr>
<td>Appendix F: Study Two Participants Consent Form</td>
<td>98</td>
</tr>
<tr>
<td>Appendix G: Honesty-Humility Descriptions and Cues for Study Two</td>
<td>100</td>
</tr>
</tbody>
</table>
Appendix H: Job Description for Study Two ................................................................. 103

Appendix I: Interview Questions for Study Two Interview Videos .............................. 104

Appendix J: Study Two Interview Candidate Consent Forms ...................................... 105

Appendix K: Study Two Interview Candidates Media Release Form .............................. 109

Appendix L: Study Two Interview Candidate Media Release Forms ............................. 111

Appendix M: Material Creation Methods Section ....................................................... 113

Appendix N: Study Two HH Differences between Accuracy Correlations .................... 117

Appendix O: Study Two Cues for Use for Detection of HH in the Interview (By Facet) ........ 118

Appendix P: Proposal ..................................................................................................... 122
LIST OF TABLES

Table 1. Descriptions of HEXACO Personality Traits ................................................................. 61
Table 2. Variance accounted for by Interviewers vs Interviewees ............................................ 62
Table 3. Intraclass Correlations between SMEs for HH and Its Facets in Study One ............. 63
Table 4. Variance Accounted for by Interviewers vs Interviewees in Study One .................. 64
Table 5. Study Two Data Cleaning Steps .................................................................................... 65
Table 6. Study Two Cell Comparisons ....................................................................................... 66
Table 7. Study Two HH Accuracy Correlations per Condition (Means, Standard Deviations and d) ................................................................................................................................. 67
Table 8. Study Two Overall HH and Facet Score Correlations between Self-Reports and SME Ratings ........................................................................................................................................ 68
Table 9. Study Two Facet Accuracy Correlations per Condition (Means and Standard Deviations) ........................................................................................................................................ 69
Table 10. Study Two Question Type Facet Cell Comparisons .................................................. 72
Table 11. Study Two Probe Effect Facet Cell Comparisons ....................................................... 74
Table 12. Study Two Facet Final Cell Comparisons ................................................................ 76
LIST OF FIGURES

Figure 1. A model of the process of accurate personality judgement ........................................... 77
Interviews Instead of Self-Reports? Investigating Cues and Questions for the Detection of Honesty-Humility in Employment Interviews

In employee selection, organizations can select employees based on a multitude of job-relevant predictors. One of these predictors, which is important for job performance, is personality (Barrick & Mount, 1991). Research has found that specific traits are predictive of job performance and a trait’s predictiveness can be contingent on the job in question. For example, Extraversion is important for specific jobs, such as sales, whereas Conscientiousness is an important predictor of performance in all jobs (Barrick & Mount, 1991). One trait that has been overlooked, but is particularly important to consider in selection, is the trait of Honesty-Humility (HH). HH is a trait within the six-factor HEXACO model of personality and it defines an individual’s level of Sincerity, Fairness, Greed-Avoidance and Modesty (Lee & Ashton, 2004). HH is important in selection for three reasons. First, HH can be a predictor of performance in certain jobs. Specifically, high levels of HH can predict success in jobs that involve caregiving (Johnson, Rowatt, & Petrini, 2011). Second, a recent meta-analysis found that HH can help to predict task performance on the job (Lee, Berry & Gonzalez-Mulé). Third, individuals with low levels of HH are more likely to engage in behaviours such as theft, absenteeism, workplace delinquency (Lee, Ashton, & de Vries, 2005) and unethical business decision-making (Lee, Ashton, Morrison, Cordery, & Dunlop, 2008). As a result, hiring managers may want to be able to detect a candidate’s level of HH for a specific job, but more often they may want to detect HH to decrease the negative behaviours associated with low levels of this trait. To be able to accomplish either of these goals, hiring managers need to be able to accurately assess HH in their selection system.
The most common methods to assess the construct of HH, and those similar to HH (such as one’s integrity), are self-report personality measures and integrity tests. It should be noted that integrity and HH are not the same constructs, however there is some overlap. There is a correlation of .48-.63 between integrity tests and HH, depending on the type of test used (Marcus, Lee & Ashton, 2007). Integrity can be defined as acting according to a set of moral values (Becker, 1998) and the overlap between integrity and HH likely relates to the Sincerity and Fairness components of HH (those that relate to honesty) which align with one’s values. HH is unique, however, in its ability to capture Greed-Avoidance and Modesty which tap into an individual’s humility.

Personality and integrity tests have previously been viewed as effective and efficient ways to assess HH and integrity, respectively. However, Morgeson et al. (2007) noted a number of problems with the self-report measures of these constructs when they are used within the selection context. These problems include negative perceptions of the selection process caused by the inclusion of these tests (Rosse, Miller, & Stecher, 1994) as well as the fake-ability of these measures, which can be defined by a candidate’s ability to present themselves in an ideal light on a measure (Berry, Sackett, & Wiemann, 2007; Topping, & O’Gorman, 1997). Because of these issues, Morgeson et al. (2007) suggested that alternatives to self-reports should be investigated to detect personality in selection.

A possible alternative to self-report measures for detecting HH in selection is to assess HH during the employment interview. Interviews could be a more ideal method for detecting personality traits such as HH because they are widely used and accepted by organizations for use in candidate selection, and applicants tend to view interviews more positively than personality tests (Steiner & Gilliland, 1996). Also, research has found that faking in general occurs to a
lesser extent in interviews than on personality tests (Van Iddekinge, Raymark, & Roth, 2005). Less response distortion in interviews may result from candidates’ behavioural descriptions being separated from the evaluation of those behaviours (Levashina, Hartwell, Morgeson, & Campion, 2014). When self-reports are used, candidates will rate themselves on a behaviour, and may fake their response depending on how they think that behaviour is perceived by others. However, when interviewers assign ratings in interviews there is a separation of behavioural descriptions (given by the candidate) and ratings (given by the interviewer). Therefore, applicants have less control over their scores when an interviewer is responsible for assigning them, making it more difficult for applicants to engage in faking in the interview. In addition to the benefits of reduced fake-ability, meta-analyses have found that observer ratings of personality are more predictive of job performance than are self-reports (Connelly & Ones, 2010; Oh, Wang, & Mount, 2011). Therefore, in the context of selection, observer ratings made through interviews may be uniquely predictive as compared to self-reports measures. Interviews may be a more expensive and time-intensive method, compared to self-reports; however, given their overall benefits, including positive candidate reactions, being less fakable and predicting job performance, interviews could be a viable alternative for assessing personality in selection.

For interviews to be a useful tool to detect personality in selection, they must be effective enough to merit the extra time and potential cost, relative to self-report personality measures. Fortunately, research conducted by Powell and Goffin (2009), as well as Powell and Bourdage (2016), has shown promise for the detection of personality in interviews. The problem, however, is that researchers have yet to examine the detection of HH in interviews. Therefore, the overarching research question that this research will be looking to answer is: can the personality trait of HH be detected in the interview?
The HEXACO Model of Personality

Personality is often thought of in terms of five traits within the Big Five model of personality (Goldberg, 1990; Goldberg, 1992). However, research has recently discovered that personality is more robustly measured by six traits (Ashton, Lee & Son, 2000). These traits are encompassed in the HEXACO model of personality and are categorized as the following: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. The latter five traits are encompassed within the Big Five, whereas Honesty-Humility is unique to the HEXACO. Each trait is described in more detail in Table 1. Honesty-Humility is broken into the facets of Sincerity, Fairness, Greed-Avoidance and Modesty. With these new, previously untapped facets, Honesty-Humility captures elements of personality that had yet to be properly encompassed within the Big Five. These nuances within the HEXACO have also enabled the HEXACO to gain predictive validity over the Big Five in some organizational concepts. For example, the HEXACO accounts for more variance than the Big Five in workplace delinquency, likelihood to sexually harass (Ashton & Lee, 2007), unethical business decision-making (Ashton & Lee, 2008), and job performance in specific jobs, such as those involving caregiving (Johnson et al., 2011). As a result, there are benefits to using the HEXACO personality model in the workplace and examining the trait of HH.

Funder’s Realistic Accuracy Model

For interviewers to be able to examine HH in their selection system, they must be able to detect the HH level of the candidate. To detect any personality trait, an interviewer goes through several steps. These steps are encompassed in the Realistic Accuracy Model (RAM) proposed by Funder (1995) and depicted in Figure 1. According to the RAM, four stages must occur for personality to be accurately detected. In the interview context, the first two stages are based on
the interviewee. An interviewee must emit cues of a personality trait (the relevance stage) in front of the interviewer (the availability stage). The last two stages occur at the level of the interviewer. The interviewer must pay attention and notice the personality cues emitted by the interviewee (the detection stage), and the interviewer must know that the cues are related to a specific personality trait and use the cues to make that specific trait attribution (the utilization stage). The cues that the interviewer will use in this process can be based on the candidate’s behaviour or what they say in the interview and through these stages the cues can lend information to the interviewer about the candidate’s personality.

The RAM model for personality detection is useful for the overall detection of personality traits, however its stages are also important because some traits are easy to detect, while others are more difficult. Funder (1995) stated that whether a trait is easy or difficult to detect depends on four components.

The first component Funder (1995) proposed is visibility. If a trait is highly visible and it has cues that are presented easily and often, it is easier to detect. An example of a highly visible trait is Extraversion, as this trait can be easily detected through simple interactions with others. A trait that is not as visible is the trait of Openness to Experience (Connelly & Ones, 2010). This trait encompasses factors such as inquisitiveness and being imaginative, and these factors are often not inherently visible.

The second component is frequency (Funder, 1995). As cues are presented with higher frequencies, they become easier to detect. This effect is due to increased availability of the trait’s cues, as when cues are being presented at higher frequencies, they are more readily emitted and therefore, more readily available for detection by a judge.
The third component is whether the trait is presented ambiguously (Funder, 1995). If cues of a trait are ambiguous, that trait can be difficult to detect because judges may be unsure if the cues an individual is emitting are related to that specific trait or not. For example, if someone is not very talkative, they might be introverted, or the situation may have made them feel nervous.

The fourth and final component relates to how positively or negatively levels of a trait are judged by others (Funder, 1995). If an individual is going to be evaluated negatively because they have a certain level of a trait, they will try to avoid displaying cues of that trait. On the other hand, if they know they are going to be evaluated positively by displaying cues of a trait, they are more likely to do so. For example, if people are low on agreeableness, they often would not be regarded positively in normal day-to-day interactions. Therefore, they may try to appear more pleasant and easy to get along with.

To summarize, the degree to which a trait is visible, presented frequently, non-ambiguous and evaluated is important for the detection of that trait and the more of each component that is present, the easier a trait will be to detect (Funder, 1995).

All the factors specified above affect the ability of others to detect HH. The problem with HH is that it is what Funder would characterize as a ‘difficult trait’, meaning that HH is hard to detect. For the trait of HH, some facets, such as one’s Sincerity, are not always inherently visible to others and can be quite ambiguous (it is often hard to tell if someone is truly being sincere). In addition, behaviours associated with HH are not always presented frequently, e.g. displays of Modesty or lack thereof. Finally, people tend to be evaluated negatively if they are greedy, insincere and unfair, and therefore, individuals generally avoid openly indicating if they are low on HH. These factors combined make the trait of HH hard to detect.
Research provides support for HH being a ‘difficult trait’ to detect. Without ample time to get to know someone, it is difficult to judge their HH level (Ashton & Lee, 2010). Indeed, the correlation between acquaintances’ self- and other-reports of HH is only .22, and it typically is not until individuals are friends, family members or in romantic relationships (self-other correlations of .30, .49 and .60, respectively) that they can judge the others level of HH with some degree of accuracy (de Vries, Lee, & Ashton, 2008). Therefore, when individuals have not spent a great deal of time getting to know each other, HH is a difficult personality trait to judge. With time, people can judge HH with greater accuracy, due to increased exposure to that individual in different situations, and consequently, increased chances to view HH cues emitted from that individual over time. A problem arises, therefore, when one is considering the detection of HH in the employment interview context, because time with a candidate is short and confined to a specific situation. Therefore, interviewers need to ensure they can accurately detect HH in this short time frame.

**Accuracy of Observer Personality Ratings**

In personality detection, Schmid Mast, Bangerter, Bulliard, and Aerni (2011) define personality judgement accuracy in terms of the correlations between self-report and other/judges’ reports of personality. This definition is consistent with the definition of convergent validity: the correlation between two different ways of measuring targets’ personality (Paunonen, 1984). In the person-perception literature, however, these correlations constitute personality judgement accuracy. The accuracy correlations between self-reports and other reports are used to determine how similar a judge’s rating is to individuals’ true score of personality, or their ‘true’ level of personality traits.
**True Scores of Personality**

According to the Classical Test Theory of measurement, an observed measure of a construct, is made up of a “true score” and error (Novick, 1966). For example, every individual has what is classified as their true score of each personality trait. This true score indicates the expected level/value that would be produced over an infinite number of replications of a specific personality measure for a specific person (Lord & Novick, 1968). Unfortunately, this score is never able to be measured (Lord & Novick, 1968) and therefore, attempts to measure a specific construct can only provide an approximation or estimate of one’s true score. There is always error present in measurement and this causes the observed score to differ from the true score (Novick, 1966). Therefore, although getting an exact measure of one’s true score is impossible, measures of personality are created to attempt to measure an individual’s true score of personality with as much accuracy as possible. These measures will only, however, be able to provide an estimate of the true score and not the value of the “true score” itself.

The most commonly used tools to estimate one’s true score of personality are self-reports and observer reports. Neither is a perfect form of measurement. Both self-reports and observer reports will tap into individual’s response biases, which cause error and therefore affect the measures ability to detect an individual’s true score (Dodorico McDonald, 2008). However, within the personality judgement literature, self-reports are often used as a measure of the estimated true score. As mentioned above, Schmid Mast et al. (2011) uses self-reports as the estimated true score measure in their definition of personality judgement accuracy. Research has also shown that self–other agreement is an accurate criterion to use when examining personality judgement accuracy (Funder & Dobroth, 1987), and therefore self-reports are judged as to be a reasonable measure to estimate an individual’s true score. In addition, research states that self-
reports are an accurate tool to use when estimating one’s true score of personality because individuals themselves should have the most knowledge about the behaviours they engage in and the motivations behind those behaviours (Paulhus & Vazire, 2007). This literature provides support for the use of self-reports as a measure of one’s estimated true score and will be one method of estimated true score measurement (the criterion) in this research.

Although self-reports do have merit as an estimated true score measure, research can benefit from increased accuracy and a more complete depiction of a construct by using more than one method of estimated true score (or criterion) measurement (Dodorico McDonald, 2008). To take advantage of this benefit, expert observer reports will also be utilized in this research to estimate an individual’s true score of personality. These experts have been trained on the trait of HH, giving them an advantage in their ability to detect this trait, as compared to observers who are more unaware of the uniqueness of this trait. This training should also allow them to make judgements that more closely align with the ‘true score’ of an individual’s personality as well. Using both self and expert observer reports as measures of one’s estimated true score will help provide a more thorough analysis of accuracy in HH judgement.

Detecting the Estimated True Score of HH in the Job Interview

For an interviewer to accurately assess HH during an interview, the candidate must emit cues so that the interviewer can detect and utilize these cues to make accurate judgements. The question then arises: to what extent does HH cue emission and accurate detection occur in the job interview? To examine this issue, I will conduct two studies: one to examine cues of HH emitted in real interviews, and another to examine whether the type of interview questions can increase HH detection accuracy.
Study One

Interviewers Judgements of Honesty-Humility

To rate personality in an interview, interviewers must detect cues emitted by the candidate. Cues of a candidate’s personality could come from many different sources, including what the candidate says before, during or after the interview, how the candidate speaks, and/or the candidate’s non-verbal communication. In the context of this research, cues of personality will focus on information the interviewee verbally provides in their answers to interview questions because it is the most explicit information and requires the lowest degree of inference for making judgements.

Ideally, the interviewer would notice the cues provided to them by the candidate and then make a rating of a personality trait (e.g., HH) that approximates the candidate’s “true score”. However, in a more realistic situation, an interviewer’s detection will be imperfect. Error present in detection will affect the degree to which they can accurately judge a candidate’s personality. It is not known, however, with what level of accuracy interviewers can rate a candidate’s personality. Previous research has examined the accuracy of strangers and acquaintances in judging another’s personality (de Vries et al., 2008), however the interview is a very specific context that could affect accuracy. It is important to determine interviewer accuracy, because if interviewers can be accurate at detecting HH, it would demonstrate that detection of HH in the interview is feasible. Therefore, the first research question for this study is:

R1: Can interviewers trained to detect HH judge HH with some degree of accuracy?
**Between-Rater Accuracy**

Although individual interviewer accuracy is important, it is also important for interviewers judging HH to produce similar judgements to one another and therefore be reliable. Reliability in this context refers to inter-rater reliability, the variation in judgements between more than one rater when judging the same subject (Koo & Li, 2016). Often in a selection system more than one interviewer will be used to conduct interviews. If this is the case for a personality/HH-based interview or interview containing questions relating to HH, the interviewers must be able to produce similar scores. If the scores are not similar, then candidates will be selected at different rates simply due to the interviewer(s) who conducted the interview.

To avoid this problem, research needs to investigate how much variance there is between individual interviewers in HH detection and the reliability of their scores. Therefore, the second research question of this study is:

R2: Can interviewers trained to detect HH judge HH reliably?

**Generalizability Theory**

A method that can be used to judge the variance between interviewers, and test one form of the reliability of raters, is a generalizability study. As specified by Shavelson and Webb (1991), generalizability (G) theory (tested within a generalizability study) relates to how dependable behavioural measurements are. Its main goal is to determine the variance accounted for by various sources, such as items, raters, targets, and occasions. G theory partitions out the variance from various sources to allow researchers to determine the answer to questions like: how many items or raters are needed to obtain dependable or generalizable scores? Classical test theory is a branch of G theory, with G theory being able to account for more than one source of error/variance in one analysis. For example, in the research explained here, there are four sources
of variability. These sources include variability due to the interviewees levels of HH (which is classified as the object of measurement), interviewers’ variability between each other in rating an interviewee’s level of HH, the interaction of interviewer and interviewee variability, and error. G theory can calculate the individual variability of interviewees’ HH and interviewer’s ratings of HH. Ideally the main source of variability for this specific study would come from differences in the interviewees’ levels of HH. G theory can help examine if this is the case and therefore answer this study’s final research questions of:

R3: In interviews, what percentage of variance is due to differences in interviewees’ levels of HH and what percentage is captured by differences in interviewers’ ratings of interviewees’ HH levels?

Using Interviewer Accuracy and Cues for Detecting Honesty-Humility

Interviewer consistency and accuracy information are important to determine the reliability and validity of interviewer judgements of HH. However, this information can also be used to help determine relevant cues of HH. If interviewers are accurate and reliable at detecting HH, then the cues they use to detect HH may be valuable for use by future interviewers to detect HH in the interview. Therefore, information obtained about accuracy and reliability may help to determine useful cues of HH in employment interviews.

Cues for Detecting Honesty-Humility

The Realistic Accuracy Model’s stages revolve around the elicitation and detection of personality trait cues. Without these cues, personality detection cannot occur. In addition, for an interviewer to be able to detect a difficult trait like HH in the short time frame of an interview, the interviewer must understand what cues to look for. However, research has yet to fully identify good cues of HH, particularly in the selection context. Lee and Ashton (2012) have
suggested some valid signs of HH such as those who try to beat the system, engage in instrumental ingratiation, gamble, are sexually unfaithful, partake in conspicuous consumption, have an above-the-law mentality, and have contempt for other groups. The problem with many of these cues is they are hard to detect in an employment interview. For example, it is very unlikely that someone is going to discuss sexual infidelity or a gambling problem in an interview. In addition, if someone tries to beat the system, it is unlikely that they are going to make that fact explicit in a job interview. Therefore, not only do more cues of HH need to be identified, interview-specific cues needed to be identified.

Some possible cues for identifying HH could be drawn from the integrity literature, because integrity tests seek to discover individuals who are dishonest and lack integrity and typically engage in behaviours such as theft (Berry et al., 2007). Trying to tap into these traits can be done by inquiring about concepts such as punitiveness and projectiveness. Punitiveness defines how likely someone is to want to punish others for engaging in wrongful acts, and projectiveness is defined as someone projecting their personality on to others (assuming that they are similar to others; Cunningham, Wong & Barbee, 1994). These concepts relate to honesty, because those who are more honest tend to be more punitive and those who are dishonest tend to project their dishonesty onto others; they assume that others must have similar levels of dishonesty (Cunningham et al., 1994). Although these cues may be helpful to detect HH, it is important to remember that integrity and HH do not overlap completely (e.g. one’s level of modesty is not necessarily related to their integrity), and therefore cues of HH may somewhat differ from cues of integrity.

Because the knowledge we have of interview-specific HH cues is scarce, it is important to find cues of HH in the interview. These cues, if demonstrated to be valid cues of HH, could
then be used by an interviewer to help them accurately detect the HH level of a candidate.

Therefore, the third research question for this study is:

**R4**: What are cues of HH that can be elicited in employment interviews?

**Contributions**

This study is going to add to the current literature by investigating interviews as a possible avenue to detect HH. It demonstrates the accuracy and reliability of interviewers in detecting HH as well as investigates possible cues for detecting HH in the job interview context.

**Methods**

**Participants.** Eight-two University of Guelph students who were applying for a research assistant job in Dr. Powell’s lab were interviewed. Eighty-three percent identified as female and 17% identified as male. Ages ranged from 18 to 41. Forty-three of the candidates were Caucasian (52%), two were Aboriginal/First Nations/Metis (2%), five were Arab (6%), five were Black/African/Caribbean (6%), 12 were South Asian (15%), five were Southeast Asian (6%), and 10 classified their ethnicity as other (13%).

**Interviewers.** The interviewers consisted of four research assistants in Dr. Powell’s lab. They were trained to administer the interview questions, ask probes, and score the candidates on anchored rating scales. These scales were only used for hiring purposes and were not examined for the purpose of this research. All the interviewers were female.

**Subject Matter Experts.** The SMEs that watched the videos for this study included myself and three research assistants selected by Dr. Powell. All raters received training to become experts on the trait of Honesty-Humility and all raters were female (training is described below under materials).
Procedure. The interviewees were videotaped as they completed an interview for a research assistantship (see Appendix A for the job description). They answered four questions and were probed to provide further information, if deemed necessary by the interviewer. See Appendix B for the interview script. Following the interview, the participants completed a self-report of their personality. This data was originally collected for another study by Dr. Powell. Other items were also collected for Dr. Powell’s study that will not be used for the purpose of this current study; specifically, the Honest Impression Management Scale (HIMS), the Interview Faking Behaviour – Short scale (Bourdage, Roulin & Tarraf, 2017), the Measure of Anxiety in Selection Interviews (MASI; McCarthy & Goffin, 2004), STAR (Powell, Roulin, & Bourdage, 2018), and the Machiavellian Personality Scale (MPS; Dahling, Levy & Whitaker, 2009).

Following the collection of the data specified above, SMEs were trained to become experts in the trait of HH. After training, the SMEs watched the videotaped interviews and rated each of the candidates on a Relative Percentile Method (RPM) scale. They rated each of the candidates individually. Each candidate was rated on their overall level of HH as well as each facet of HH (Sincerity, Fairness, Modesty and Greed-Avoidance). In addition to rating the candidate, SMEs recorded any cues of HH they thought the interviewee emitted in the interview and indicated which facet they thought corresponded with each cue.

Materials.

Training. Training for the SMEs consisted of them reading descriptions of HH and its facets, view self-report items that correspond with HH, as well as read chapters 1-5, and 9-10 of Lee and Ashton’s (2012) book on HH. They also completed a test to demonstrate their knowledge of HH. This test consisted of a list of the HEXACO-100 items. The SMEs had to indicate which items they thought corresponded with HH and each of its facets. Any mistakes
were identified and corrected by explaining to the SMEs which items should have been selected and why. It should be noted that the SMEs made at most two mistakes on the test. The training, in its entirety, took SMEs approximately two hours. The training document provided to SMEs is provided in Appendix C.

**Relative Percentile Method (RPM) Scale.** The scale used by SMEs to rate each candidate’s level of HH was an RPM scale, which has been judged as advantageous in past research (Goffin, Jelley, Powell, & Johnston, 2009). This scale ranges from 0-100 (0 = very low, 50 = average, 100 = very high). SMEs rated each candidate by deciding whether they thought the candidate was very low, average or very high on HH when compared to others in the general population. A depiction of the scale as well as instructions given to the SMEs are provided in Appendix D.

**Personality.** Personality was measured by the self-report 60-item HEXACO (Ashton & Lee, 2009). This measure has an internal consistency reliability ranging from .73 to .80 for each of the personality trait scales (Ashton & Lee, 2009). The internal consistency reliability of the HH scale specifically is .76 (Ashton & Lee, 2009). The items of the HEXACO are presented on a 5-point Likert scale (1 = *strongly disagree*, to 5 = *strongly agree*). This measure was used to obtain self-reports of HH from the interviewees and is used in this study to reflect an individual’s estimated true score of personality. This measure was filled out by interviewees’ after their interview, in a separate room. They were informed that the results from this survey would not affect their chances of getting the position to encourage honesty in responding.

**Results**

**SME Accuracy.** The accuracy of HH detection by the SMEs was assessed in terms of specific personality trait accuracy as specified by Davis and Kraus (1997). Schmid Mast et al.
(2011) termed this type of accuracy as the correlation between the judges’ ratings and interviewees’ self-reports of a single personality trait across all of targets in the employment interview. This type of accuracy examines how well the judge detected variation between the interviewees on a specific personality trait, here being HH and its facets.

Specifically, accuracy correlations were calculated between interviewee self-reports and SME ratings to answer R1: Can interviewers trained to detect HH judge HH with some degree of accuracy? Each SME had an accuracy correlation, and an accuracy correlation was also calculated with an average of the SME scores. Overall, accuracy correlations for HH and its facets were poor. The correlations for SME three for HH and Modesty, and the SME average were the only correlations that had confidence intervals that did not cross zero, $r = .26$, 95% CI [.04, .45], $r = .30$, 95% CI [.08, .49], and $r = .28$, 95% CI [.06, .47], respectively (Table 2). SMEs most accurately judged overall HH and the facet of Modesty. It is important to consider, however, that even though some of the correlations had confidence intervals that did not cross zero, the low ends of the confidence intervals were very close to zero, so it is difficult to determine if these effects are meaningful. To conclude, the interviewers are not able to judge HH and its facets with much accuracy.

**Reliability.** Two-way random effects model intraclass correlations (ICCs) were calculated as the first form of reliability for the SMEs. These correlations demonstrate the reliability present when only a single SME rated the candidates’ HH and facet levels, as well as the reliability when four SMEs rated the candidates. Random effects ICCs allow for generalization of the reliability to other raters (Koo & Li, 2016). Therefore, the ICCs can help to determine the consistency of raters when making HH judgements.
Based on benchmarks by Koo and Li (2016) moderate reliability is determined by ICCs between .50 and .75, good ICCs range from .75 to .90, and excellent reliability occurs when ICC values are .9 and up. Using these benchmarks, the ICCs for HH and its facets can be judged as moderate when all raters scores were analyzed, with the lowest ICC resulting from Fairness ($ICC = .59, 95\% \ CI [.42, .72]$) and the highest for Sincerity ($ICC = .75, 95\% \ CI [.64, .83]$; Table 3). These analyses demonstrate that the raters can judge HH and its facets reliably, answering R2.

A generalizability study (G-Study) was conducted to examine a second form of reliability for the SME ratings. This G-study examined the sources of measurement error within the study design. As this is a one-facet design (using only interviewer ratings) the universe was defined by four sources of variability, the variance examined was among interviewees ($o$), raters ($i$), the interaction of interviewees and raters ($o \times i$) and random error. In the G-study, ANOVA was used to partition and examine each of the variances.

The G-study was used to test R3, which asked what percentage of variance is due to differences in interviewees’ HH and what percentage is captured by differences in interviewers’ ratings of interviewees’ HH? Ideally variance due to differences in interviewees’ HH levels would be larger than the variance accounted for by differences in interviewers’ ratings, because interviewees are inherently different from one another in their levels of HH.

The variance component for interviewees was of a moderate size. For HH, interviewees accounted for 35.3% of variance. For the facets, interviewees accounted for 36.1% of variance for Sincerity, 21.8% for Fairness, 33.8% for Greed-Avoidance, and 33.8% for Modesty. Additionally, the variability attributed to raters had a small effect on the percentage of total variation. Rater variance accounted for 10.2% of variance in HH, 15.1% of variance in Sincerity, 16.9% of variance in Fairness, 9.6% of variance in Greed-Avoidance, and 6.7% of variance in
Modesty. The residual variance had the largest contribution of total variance for both HH and its facets (Table 4).

**Cues.** For the cues to be judged as accurate for detecting HH three criteria were considered. First, more variance in the scores from the G-study must come from the interviewees as compared to the interviewers to demonstrate adequate reliability. Second, the ICCs needed to demonstrate adequate reliability values. Third, SME accuracy correlations needed to be greater than .5. This information demonstrates that there is not more error than true scores present, and therefore, that the SMEs are providing valid ratings of the candidate’s HH.

More variance in the scores from the G-study came from the interviewees than the interviewers and together the raters demonstrated moderate reliability (Table 2). Moderate reliability was also determined through the ICC values (Koo & Li, 2016; Table 3). This information demonstrates that the cues provided by the SMEs could be useful to detect HH in the interview. However, even though three of the correlations had confidence intervals that did not cross zero, the SMEs accuracy correlations with the candidates’ self-reports were not greater than .5 (Table 4). This information demonstrates that there is more error than true scores present in their ratings. Because of this, the cues cannot be used to demonstrate cues that could be useful in an interview to detect HH and its facets.

**Discussion**

The first study examined a variety of research questions. First, it examined the ability of interviewers to judge HH accurately (using self-reports as estimated true scores). Second, it tested whether interviewers can judge HH reliably by examining ICCs and what variance is accounted for by differences in interviewees’ scores and interviewers’ ratings of HH. Third, and finally, possible cues of HH that were emitted by interviewees were investigated. This study
demonstrated that although the interviewers’ can reliably judge HH and its facets (i.e. there is agreement in their ratings), the accuracy (relative to self reports) of interviewers is very low. Because the interviewers demonstrated moderate reliability, it can be said that the interviewers are likely to produce similar scores when judging the HH level of a candidate, and therefore likely to select candidates at a similar rate. However, although the scores may be similar, they may not be accurate, based on the overall low interviewer accuracy. This low accuracy would cause issues in a selection system as the candidates could be incorrectly selected based on the trait of HH. In addition, because of this low accuracy, it cannot be determined whether the cues used by the SMEs to judge HH were accurate cues of HH in a job interview.

This study has helped to demonstrate that although accuracy of the SMEs was low, they are able to produce reliable judgements. This study was limited, however, in that the questions asked in the interview were not created in a manner meant to elicit HH and, therefore, the candidates may not have emitted enough accurate HH cues for the interviewers to accurately detect this trait. Therefore, the next study will look at the types of questions asked in the interview to increase the HH judgement accuracy of interviewers.

**Study Two**

Study One demonstrated that even with training on the trait of HH, it can be quite hard to detect HH in interviews. Nonetheless, improving the detection of HH could be accomplished by modifying the design of the interview itself. To look at this, Study Two will assess conditions that improve the accuracy of HH detection in the interview.

For accurate HH detection to occur, the questions asked in the interview must elicit HH cues from candidates, thus increasing the chances that the judge will detect and utilize those cues. Therefore, it is important to know which types of questions can best attain this goal. The pursuit
of this goal has been discussed by Levashina et al. (2014) who stated that research needs to be done on the design of personality-based interviews and one of the areas that needs to be examined specifically is types of questions asked. By examining these personality-based interview questions, types of questions can be identified that best elicit trait-relevant information and information consistent with how an individual would act day-to-day, so the interviewer can accurately detect the candidate’s personality. Therefore, the research question that will be addressed in this study is: what types of questions lead to the greatest degree of HH detection accuracy?

**Situation Strength**

One important factor to consider in designing questions to elicit HH cues is the concept of situation strength. Situation strength is defined by external cues provided in a situation that signal the desirability of certain behaviours; these situational cues cause psychological pressure on individuals to act in a similar way to others, regardless of how they themselves would typically act (Meyer, Dalal & Hermida, 2010). By this definition, if a situation elicits many cues of behavioural desirability, the situation will be strong, and personality will not play a large role in behaviour. Instead, individuals will behave in a way that the situational cues tell them to behave. An example of situational strength is an extravert in a quiet office. If they are silent, it is not because they typically behave this way based on their personality. Instead, this behaviour is likely because the situation signals to them that it is not appropriate to be loud and outgoing. If situation strength were to be applied to the trait of HH, and a strong situation signalled to an individual that high HH is most appropriate, then that individual would feel as though they needed to appear sincere, fair, greed-avoidant and/or modest, regardless of if they are high on
these facets and consequently high on HH. The situation therefore constrains what HH cues an individual will elicit, based on what that individual thinks is appropriate and/or desired.

Because of how situation strength affects behaviour, it could affect the availability of HH cues in an interview. If the interview questions create a situation in which the candidate is restricted in the possible answers they can give, it may affect the cues the candidate can emit in response to those questions. In other words, the questions may signal to the candidate how they should respond, and therefore genuine HH cues may not be elicited. Based on this idea, situation strength seems important for a personality interview. Despite this evident importance, research has not examined how the questions asked in an interview could affect situation strength and therefore personality cue availability and overall detection. My research will look at varying the situation strength in interview questions. My aim is to demonstrate the possible effect that situation strength may have on HH detection.

**Structured Interviews vs Unstructured Interviews**

To manipulate situation strength in interview questions within this research, an interview type must be selected, as the type of interview used affects the types of questions asked. The two main types of interviews are structured interviews and unstructured interviews. Based on the concept of situation strength, it should follow that unstructured interviews are beneficial for personality, as they present a more open-ended question format to allow the natural elicitation of personality. Research supports this idea. Blackman (2002) found that unstructured interviews lead to better judgement of personality when compared to more structured interviews. With this in mind, it might seem best to look at types of interview questions within the unstructured interview for the purpose of this research. However, unstructured interviews are less reliable and valid tools than structured interviews (Levashina et al., 2014). Therefore, it is valuable to
investigate the questions asked in structured interviews to try to increase the potential for personality traits, such as HH, to be detected in these interviews. As a result, this research will use structured interviews and try to decrease the situation strength of the questions in the interview (with regards to HH).

**Structured Interview Questions**

Within structured interviews there are many different types of questions that can be asked so it is important to ensure the best type is used for this research’s purposes. The questions asked are typically broken into three categories, behavioural questions, situational questions, and technical job knowledge questions. Job knowledge questions ask about aspects of the job that candidates are expected to know to properly perform the job, behavioural questions ask candidates about previous behaviour on the job, and situational questions ask how a candidate would behave in a hypothetical job situation (Levashina et al., 2014). Behavioural and situational questions can be created to tap into a variety of knowledge, skills and abilities, including one’s personality.

For the purpose of this research, detecting personality, behavioural questions may have some overall benefits. Some research has found that behavioural questions measure personality traits (Krajewski, Goffin, McCarthy, Rothstein & Johnston, 2006), and that they are structured like personality inventories because they are past-oriented (Levashina et al., 2014). In addition, behavioural questions have slightly higher criterion validity than situational questions (Levashina et al., 2014). Therefore, it may be beneficial to use behavioural questions in a personality interview. Following this, behavioural questions will be used in this research.
General vs Specific Personality Questions

When creating behavioural interview questions to elicit personality cues, these questions can be created to tap into a specific trait (e.g., Van Iddekinge et al., 2005). This approach intuitively makes sense; if you want to know the answer to a question, you ask a specific question to elicit that answer. Tying this to personality, if you are trying to detect a personality trait, it makes sense that you would ask a specific question to tap into this trait. A problem with this approach, however, is that interview questions created to tap into specific personality traits may create a strong situation. This strong situation may result because the question provides cues to candidates as to how they should answer. For example, if an interviewer is trying to elicit cues of Extraversion from a candidate and they ask: “Tell me about a time when you had a work environment you enjoyed. Describe your interactions with your co-workers.” This question will likely to only elicit a small range of answers, and depending on the organization, the candidate may feel that they need to appear extraverted or introverted to get the job. However, if the question was created to be more general, therefore decreasing situation strength, sincere personality cues may be elicited more readily. If a candidate was asked “Tell me about a time when you had a work environment you enjoyed. What about this environment made it enjoyable?” the question could elicit cues of Extraversion but could also elicit cues of other traits such as Conscientiousness. Simply by eliminating one constraint, the focus on co-workers, the question becomes broader and more readily able to elicit a variety of personality cues. For the trait of HH, if questions evidently tap into HH’s facets, candidates will likely feel as though they need to answer in a certain way. Because of how HH is evaluated, that being low on HH is viewed negatively, people will likely feel as though they need to answer to appear high on HH to get the job. However, if the question is general enough to allow HH cues to be expressed more
naturally, then the interviewer should be able to make a more accurate judgement of HH. In more
general terms, questions created to tap into multiple personality traits increase the range of
possible answers the candidate can give and, therefore, the number of genuine personality cues
that can be elicited.

The general personality questions described above can be compared to those asked in
unstructured interviews. These types of questions create an environment where the interviewee
can speak more freely because they feel less constraints on how they should answer. Tying this
back to the RAM (Funder, 1995), general personality-tailored questions should elicit more cues
relevant to the individual’s “true” level of specific personality traits and therefore make these
cues available to the interviewer for them to judge. Consequently, in an interview, using general
behavioural questions (versus specific) should lead to better HH detection overall.

H1: Interviewers will have greater accuracy in detecting HH in response to general
personality-tailored behavioural interview questions as compared to behavioural
interview questions that tap specifically tap into HH.

Probing

Decreasing the strength of the situation in an interview question is one way to help elicit
personality information from a candidate. Another way to elicit personality-related cues could be
through using probes. Probes are a type of follow-up question used to obtain a more elaborate
response from an interviewee if their answer originally seemed vague or incomplete (Levashina
et al., 2014). Research has found that probing may help to increase the accuracy of information
gathered in an interview (Motowidlo et al., 1992; Schmidt & Conaway, 1999). There is some
support that this increased accuracy also applies to personality interviews. Blackman (2002)
found that in interviews that included more probing, interviewers made more accurate
personality judgements. In Blackman’s research the type of interview that included more probing was an unstructured interview, however, this finding may apply to structured interview questions as well. By asking probing questions, interviewers may obtain more accurate information about the behaviour the interviewee engaged in and the reasons behind their behaviour. In other words, the probes may help the interviewer to elicit more trait-relevant cues. Specifically, this research is going to focus on asking probes that inquire about an applicant’s reasons for engaging in certain behaviours and what they were feeling when they engaged in these behaviours. By asking these types of questions candidates must reflect on why they engaged in their behaviours and this could help to elicit cues of HH. For example, if a candidate is asked why they chose a certain approach, they may make a comment along the lines of ‘I did this because I wanted to help my teammates.’ This response could tap into the facet of Fairness and therefore cause higher, and more accurate ratings of that interviewee’s HH. If the probes elicit more cues from interviewees, then the interviewer should be able to make a more accurate judgement of the interviewee’s HH. This research will focus on adding probes, but keeping probes consistent for each question, which will add structure to the interview, while still increasing opportunities for cues to be elicited similar to unstructured interviews.

H2: The use of probes (versus no probes) will lead to increased accuracy in HH detection.

Probing may be important for eliciting more personality cues, however it may be especially important for general personality-tailored questions. For these questions, candidates may provide an array of different information, tapping into a variety of traits, because the questions are quite broad. Although this is the goal of asking this type of question, the information the candidate discusses may stray away from information that will provide the cues that the interviewer desires. If this happens, it could be difficult for the interviewer to detect the
trait they were hoping to elicit. By adding probes, the interviewer can give the candidate more opportunities to emit relevant personality cues. Also, probes may help to delve deeper into what the candidate has said to ensure the interviewer is not making an inference as to what the candidate means by their statements. For example, if a candidate were to say they were influencing their friend to do something, an interviewer may assume that the candidate is lower on HH. However, by asking a probe, the interviewer may find out that the candidate is trying to influence their friend to do something because the candidate knows it will benefit their friend in the long run. Therefore, the interviewer may not judge the candidate as low on HH simply because a probe was asked, and the interviewer is able to make a better judgement as a result. In sum, probes may help the interviewer obtain more accurate personality cue information from general personality questions and therefore make the interviewer better at detecting HH.

H3: General personality-tailored behavioural questions in combination with probing will lead to the most accurate HH detection as compared to any other condition.

Contributions

This study is going to contribute to the current literature by investigating what questions could increase the detection of HH in interviews. It demonstrates the effect of general compared to specific questions as well as the effect of probes on the detection of HH. It also demonstrates which combination of questions and probes is best for overall HH detection. In addition, because accurate cues from Study One could not be attained, the videos in Study Two will also be examined for cues of HH in the interview.

Methods

Participants. Participants were recruited for this study through Amazon Mechanical Turk (M-Turk). Based on a power analysis, 788 participants were needed for this study (using a
“small” Cohen’s d of 0.2). After data cleaning (see Table 5), 933 participants were obtained. Out of this sample, 42% identified as female, 57% identified as male and 1% identified as other. Ages ranged from 18 to 70. The participants indicated they identified as the following: 696 as Caucasian (75%), three as Indigenous (.3%), 73 as Black/African/Caribbean (8%), one as Arab (.01%), 12 as South Asian (1%), 53 as Southeast Asian (6%), 42 as Latin American (5%), and 53 as other (6%). The education level of the participants varied widely. One participant had some high school education (.01%), 111 had completed high school (12%), 223 had some college or university education (24%), 24 had apprenticeship training/training within trades (3%), 382 completed college or university education (41%), 44 had some graduate education (5%), 107 have completed graduate education (11%), 36 had professional degrees (4%) and 5 did not indicate their education level (1%). Three hundred and five participants had previous experience giving interviews (35%) while 603 did not (67%). Five participants did not respond to this demographic item.

**Procedure.** Participants were provided a link to the study’s survey on M-Turk (see Appendix E for the M-Turk HIT Description). When participants first opened the survey, they completed an online consent form (see Appendix F). Once they provided their consent, they completed an audio check, which involved the participant listening to a short, 2-second sound clip. This audio check was provided to ensure the participants’ sound was working and that they would be able to hear the videotaped interviews that were to follow and therefore be able to provide ratings of the candidates’ HH. Participants were first asked if their speakers were on, and then were able to play the clip. They then answered a question corresponding to what they heard in the clip. If they did not answer the question correctly the survey ended. Compensation was not rewarded to participants who did not pass the audio checks.
Once participants passed the audio checks, they were given descriptions of HH and each of its facets. HH was labelled as ‘Composite Trait’, to deter participants from focusing on the concept of honesty when rating the candidates in the videos. This strategy was employed to help to ensure that participants focused on making ratings based on the definitions provided to them.

Following this, the participants were randomly assigned to one of the four conditions based on a 2x2 study design. The conditions were split based on question type (general vs specific) and probing (probing vs no probing). Once the participant was randomly assigned to one condition, they watched the five videos that corresponded with their condition. They rated each interviewee’s level of HH by providing a general HH score and a score for each of the HH facets (Sincerity, Fairness, Modesty and Greed-Avoidance). The ratings were done on an RPM scale. The definitions of HH and its facets were presented above each scale to ensure that the participants did not forget the definition. They were also given two example cues for each facet to help clarify how each facet may be presented in the interview. These cues were extracted from SME cues from Study One, based on cues that fit most closely with the definition of each of the facets. See Appendix G for the descriptions and cues provided to the participants.

Following the rating of each video, attention checks were presented that corresponded to the video’s content (e.g., what was the candidate’s job?). These attention checks were used to make sure the participants paid enough attention to be able to rate the candidates HH. If the attention check was not answered correctly, the survey would end. Participants had to pass all five attention checks to receive payment and for their data to be analyzed. Once all the attention checks were passed, participants completed demographic questions and received $4.00 for compensation. Participants took, on average, a total of 33 minutes to complete the survey.
**Interview Videos.** Seventeen mock interviews were videotaped of candidates answering four interview questions (see Appendix H for the mock job they applied for). They answered two general personality-tailored behavioural questions and two behavioural questions tapping specifically into HH (see Appendix I). Each question was followed by a set of standard probes (see Appendix I). Following the interview, the candidates each completed a self-report of personality and rated their personality on an RPM scale. These measures acted as estimated true score measures for this study. For the consent and media release forms completed by these candidates and the ads used to recruit them, see Appendix J, K and L.

SMEs watched these videos and provided expert ratings on an RPM scale as well as wrote down cues that they noticed while watching the videos. These ratings were used as a form of estimated true score measurement and the cues they provided were used to investigate useful cues of HH in the interview.

The 17 videotaped interviews were narrowed down to a total of five (for the process of selecting these videos see Appendix M). Each of the five videos were split into four clips, each for one condition. First the videos were split based on question type (general vs specific), and then videos were split to either contain probes or no probes; this resulted in four videos total per videotaped interview. Four conditions (each one containing all five candidates) were therefore created: general questions and no probes, general questions and probes, specific questions and no probes, specific questions and probes (for a more in-depth explanation of the video selection and candidate demographics see Appendix M).

**Measures.**

*Relative Percentile Method (RPM) Scale.* The scale used by participants to rate the candidates in the videos was an RPM scale, similar to the one used in Study One. This scale
ranges from 0-100 (0 = very low, 50 = average, 100 = very high). Participants rated the candidates by deciding whether they thought they were very low, average, or very high on HH and each of its facets. The RPM scale type has been judged as advantageous in past research (Goffin et al., 2009).

Demographics. Participants were asked to complete questions corresponding to five demographic variables: age, education, gender, ethnicity and interview experience. For interview experience, participants were asked if they had any experience with interviewing as well as the context of their interviewing, and approximately how many interviews they had conducted in the past.

Analytic Approach

In addition to examining the hypotheses of Study Two, the videos created in Study Two were also used to investigate cues of HH. SMEs in Study One did not produce great enough accuracy for the HH cues to be examined and therefore Study Two’s videos will be used to answer R3 from Study One: what are cues of HH presented in the interview?

Participant Accuracy. Participant accuracy was assessed based on the Schmid Mast et al. (2011) definition used in Study One. These specific trait correlations were calculated between an interviewee’s measure of estimated true score and the participant’s RPM rating for HH. Each of the facet’s accuracy correlations were also calculated in this manner. Three different measures of estimated true score were used for HH and its facets, self-report scores from the HEXACO-200, self-report RPM ratings and SME RPM ratings, and therefore three different participant accuracy correlations were calculated for each facet and HH. By using the self-reported RPM scores as an estimated true score measure, the participants’ ratings were compared to self-reports on a similar scale. Also, when SME ratings were used as a measure of the estimated true score
the accuracy correlation that was obtained was specific to the interview context, as opposed to HEXACO-200 score which is based on the interviewee’s knowledge of their behaviour over their entire lifetime. The facets were included in exploratory analyses to explore if certain facets were harder to judge than others.

**Condition Comparison.** Five cell comparisons were conducted, using t-tests, to test hypotheses one through three. The cells that were compared are depicted in Table 6. To test H1 (Interviewers will have greater accuracy in detecting HH in response to general personality-tailored behavioural interview questions as compared to behavioural interview questions that tap specifically tap into HH), two cell comparisons were conducted. First, and separately, cell one (general, no probes) and three (specific, no probes) were compared, and then cell two (general, probes) and four (specific, probes). For this hypothesis to receive complete support, cell one must have a significantly higher mean participant accuracy correlation compared to cell three, and cell two must have a significantly higher mean participant accuracy correlation than cell four. Achieving these differences would demonstrate that general questions were better than specific questions when probes were present, as well as when they were not present.

To test H2 (The use of probes (versus no probes) will lead to increased accuracy in HH detection), two cell comparisons were conducted. These comparisons were between cell one (general, no probes) and two (general, probes), and then cell three (specific, no probes) and four (specific, probes). For hypothesis two to receive full support, cell two must have a significantly higher mean participant accuracy correlation compared to cell one, and cell four must have a significantly higher mean participant accuracy correlation than cell three. This difference would demonstrate that by asking probes, accuracy increased, both when the initial interview question was a specific question and when it was a general question.
To test H3 (general questions with probes are best), three cell comparisons were considered. First, cell one (general, no probes) and cell two (general, probes), then cell two (general, probes) and cell four (specific, probes), and finally cell two (general, probes) and cell three (specific, no probes). The final cell comparison, cell two and three, was conducted solely to test this hypothesis, while the other cell comparisons help to test other hypotheses as well (as stated above). For this hypothesis to be supported, cell two would have to have a significantly higher mean participant accuracy correlation in all the cell comparisons. This difference would demonstrate that the general probe condition is indeed better than each other condition.

Bonferroni corrections were run on the planned contrasts to reduce the chance of Type 1 error occurring when running multiple comparisons. Fifteen comparisons were corrected for, to account for all the planned contrasts run.

**SME Accuracy.** Accuracy of HH detection was the same as accuracy of SMEs in Study One. Accuracy correlations will be calculated for HH and each of its facets.

**Cues.** Like Study One, SMEs in Study Two must have accuracy correlations of greater than .5 with the candidates’ self-reports for the cues to be sorted and categorized. The cues will help to demonstrate cues that could be useful in an interview to detect HH and its facets.

**Results**

Multiple planned contrasts were conducted to test the hypotheses. The omnibus interaction tests were significant for each of the outcome variables (the participant accuracy correlations with each of the estimated true scores), demonstrating that the planned contrasts should be conducted, $F_{\text{HEXACO-200}}(1, 929) = 47.74, p < .001$, $F_{\text{SelfRPM}}(1, 929) = 35.57, p < .001$, $F_{\text{SME}}(1, 929) = 23.35, p < .001$. See Table 7 for cell means.
**Condition Comparison.**

**Question Type.** Question type was examined to test H1: interviewers will more accurately detect HH when interviewees answer general personality-tailored behavioural interview questions as compared to behavioural interview questions that tap specifically into HH.

**Self-Reports (HEXACO-200).** H1 was first examined using self-reports from the HEXACO-200 as a measure of a candidate’s estimated true score. In the probe condition, HH detection accuracy was higher in the general question condition ($M = .31, SD = .48$) than the specific question condition ($M = .06, SD = .46$), $d = .53, 95\% CI [.34, .73], t(410) = 5.41, p < .001$. Therefore, general questions led to greater HH detection accuracy than specific questions when probes were asked. When examining question type for the no-probe condition, the general question condition ($M = .03, SD = .46$) did not lead to greater HH detection accuracy when compared to the specific question condition ($M = .21, SD = .48$), $d = -.38, 95\% CI [-.55, -.21], t(519) = -4.32, p = 1.00$. Therefore, H1 received partial support, but only for the probe condition.

**Self-Reports (RPM).** H1 was then tested using self-reports from the RPM scale as a candidate’s estimated true score. In the probe condition, the general question condition led to greater HH detection accuracy ($M = .33, SD = .46$) than the specific question condition ($M = .08, SD = .47$), $d = .52, 95\% CI [.32, .72], t(410) = 5.28, p < .001$. However, in the no-probe condition, the general question condition ($M = .07, SD = .45$) did not lead to greater HH detection accuracy compared to the specific question condition ($M = .19, SD = .50$), $d = -.27, 95\% CI [-.44, -.10], t(501) = -3.04, p =1.00$. These overall findings replicate what was found using the HEXACO-200 as a measure of a candidate’s estimated true score. Therefore, H1 received support in the probe condition, but not in the no-probe condition.
SME Reports. Finally, H1 was tested using SME reports as the estimated true score measure. In the probe condition, the general question condition ($M = .31, SD = .48$) led to greater HH detect accuracy compared to the specific question condition ($M = .11, SD = .47$), $d = .42$, 95% CI [.22, .61], $t(410) = 4.25, p < .001$. In the no-probe condition, the general question condition ($M = .11, SD = .48$) did not lead to greater HH detection accuracy when compared to the specific question condition ($M = .22, SD = .48$), $d = -.22$, 95% CI [-.39, -.05], $t(519) = -2.54$, $p = 1.00$. Again, these findings replicate the findings above (when using self-reports as the measure of estimated true score). H1 received partial support, but only in the probe condition.

Probes vs No Probes. The effect of probes was examined to test H2: asking probes will increase an interviewer’s ability to accurately detect HH.

Self-Reports (HEXACO-200). H2 was first tested using self-reports from the HEXACO-200 to measure a candidate’s estimated true score. In the general question condition, the probe condition ($M = .31, SD = .48$) led to greater HH detection accuracy when compared to the no-probe condition ($M = .03, SD = .46$), $d = .60$, 95% CI [.41, .78], $t(469) = 6.41, p < .001$. Therefore, greater HH detection accuracy occurs with general questions when probes are asked as compared to when they are not. In the specific question condition, the probe condition ($M = .06, SD = .48$) did not increase HH detection accuracy over the no-probe condition ($M = .21, SD = .48$), $d = -.32$, 95% CI [-.50, -.13], $t(460) = -3.38, p = 1.00$. Therefore, H2 received partial support, but only in the general question condition.

Self-Reports (RPM). H2 was next tested using self-reported RPM scores as the measure of estimated true score. For the general question condition, the probe condition ($M = .33, SD = .46$) led to greater HH detection accuracy compared to the no probe condition ($M = .07, SD = .45$), $d = .57$, 95% CI [.38, .76], $t(469) = 6.12, p < .001$. However, in the specific question
condition, the probe condition \((M = .08, SD = .47)\) did not lead to greater HH detection accuracy when compared to the no-probe condition \((M = .19, SD = .50)\), \(d = -0.23\), 95% CI \([-0.41, -0.04]\), \(t(460) = -2.45, p = 1.00\). These findings replicate the results found using the self-report scores from the HEXACO-200. Overall, H2 received partial support, but only for the general question condition.

SME Reports. Finally, H2 was tested using SME reports as the measure of estimated true score. In the general question condition, the probe condition \((M = .31, SD = .48)\) had greater HH detection accuracy compared to the no-probe condition \((M = .11, SD = .48)\), \(d = 0.42\), 95% CI \([0.23, 0.60]\), \(t(469) = 4.50, p < .001\). However, in the specific question condition, the probe condition \((M = .11, SD = .47)\) did not lead to greater HH detection accuracy when compared to the no-probe condition \((M = .22, SD = .49)\), \(d = -0.22\), 95% CI \([-0.40, -0.03]\), \(t(460) = -2.34, p = 1.00\). These findings replicated those produced in the comparisons using self-reports as a measure of estimated true score. Therefore, H2 again only received partial support, for general questions specifically.

Condition Effectiveness. The cell of general questions with probes was compared to all other cells to test H3: general personality-tailored behavioural questions in combination with probing will lead to the most accurate HH detection.

Self-Reports (HEXACO-200). H3 was first tested using the self-reports from the HEXACO-200 as the measure of estimated true score. Based on the findings presented above, the general question, probe condition \((M = .31, SD = .48)\) led to greater HH detection accuracy over the specific question, probe condition \((M = .06, SD = .48)\), and the general question, no probe condition \((M = .03, SD = .46)\). However, the general question, probe condition was not
significantly different from the specific question, no probe condition ($M = .21$, $SD = .48$), $d = .21$, $95\% CI [.02, .40]$, $t(446) = 2.20$, $p = .21$. Therefore, H3 was not supported.

**Self-Reports (RPM).** H3 was then tested using the self-reported RPM ratings as the measure of estimated true score. Based on the findings presented above, the general question, probe condition ($M = .33$, $SD = .46$) led to greater HH detection accuracy compared to the specific question, probe condition ($M = .08$, $SD = .47$), and the general question, no probe condition ($M = .07$, $SD = .45$). In addition, the general question, probe condition was significantly different from the specific question, no probe condition ($M = .19$, $SD = .50$), $d = .28$, $95\% CI [.09, .46]$, $t(446) = 2.89$, $p = .03$. Therefore, H3 was supported.

**SME Reports.** H3 was finally tested using SME reports as the measure of estimated true score. Based on the findings presented above, the general question, probe condition ($M = .31$, $SD = .48$) led to greater HH detection accuracy over the specific question, probe condition ($M = .11$, $SD = .47$), and the general question, no probe condition ($M = .11$, $SD = .48$). However, the general question, probe condition was not significantly different from the specific question, no probe condition ($M = .22$, $SD = .49$), $d = .19$, $95\% CI [.01, .38]$, $t(446) = 2.04$, $p = .31$. Therefore, H3 was not supported.$^{11}$

**Exploratory Analyses.** As hypotheses in this research were only made at the overall HH level, the facets were analyzed in an exploratory manner. The omnibus interaction tests using the self-reports from the HEXACO-200 as the estimated true score produced significant results for all facets and therefore cell comparisons were run for each of the facets, $F_{Sin}(1, 929) = 22.93$, $p < .001$, $F_{Fair}(1, 929) = 33.30$, $p < .001$, $F_{GA}(1, 929) = 14.64$, $p = .002$, $F_{Mod}(1, 929) = 16.13$, $p = .001$. The omnibus interaction tests were also significant for all facets when SME-reports were

---

$^{11}$ These results were also examined by testing the differences between each of the accuracy correlations. For these results see Appendix N.
used as a measure of estimated true score, $F_{Sin}(1, 929) = 12.42, p = .007$, $F_{Fair}(1, 929) = 9.27, p = .04$, $F_{GA}(1, 929) = 14.66, p = .002$, $F_{Mod}(1, 929) = 14.71, p = .002$, allowing for cell comparisons to be run for all the facets. However, when the self-reports from the RPM scale were used, the omnibus interaction test was only significant for Fairness and therefore only comparisons for this facet were run, $F_{Fair}(1, 929) = 38.14, p < .001$. For cell means, refer to Table 9.

**Question Type.** To examine the effects of question type on each of the facets, exploratory two-sided t-tests were conducted. Cell one was compared to cell three, and cell two was compared to cell four.

**Self-Reports (HEXACO-200).** The first cell comparisons for question type were examined using the self-reports from the HEXACO-200 as a measure of the candidate’s estimated true score. When comparing cell one and three, the results for Sincerity, Greed, Avoidance and Modesty replicated the findings from HH (see Table 10). There was no difference, however, in question type for the no-probe condition for the facet of Fairness, as the confidence intervals crossed zero (see Table 10). When comparing cell two and four, the results from overall HH were replicated for Fairness and Sincerity (see Table 10). It should be noted, however the lower end of the confidence interval for Sincerity is very close to zero. The other facets produced negligible effect sizes (see Table 10).

**Self-Reports (RPM).** Question type was next examined using self-reports from the RPM scale as the measure of estimated true score. When comparing cell one and three, there was no difference for Fairness as the effect was very small and had confidence intervals that overlap zero (see Table 10). When comparing cell two and four, the results of overall HH were replicated (see Table 10).
SME-Reports. Finally, question type was examined using SME reports as the estimated true score. When comparing cell one and three, the results of overall HH were replicated for Greed-Avoidance and Modesty (see Table 10). Sincerity produced an effect with confidence intervals that overlapped zero and Fairness had a lower end confidence interval of zero and therefore the effects are likely small or not present (see Table 10). When comparing cells two and four, the results of overall HH were replicated for Sincerity and Fairness. It should be noted, however that the confidence interval for Fairness approaches zero. The other facets produced confidence intervals that overlapped zero and therefore, the effects are likely not present (see Table 10).

Probes vs No Probes. To examine the effects of probes, exploratory two-sided t-tests were run on each of the facets. Two cell comparisons were conducted: cell one with cell two, and cell three with cell four.

Self-Reports (HEXACO-200). The first cell comparisons were run using self-reports from HEXACO-200 as a measure of the candidate’s estimated true score to test the effects of probes. Comparing cell one and two, all facets replicated the results found for overall HH (see Table 11). When comparing cell three and four, the results for overall HH were replicated for Sincerity (see Table 11). For the other facets, they had confidence intervals that overlapped zero.

Self-Reports (RPM). The effect of probes was then tested using self-report ratings from the RPM scale as the estimated true score. Comparing cell one and two, the results for overall HH were replicated for Fairness. When comparing cell three and four, the confidence intervals for Fairness overlapped zero indicating that there is likely no effect present (see Table 11).

SME Reports. Finally, to test the effects of probes, SME reports were used as a measure of estimated true score. Comparing cell one and two, all facets replicated the results found for
overall HH (see Table 11). When comparing cell three and four, all the effects had confidence intervals that overlapped zero and therefore an effect is unlikely (see Table 11).

**Condition Effectiveness.** The cell of general questions with probes was compared to other cells to test if it is the best condition for each facet’s detection.

*Self-Reports (HEXACO-200).* The effectiveness of the general question, probe condition was first tested using the self-reports from the HEXACO-200 as the measure of estimated true score. Based on the findings presented above, for Sincerity and Fairness, the general question, probe condition led to greater HH detection accuracy over the specific question, probe condition, and the general question, no probe condition. For Fairness, when comparing the addition cells (cell two and three) the general question, probe condition also had greater facet detection accuracy over the specific question, no-probe condition (see Table 12). For Sincerity, when cell two and three were compared, the specific question, no-probe condition had greater facet detection accuracy over the general question, probe condition (see Table 12). Although it should be noted that the confidence interval approaches zero. For Greed-Avoidance and Modesty, the findings presented earlier demonstrate that the general question, probe condition produced greater facet detection accuracy over the general question, no-probe condition, but not over the specific question, probe condition. When cell two and three were compared, both facets had effect sizes with confidence intervals that overlapped zero and therefore an effect is unlikely (see Table 12).

*Self-Reports (RPM).* The effectiveness of the general question, probe condition was then tested using the self-reports from the RPM as the measure of estimated true score. For Fairness, the general question, probe condition led to greater HH detection accuracy over the specific question, probe condition, and the general question, no probe condition. In addition, when
comparing cells two and three the general question, probe condition had greater facet detection accuracy over the specific, no probe condition (see Table 12).

*SME Reports.* The effectiveness of the general question, probe condition was finally tested using the SME reports as the measure of estimated true score. Based on the findings presented above, for Sincerity and Fairness, the general question, probe condition led to greater HH detection accuracy over the specific question, probe condition, and the general question, no probe condition. For Sincerity, when comparing the addition cells (cell two and three) the general question, probe condition also had greater facet detection accuracy over the specific question, no-probe condition (see Table 12). However, it should be noted that the confidence intervals approach zero. For Fairness, when cell two and three were compared, the confidence interval overlapped zero and therefore it can be presumed that an effect is not likely present (see Table 12). For Greed-Avoidance and Modesty, the findings presented earlier demonstrate that the general question, probe condition produced greater facet detection accuracy over the general question, no-probe condition, but not over the specific question, probe condition. When cell two and three were compared, both facets had effect sizes with confidence intervals that overlapped zero and therefore an effect is unlikely (see Table 12).

**Cues.** As opposed to Study one, SMEs in Study Two produced some accuracy correlations greater then .5 (see Table 8). As a result, the cues used by SME three for all facets, other than Fairness, as well as SME six for the facets of Greed-Avoidance and Modesty were presented (Appendix O) to demonstrate cues that are useful to detect HH and its facets in an interview. Examples of cues for each of the facets are as follows: Sincerity, “I took advantage of the fact that he had kids to convince him to donate” and “I value relationships for much more than a way to get something out of someone”; Greed-Avoidance, “I want to give back to others”
and “I am passionate about the company and want them to do well”; and Modesty, “I recognize that I was wrong in the situation” and “I was upset because I thought I was right and was getting told the opposite”. This study only failed to provide useable cues for the facet of Fairness. However, because no accuracy correlations produced for Fairness were above .5 the cues provided by SMEs cannot be judged as accurate.

Discussion

Study Two tested three hypotheses. H1 proposed that interviewers would be the most accurate in detecting HH when candidates respond to general personality-tailored behavioural interview questions as compared to behavioural interview questions that tap specifically into HH. H2 proposed that the use of probes would lead to increased HH detection accuracy. And finally, H3 proposed that greatest HH detection accuracy would result from asking general personality-tailored behavioural questions and probes. This study found that although when probes are asked, asking general questions will lead to increased HH detection accuracy, when probes are not asked, general questions do not. By examining the effect size and its direction, it appears that specific questions actually result in better HH detection than general questions when probes are not asked. The effect was also smaller when examining specific questions. Therefore, H1 was only supported for the probe conditions. When examining the use of probes, this study demonstrated that probes are most effective for general questions, but not specific questions. By examining the direction of the effect size, it appeared that it is better to not ask probes following specific questions to achieve greater HH detection accuracy. These effects gave partial support to H2, but only for general questions. Finally, this study demonstrated that when detecting overall HH, two of the conditions were similar in their ability to increase HH detection accuracy. General questions in combination with probes, as well as specific questions without probes were
not significantly different from one another in their ability to improve HH detection in the interview, for self-reports using the HEXACO-200 and with SME reports, rejecting H3. However, H3 was supported for when self-reports using the RPM scale were the measure of estimated true score.

This study also allowed interview specific HH cues to be further explored. Two SMEs provided accuracy scores of over .5 and therefore helped to provide cues that can be used in the future for the detection of HH in job interviews. Fairness, however appeared quite hard to detect accurately and therefore cues were not gleaned from this study for the facet of Fairness.

**Exploratory Analyses.** When the facets were explored, and effects were present, they often replicated the results of overall HH, but varied slightly depending on the measure of estimated true score used. Most of the variations in findings were present when examining the condition that was best for detecting each facet. However, if one or two conditions led to the greatest accuracy, the conditions were general questions with probes or specific questions without probes. It appears that there may be some nuances about each of the facets that affect their detection in the interview, compared to overall HH.

Overall, all of these results demonstrate that when trying to detect HH and its facets, aside from the facet of Fairness and HH when self-report RPM scores were used (which are best detected when candidates are asked a general question and followed up with probes), that general or specific questions can be used to detect HH, but these question types must be used specifically in conjunction with probes and no-probes, respectively.

**General Discussion**

This research used Funder’s Realistic Accuracy Model (RAM; 1995) to answer the question of: can we detect Honesty-Humility in the employment interview? To answer this
question, two studies were conducted. Study One tested all four stages of the RAM by examining possible cues of HH in employment interviews (tapping into the relevance and availability stages of Funder’s Model) and by assessing the reliability and validity of Subject Matter Experts’ ratings of HH (tapping into the detection and utilization stages of Funder’s Model). Study Two then tested the relevance and availability stages further by investigating if general personality-tailored behavioural questions allowed interviewers to detect HH more readily than HH-specific questions and if probing increased HH detection accuracy. This research built upon previous research conducted by Powell and Goffin (2009) and Powell and Bourdage (2016) who investigated the detection of personality traits in the job interview, as well as answered the call for research put out by both Morgeson et al. (2007) and Levashina et al. (2014).

The results of this research determined, using reliability analyses, that SMEs were able to reliably rate HH, however the validity of these ratings varied. In Study One, none of the SMEs were able to achieve accurate judgements, but of the traits rated, SMEs most accurately rated HH and the facet of Modesty. In Study Two, only two of the SMEs were able to obtain accurate judgements of HH, and neither of these SMEs were able to provide accurate ratings for the facet of Fairness. For Study Two, Greed-Avoidance and Modesty were judged most accurately. This finding lends itself to the assumption that candidates may not view Greed-Avoidance and Modesty as traits that are judged negatively (i.e. how these facets are evaluated might be different than other facets of HH), and therefore they readily emit cues of these facets in the interview. However, by examining the cues recorded by the SMEs, cues were mostly recorded for Sincerity and the least number of cues were recorded for Greed-Avoidance. Therefore, there may be something about how the trait of Greed-Avoidance is presented by a candidate in an interview that affects accuracy without needing many cues to be presented. For example, cues
presented may have a stronger salience and effect on judges, causing the judges to more accurately rate Greed-Avoidance. Also, there may be cues, that were not recorded by the SMEs that may have also played a role in the SMEs judgement. Future research should examine Greed-Avoidance and Modesty, to determine what makes them easier to detect in the interview.

It should be noted, that although accuracy of over .5 was only achieved by certain SMEs, many of the SME ratings were able to reach a correlation of .30 which is equal to the correlation obtained when friends rate each other’s HH (de Vries et al., 2008). In some cases, the SMEs were even able to achieve a correlation that approximated that of what romantic partners can rate each other (a correlation of 0.60; de Vries et al., 2008). Therefore, even if the correlations appear to be low, they are quite good given the correlations produced when individuals know each other very well.

The findings on accuracy and reliability from both studies demonstrate three important things. First, they provide support for previous research that stated that HH is hard to detect (Ashton & Lee, 2010). Second, Fairness appears to be quite hard to judge accurately in the interview. This difficulty may be a particular issue, as Fairness is likely an important trait for job performance, especially for jobs that involve being in a managerial position. Managers need to be fair and be judged as fair by their employees for their employees to be satisfied and committed to the organization (Simons & Roberson, 2003), therefore it would be an important facet to hire for. The questions asked in the interview in Study Two may have affected the ability for Fairness to be elicited and detected. The questions that were asked tapped into conflict, persuasion, building long-term relationships and obtaining a personal benefit through a relationship (see Appendix I). These questions appear to provide the opportunity to tap into the facet of Fairness, however if candidates were focussed on other aspects of the question they may
not have emitted cues specific to their Fairness level. Referring to the RAM, without the emission of cues from candidates in the availability stage, accurate detection cannot occur (Funder, 1995). Therefore, the answers the candidates gave in response to the questions may have affected the SMEs ability to accurately rate this facet. As a result, future research should examine how questions can be asked to improve Fairness detection accuracy and why Fairness is a particularly hard facet to judge. Opposing the previous explanation, but plausible as well, is that the candidates may have been aware of the importance of Fairness in social interactions (e.g. the interview) and therefore each tried to appear fair in the job interview in response to the questions. This kind of responding would reduce variability between candidates and make all candidates appear high on Fairness. As a result, the accuracy of Fairness would not be judged accurately, as not all the candidates are high on this facet. When examining the ratings given by SMEs, it appears as though a slightly higher degree of candidates were given high scores on Fairness as compared to other traits, giving this explanation further plausibility. Third, contrasting the results found in Study One and Study Two, support is provided for ensuring that good questions are asked in job interviews. It was only when questions were used that were tailored for this research that some of the SMEs were able to achieve a moderate level of accuracy in judging HH and its facets.

In addition to the findings on reliability and accuracy, this research demonstrated that, when trying to detect a candidate’s overall level of HH, general questions are best for HH when they are followed up by probes. However, when specific questions are asked, raters will achieve higher HH detection accuracy when probes are not presented. These results appear to support the concept of situation strength (as defined earlier by Meyer, Dalal & Hermida, 2010), but may also relate to trait activation theory. It appears that, because general questions with probes and
specific questions without probes were not significantly different in their ability to elicit cues, either question could be asked to achieve effective HH detection. For general questions it appears that they allow the interviewer to create a weak situation that does not signal to the candidate how they should answer. However, because general questions are not very effective without probes, follow-up questions must be asked after a general question to either elicit more information from the candidate (naturally allowing the candidate to emit HH cues) or to guide the candidate to elicit the cues necessary for HH detection. When specific questions are used, the opposite is true. Specific questions appear to tap into trait activation theory, which states that for a trait to be expressed it must be activated by trait relevant cues (Tett & Guterman, 2000). In other words, specific questions allow for a trait to be elicited because the candidate becomes aware of what information is necessary to answer the question, and when probes are not asked, these questions allow cues to be targeted in a short amount of time. However, when probes are asked, the ability of a rater to accurately detect HH declines. This decline may occur because, when more questions are asked through probing, more opportunities are presented for the candidate to fake as they are now aware of what the interviewer is looking for and can try to tailor their answer further to appear in a positive light and get the job. Future research should examine if this explanation holds true by investigating if specific questions allow for more candidate faking. This research would help to further understand why probes for specific questions led to poor HH detection. Overall, the findings support the relevance and availability stages of Funder’s (1995) model, as they demonstrate the importance of the emission of cues for accurate personality trait detection.

For overall HH, both general personality questions with probes and specific questions without probes appear to provide better opportunities for HH detection. When self-reports from
the HEXACO-200 and SME reports were used as a measure of the estimated true score; these conditions were not significantly different, however when self-reports from the RPM scale were used the conditions were significantly different and general questions with probes were the best condition compared to the other three. That being said, although a significant difference was not present for some of the estimated true score measures, the accuracy correlations for general questions with probes condition was the highest. Therefore, it may be best to use general questions with probes when trying to detect HH in interviews. This research should be replicated to ensure that these results hold true. It should also be noted, that if HR managers are looking to hire people based on specific facets of HH, different types of questions may need to be used. As stated following Study Two, the findings for the facets were not as clear as the findings for HH. It appears that for some facets, asking general questions with probes is best, while for others specific questions with no probes is best. In addition, for facets like Greed-Avoidance and Modesty, no one condition or even two conditions were best. In cases like these, future research needs to be done to delve into the differences between the facets and how best to elicit them individually. As the results based on the facets were exploratory, future research should examine the facets in a confirmatory manner to explain these findings as well.

Another crucial aspect of this research is the accuracy of interviewers. When examining the accuracy correlations produced by each condition, overall the mean accuracy correlations appeared to be small. However, though they may be quite small in certain conditions, the general question with probes condition was able to achieve the correlation that is obtained when friends rate each other’s HH (0.3; de Vries et al., 2008). This correlation is quite impressive considering the videos were often quite short (with some of the videos in this condition being no more than three minutes long). Therefore, accuracy would likely be even greater if the videos were of a
greater length. That being said, this accuracy could still be improved, and from a practical standpoint, this correlation should ideally be higher if we are suggesting HH be detected within the interview in organizations. One thing that should be considered, therefore, is training. The participants rating the videos were only provided small amounts of information about the trait of HH and each of its facets. Therefore, it may be understandable as to why the accuracy correlations were quite low in many cases. It should be considered that the SMEs, who were provided training, were able to obtain higher accuracy correlations than the participants. Therefore, future research should examine the accuracy that can be obtained when using an effective interview question method while also providing raters with training. This research would be able to look at a greater sample size than was provided by the SMEs as well.

Individual difference variables are something that should be also examined. Evidently, there were some SMEs that were more accurate at detecting HH than others, and therefore there may have been traits or attributes of those SMEs that made them more accurate. Future research should investigate what individual difference variables make an individual better at detecting HH in the job interview, such as working memory or cognitive ability. These variables could be important when looking to select interviewers to ensure that they can most accurately detect this trait. Previous research by Powell and Bourdage (2016) found that dispositional intelligence was important in detecting personality, and therefore there are likely more individual differences that can affect personality detection, especially for a difficult trait to detect such as HH.

Another important thing to note, is that three different types of estimated true scores were used in Study Two. Although these different measurement methods did not cause many differences for overall HH, differences were produced for the facets. These differences may be a result of the kind of estimated true score used. Self-reports from the HEXACO and RPM are
broad personality measures, so when the candidate is reflecting on their personality, it is based on their behaviour over their lifespan. However, the SME reports were context specific; they were based on how personality was portrayed in the job interview. Therefore, although the SME accuracy was not always overly high, it may be a good judgement of interview specific personality. In addition to this, it is hard to say how accurate people are at being able to rate themselves on an RPM scale. This method requires a greater level of self-reflection than answering items within the HEXACO-200 as the HEXACO-200 focuses more specifically on like, dislikes and specific behaviours, as opposed to rating yourself on a broad definition. The self-report scores on the RPM scale did differ greatly in some instances from those on the HEXACO-200, which may give some air of caution when interpreting these results compared to the other measures of estimated true score.

**Limitations and Future Research**

**Study One.** As with all research, these studies were not without their limitations. The biggest limitation of Study One is that the cues were unable to be deemed accurate for use by interviewers. Although cues were recorded for all 82 videos, because none of the SMEs were able to achieve an accuracy correlation of .5, the cues were not usable. This sample was unique because it would have provided a sample of cues from a high-stakes interview, but unfortunately, given the accuracy correlations, this was not the case. The accuracy may, however, have been affected by the questions asked in the study, and they may not have elicited enough cues from the candidates to allow for accurate HH detection. Another limitation was that all the SMEs were female and of a similar age. These similarities may have increased the likelihood of them producing similar and reliable scores. It is important for future research to ensure that good reliability values can be obtained from interviewers that may be less similar from one another.
**Study Two.** The experimental nature of this design helps to give strength to the results found in this research. However, there are limitations to this study as well. The interviews used for this study were mock interviews and therefore, the answers provided by the candidates may have differed slightly from how the candidate would respond in a real interview. Although, given the low stakes situation of these interviews, the results found from this study likely demonstrate the best that can be expected from interviewers in detecting HH as the candidates are likely being most honest in this type of interview situation. Future research should examine accuracy of HH detection in more high stakes interviews. The candidates in these videos also had a small range of HH scores. The normal population would have scores ranging from one to five on the HEXACO scale for HH, whereas the participants ranged from only 2.00 to 3.91. This range may have affected the mean accuracy of scores as those who are very high or very low on HH may be easier to judge. Although these extremes are rarer in the general population, because there was not as much range as would be expected in the normal population, participant accuracy may have been affected and they may have rated the candidates in the videos with more of a general population range in mind. It should be noted, however, that it may take a very large sample to be able to obtain scores on the far ends of the distribution.

Of the candidates in the videos, the males both scored lower on HH than the females (see Appendix M). Although this is a general trend in the population (Lee & Ashton, 2004), future research should be conducted with more of a mixed sample to ensure that ratings of each gender were not judged accurately simply as a result of a particular gender stereotype; females will not always be higher on HH than males in a selection system.

Only judging HH accuracy based on five videos also limits the generalizability of this study. Research would need to replicate these findings to ensure they can be generalized to the
broader population. As well, the candidates were university students and may not have as much experience as others in the workforce. This lack of experience could affect their interviewing skill, which may, in turn, affect their responses and consequently, the cues they emit. Therefore, it is important for this research to be replicated with more experienced job candidates to see if there is any difference in results.

Given that the participants were recruited for this study from M-Turk, they may not have paid as close attention as an actual interviewer when trying to detect a candidate’s HH. This lack of attention would likely have decreased the accuracy of the participants as they would not be able to notice all of cues emitted by the candidates. Attention checks were included to try to ensure that participants were paying close enough attention to notice specific details in the video, and therefore also paying attention to HH cues, however future research should examine this phenomenon with real interviewers to judge their levels of HH detection accuracy. The interviews included in this study also only included two questions per condition. This number of questions limits the ability to explore the effects of these questions. In addition, the accuracy correlations obtained are a factor of the questions asked, so future research should look at asking different questions of the same type and see if the results replicate and examine the accuracy of the raters. Another consideration for the accuracy of the participants is that the videos they watched were quite short. Across the conditions the longest video was approximately five minutes and some videos were less than a minute. The videos for this study had to be short so the attention span of the participants was not a great concern, as it was an online study, however real interviews would evidently be much longer than this. Therefore, the accuracy of detecting HH would likely be higher when an interviewer has more time to spend with the candidate (as time
spent with an individual and HH judgement accuracy are correlated; Ashton & Lee, 2010), but future research should examine this phenomenon.

This research provided a unique contribution to the literature by providing cues that can be used to detect HH in the interview. However, there are limitations in accumulating these cues as well. First off, this study was not able to directly detect the best cues of HH. The cues that have been deemed helpful for detecting HH in interviews have been inferred from what was recorded by the most accurate SMEs. There may have been things that influenced their judgements, not tied to these cues, that were not tapped into as a result. Future research should examine less inferential ways to judge accurate cues, such as candidate watching their interview and pointing out cues that they emitted that are indicative of their HH level. Second, given the nature of the interview, the cues elicited are specific to the questions asked. It is impossible to detect all cues of HH that could be emitted in the interview and this study will only provide a small selection of possible cues. Future research should continue to examine cues to detect HH in interviews, especially for the facet of Fairness as cues for this facet were not reported in this research. Third, all the SMEs were female in this study because of which students were willing to volunteer their time. Cues that females specify, and notice, related to HH may be different than those detected by males. However, because general HH training was provided, hopefully this was not the case. Regardless, future research should also include male SMEs when trying to determine further cues of HH in the interview.

**Implications**

Despite possible limitations, this research has a variety of implications. Both studies provided theoretical contributions by offering support for the RAM. By examining the importance of reliability and accuracy, Study One demonstrated support for the detection and
utilization stages of the model. For Study Two, by demonstrating that the types of questions asked in the interview effect HH detection accuracy, the relevance and availability stages of the model received support. There are also practical implications of this research. This research demonstrated the type of questions that should be asked in interviews to best elicit HH and indicated under which question conditions probing should be used to increase HH detection accuracy. It also demonstrated the conditions under which HH detection is most accurate. Finally, it helped to uncover cues for the facets of Sincerity, Greed-Avoidance and Modesty that can be used in the future by interviewers to detect HH.

This research helps to contribute to the body of literature on personality detection, while adding unique insights to how HH can be detected in the interview. As research continues to flourish and increase HH detection, the interview may become a viable and effective way to detect HH in organizational selection systems.
References


Table 1

*Descriptions of HEXACO Personality Traits*

<table>
<thead>
<tr>
<th>HEXACO Traits</th>
<th>Low Levels</th>
<th>High Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honesty-Humility</td>
<td>Greedy, corrupt, cunning and vindictive</td>
<td>Sincere, loyal, modest and fair</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>Anxious, vulnerable and emotionally sensitive</td>
<td>Self-assured, brave, independent</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Reserved, shy and withdrawn</td>
<td>Cheerful, talkative and lively</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Irritable, stubborn and impatient</td>
<td>Calm, patient and friendly</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Disobedient, frivolous and careless</td>
<td>Meticulous, self-disciplined and organized</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>Narrow-minded, conventional and conservative</td>
<td>Creative, inventive, imaginative</td>
</tr>
</tbody>
</table>
Table 2

*Study One Overall HH and Facet Correlations between Self-Reports and SME Ratings*

<table>
<thead>
<tr>
<th>Specific Personality Trait Accuracy</th>
<th>SME 1</th>
<th>SME 2</th>
<th>SME 3</th>
<th>SME 4</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self Honesty-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Humility</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Sincerity</td>
<td>-0.09</td>
<td>-0.08</td>
<td>0.09</td>
<td>0.02</td>
<td>-0.004</td>
</tr>
<tr>
<td>Self Fairness</td>
<td>-0.01</td>
<td>0.11</td>
<td>0.16</td>
<td>-0.06</td>
<td>0.09</td>
</tr>
<tr>
<td>Self Greed-</td>
<td>-0.01</td>
<td>-0.09</td>
<td>0.05</td>
<td>0.10</td>
<td>0.03</td>
</tr>
<tr>
<td>Avoidance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self Modesty</td>
<td>0.14</td>
<td>0.21</td>
<td>0.30**</td>
<td>0.14</td>
<td>0.28*</td>
</tr>
</tbody>
</table>

* * indicates p < .05, ** p < .01, N = 82
Table 3

Intraclass Correlations between SMEs for HH and Its Facets in Study One

<table>
<thead>
<tr>
<th>Facet</th>
<th>Measure</th>
<th>ICC</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honesty-</td>
<td>Single Measure</td>
<td>.39</td>
<td>.29</td>
<td>.51</td>
</tr>
<tr>
<td>Humility</td>
<td>Average Measures</td>
<td>.72</td>
<td>.61</td>
<td>.81</td>
</tr>
<tr>
<td>Sincerity</td>
<td>Single Measure</td>
<td>.43</td>
<td>.31</td>
<td>.55</td>
</tr>
<tr>
<td></td>
<td>Average Measures</td>
<td>.75</td>
<td>.64</td>
<td>.83</td>
</tr>
<tr>
<td>Fairness</td>
<td>Single Measure</td>
<td>.26</td>
<td>.15</td>
<td>.39</td>
</tr>
<tr>
<td></td>
<td>Average Measures</td>
<td>.59</td>
<td>.42</td>
<td>.72</td>
</tr>
<tr>
<td>Greed-</td>
<td>Single Measure</td>
<td>.37</td>
<td>.26</td>
<td>.50</td>
</tr>
<tr>
<td>Avoidance</td>
<td>Average Measures</td>
<td>.71</td>
<td>.58</td>
<td>.80</td>
</tr>
<tr>
<td>Modesty</td>
<td>Single Measure</td>
<td>.36</td>
<td>.25</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>Average Measure</td>
<td>.70</td>
<td>.57</td>
<td>.79</td>
</tr>
</tbody>
</table>
Table 4

*Variance Accounted for by Interviewers vs Interviewees in Study One*

<table>
<thead>
<tr>
<th>Variance Accounted For</th>
<th>Interviewee</th>
<th>Interviewer</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>35.3%</td>
<td>10.2%</td>
<td>54.6%</td>
</tr>
<tr>
<td>Sincerity</td>
<td>36.1%</td>
<td>15.1%</td>
<td>48.8%</td>
</tr>
<tr>
<td>Fairness</td>
<td>21.8%</td>
<td>16.9%</td>
<td>61.3%</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>33.8%</td>
<td>9.6%</td>
<td>56.5%</td>
</tr>
<tr>
<td>Modesty</td>
<td>33.8%</td>
<td>6.7%</td>
<td>59.5%</td>
</tr>
<tr>
<td>N before removal</td>
<td>Reason for Removal</td>
<td>n cases removed</td>
<td>N cases remaining</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>2405</td>
<td>Failed audio check</td>
<td>120</td>
<td>2285</td>
</tr>
<tr>
<td>2285</td>
<td>Failed attention check</td>
<td>1194</td>
<td>1091</td>
</tr>
<tr>
<td>1091</td>
<td>Duration of survey did not go beyond the time needed to fully watch the videos</td>
<td>133</td>
<td>958</td>
</tr>
<tr>
<td>958</td>
<td>Duplicate participants</td>
<td>15</td>
<td>943</td>
</tr>
<tr>
<td>943</td>
<td>Outcome variable cannot be calculated (i.e. no variability in scores)</td>
<td>10</td>
<td>933</td>
</tr>
</tbody>
</table>
Table 6

*Study Two Cell Comparisons*

<table>
<thead>
<tr>
<th></th>
<th>No Probes</th>
<th>Probes</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>1. General, No Probes</td>
<td>2. General, Probes</td>
</tr>
<tr>
<td>Specific</td>
<td>3. Specific, No Probes</td>
<td>4. Specific, Probes</td>
</tr>
</tbody>
</table>
Table 7

Study Two HH Accuracy Correlations per Condition (Means, Standard Deviations and d)

<table>
<thead>
<tr>
<th></th>
<th>No Probes</th>
<th>Probes</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(HEXACO-200)</td>
<td>.03 (.45)</td>
<td>.31 (.48)</td>
<td>.60 [.41, .78]</td>
</tr>
<tr>
<td></td>
<td>((n = 272))</td>
<td>((n = 199))</td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.21 (.48)</td>
<td>.06 (.46)</td>
<td>-.32 [-.50, -.13]</td>
</tr>
<tr>
<td></td>
<td>((n = 249))</td>
<td>((n = 213))</td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td>-.38 [-.55, -.21]</td>
<td>.53 [.34, .73]</td>
<td></td>
</tr>
<tr>
<td>Self-Reports (RPM)</td>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.07 (.45)</td>
<td>.33 (.46)</td>
<td>.57 [.38, .76]</td>
</tr>
<tr>
<td></td>
<td>((n = 272))</td>
<td>((n = 199))</td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.19 (.50)</td>
<td>.08 (.47)</td>
<td>-.23 [-.41, -.04]</td>
</tr>
<tr>
<td></td>
<td>((n = 249))</td>
<td>((n = 213))</td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td>-.27 [-.44, -.10]</td>
<td>.52 [.32, .72]</td>
<td></td>
</tr>
<tr>
<td>SME Reports</td>
<td>General</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.11 (.48)</td>
<td>.31 (.48)</td>
<td>.42 [.23, .60]</td>
</tr>
<tr>
<td></td>
<td>((n = 272))</td>
<td>((n = 199))</td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.22 (.49)</td>
<td>.11 (.47)</td>
<td>-.22 [-.40, -.03]</td>
</tr>
<tr>
<td></td>
<td>((n = 249))</td>
<td>((n = 213))</td>
<td></td>
</tr>
<tr>
<td>(d)</td>
<td>-.22 [-.39, -.05]</td>
<td>.42 [.22, .61]</td>
<td></td>
</tr>
</tbody>
</table>
Table 8

**Study Two Overall HH and Facet Score Correlations between Self-Reports and SME Ratings**

<table>
<thead>
<tr>
<th></th>
<th>SME</th>
<th>SME 2</th>
<th>SME 3</th>
<th>SME 4</th>
<th>SME 5</th>
<th>SME 6</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>-02</td>
<td>.37</td>
<td>.66**</td>
<td>-.29</td>
<td>-.09</td>
<td>.69**</td>
<td>.45</td>
</tr>
<tr>
<td>Self</td>
<td>-.08</td>
<td>.33</td>
<td>.55*</td>
<td>-.10</td>
<td>-.12</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>Sincerity</td>
<td>[.54, [-.17, .70]</td>
<td>[.10, .82]</td>
<td>[-.56, .40]</td>
<td>[-.57, .38]</td>
<td>[-.18, .70]</td>
<td>[-.17, .70]</td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>-.17</td>
<td>-.13</td>
<td>.26</td>
<td>-.34</td>
<td>-.06</td>
<td>.34</td>
<td>.01</td>
</tr>
<tr>
<td>Fairness</td>
<td>[.60, [-.58, .38]</td>
<td>[-.26, [-.70, .17]</td>
<td>[.53, .43]</td>
<td>[-.17, .71]</td>
<td>[-.47, .49]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self</td>
<td>.15</td>
<td>.20</td>
<td>.69**</td>
<td>.02</td>
<td>.20</td>
<td>.60*</td>
<td>.51*</td>
</tr>
<tr>
<td>Self</td>
<td>.19</td>
<td>.48</td>
<td>.60**</td>
<td>.39</td>
<td>.09</td>
<td>.65**</td>
<td>.56*</td>
</tr>
<tr>
<td>Modesty</td>
<td>[.32, [-.02, .79]</td>
<td>[.17, .84]</td>
<td>[-.11, .74]</td>
<td>[-.40, .55]</td>
<td>[.24, .86]</td>
<td>[.11, .82]</td>
<td></td>
</tr>
</tbody>
</table>

* indicates p < .05, ** p < .01, N = 17
### Table 9

**Study Two Facet Accuracy Correlations per Condition (Means and Standard Deviations)**

<table>
<thead>
<tr>
<th>Facet</th>
<th>Source</th>
<th>Condition</th>
<th>Sincerity General</th>
<th>Sincerity Specific</th>
<th>Fairness General</th>
<th>Fairness Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>No Probes</td>
<td>-.13 (.48)</td>
<td>.09 (.44)</td>
<td>-.04 (.48)</td>
<td>-.01 (.44)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Probes</td>
<td>.00 (.45)</td>
<td>-.06 (.45)</td>
<td>.27 (.50)</td>
<td>-.04 (.45)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td>(n = 213)</td>
<td>(n = 199)</td>
<td>(n = 213)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>General</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SME Reports</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>General</td>
<td>.12 (.48)</td>
<td>.17 (.49)</td>
<td>.11 (.43)</td>
<td>.11 (.45)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Specific</td>
<td>.00 (.47)</td>
<td>.14 (.48)</td>
<td>-.09 (.45)</td>
<td>-.11 (.47)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td>(n = 213)</td>
<td>(n = 249)</td>
<td>(n = 213)</td>
</tr>
<tr>
<td></td>
<td>General</td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SME-Reports</td>
<td>.14 (.45)</td>
<td>.26 (.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greed-Avoidance (HEXACO-200)</td>
<td>General</td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.08 (.49)</td>
<td>.25 (.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Reports (RPM)</td>
<td>General</td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-.05 (.46)</td>
<td>.05 (.46)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modesty (HEXACO-200)</td>
<td>General</td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.12 (.46)</td>
<td>.31 (.47)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-Reports (RPM)</td>
<td>General</td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.14 (.47)</td>
<td>.03 (.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SME-Reports</td>
<td>General</td>
<td>Specific</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>----------</td>
<td>----------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$(n = 272)$</td>
<td>$(n = 199)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>$.10(.46)$</td>
<td>$.31(.47)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>$.28(.47)$</td>
<td>$.26(.44)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(n = 249)$</td>
<td>$(n = 213)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 10

*Study Two Question Type Facet Cell Comparisons*

<table>
<thead>
<tr>
<th>Cell One (General, No Probes) with Cell Three (Specific, No Probes)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>df</td>
<td>$d$</td>
<td>$p$</td>
</tr>
<tr>
<td><strong>Self-Reports</strong> (HEXACO-200)</td>
<td>Sincerity</td>
<td>-5.47</td>
<td>519</td>
<td>-.48 [-.65, -.31]</td>
</tr>
<tr>
<td></td>
<td>Fairness</td>
<td>-.88</td>
<td>519</td>
<td>-.08 [-.25, .09]</td>
</tr>
<tr>
<td></td>
<td>Greed-Avoidance</td>
<td>-5.79</td>
<td>519</td>
<td>-.51 [-.68, -.33]</td>
</tr>
<tr>
<td></td>
<td>Modesty</td>
<td>-5.55</td>
<td>519</td>
<td>-.49 [-.66, -.31]</td>
</tr>
<tr>
<td><strong>Self-Reports (RPM)</strong></td>
<td>Sincerity</td>
<td>-3.35</td>
<td>519</td>
<td>-.29 [-.47, -.12]</td>
</tr>
<tr>
<td></td>
<td>Fairness</td>
<td>-.59</td>
<td>519</td>
<td>-.05 [-.22, .12]</td>
</tr>
<tr>
<td></td>
<td>Greed-Avoidance</td>
<td>-4.42</td>
<td>519</td>
<td>-.38 [-.56, -.21]</td>
</tr>
<tr>
<td></td>
<td>Modesty</td>
<td>-4.10</td>
<td>519</td>
<td>-.35 [-.53, -.19]</td>
</tr>
<tr>
<td><strong>SME-Reports</strong></td>
<td>Sincerity</td>
<td>-1.39</td>
<td>519</td>
<td>-.12 [-.29, .05]</td>
</tr>
<tr>
<td></td>
<td>Fairness</td>
<td>-2.00</td>
<td>519</td>
<td>-.18 [-.35, -.00]</td>
</tr>
<tr>
<td></td>
<td>Greed-Avoidance</td>
<td>-5.35</td>
<td>519</td>
<td>-.47 [-.64, -.30]</td>
</tr>
<tr>
<td></td>
<td>Modesty</td>
<td>-4.47</td>
<td>519</td>
<td>-.39 [-.57, -.22]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cell Two (General, Probes) with Cell Four (Specific, Probes)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>df</td>
<td>$d$</td>
<td>$p$</td>
</tr>
<tr>
<td><strong>Self-Reports</strong> (HEXACO-200)</td>
<td>Sincerity</td>
<td>1.51</td>
<td>410</td>
<td>.15 [-.04, .34]</td>
</tr>
<tr>
<td></td>
<td>Fairness</td>
<td>6.66</td>
<td>398</td>
<td>.66 [.46, .86]</td>
</tr>
<tr>
<td></td>
<td>Greed-Avoidance</td>
<td>-.04</td>
<td>410</td>
<td>-.00 [-.20, 19]</td>
</tr>
<tr>
<td></td>
<td>Modesty</td>
<td>Sincerity</td>
<td>Fairness</td>
<td>Greed-Avoidance</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------</td>
<td>-----------</td>
<td>----------</td>
<td>----------------</td>
</tr>
<tr>
<td><strong>Self-Reports (RPM)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modesty</td>
<td>.39</td>
<td>410</td>
<td>.04 [-.15, .23]</td>
<td>.70</td>
</tr>
<tr>
<td>Sincerity</td>
<td>-5.08</td>
<td>409</td>
<td>-.50 [-.69, -.30]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fairness</td>
<td>7.32</td>
<td>399</td>
<td>.72 [.52, .92]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-.86</td>
<td>410</td>
<td>-.08 [-.28, .11]</td>
<td>.39</td>
</tr>
<tr>
<td>Modesty</td>
<td>-3.97</td>
<td>410</td>
<td>-.39 [-.59, -.20]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>SME Reports</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>3.54</td>
<td>410</td>
<td>.35 [.15, .54]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fairness</td>
<td>2.31</td>
<td>410</td>
<td>.22 [.04, .42]</td>
<td>.02</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>.29</td>
<td>410</td>
<td>.03 [-.17, .22]</td>
<td>.78</td>
</tr>
<tr>
<td>Modesty</td>
<td>1.12</td>
<td>410</td>
<td>.11 [-.08, .30]</td>
<td>.26</td>
</tr>
</tbody>
</table>
### Table 11

**Study Two Probe Effect Facet Cell Comparisons**

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>d</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Reports (HEXACO-200)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>3.10</td>
<td>469</td>
<td>.29 [.10, .47]</td>
<td>.002</td>
</tr>
<tr>
<td>Fairness</td>
<td>7.12</td>
<td>380</td>
<td>.68 [.49, .87]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>3.73</td>
<td>469</td>
<td>.35 [.16, .53]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Modesty</td>
<td>4.50</td>
<td>469</td>
<td>.42 [.23, .60]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Self-Reports (RPM)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>-3.77</td>
<td>469</td>
<td>-.35 [-.54, -.17]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fairness</td>
<td>7.88</td>
<td>382</td>
<td>.75 [.56, .94]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>2.34</td>
<td>469</td>
<td>.21 [.03, .40]</td>
<td>.02</td>
</tr>
<tr>
<td>Modesty</td>
<td>-2.40</td>
<td>469</td>
<td>-.22 [-.41, -.04]</td>
<td>.02</td>
</tr>
<tr>
<td><strong>SME Reports</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>3.72</td>
<td>469</td>
<td>.35 [.16, .53]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fairness</td>
<td>2.97</td>
<td>469</td>
<td>.27 [.09, .46]</td>
<td>.003</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>3.70</td>
<td>469</td>
<td>.34 [.16, .53]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Modesty</td>
<td>4.88</td>
<td>469</td>
<td>.46 [.27, .64]</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>df</th>
<th>d</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell Three (Specific, No Probes) with Cell Four (Specific, Probes)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>-3.69</td>
<td>460</td>
<td>-.34 [-.53, -.16]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Fairness</td>
<td>-.86</td>
<td>460</td>
<td>-.08 [-.26, .10]</td>
<td>.39</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-1.69</td>
<td>460</td>
<td>-.16 [-.34, .03]</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------</td>
<td>---</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td><strong>Modesty</strong></td>
<td>-1.17</td>
<td>460</td>
<td>-.11</td>
<td>[-.29, .07]</td>
</tr>
<tr>
<td><strong>Self-Reports (RPM)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>-1.34</td>
<td>460</td>
<td>-.13</td>
<td>[.31, .06]</td>
</tr>
<tr>
<td>Fairness</td>
<td>-.56</td>
<td>460</td>
<td>-.05</td>
<td>[-.24, .13]</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-.89</td>
<td>460</td>
<td>-.08</td>
<td>[-.27, .10]</td>
</tr>
<tr>
<td>Modesty</td>
<td>-2.05</td>
<td>460</td>
<td>-.19</td>
<td>[-.37, -.01]</td>
</tr>
<tr>
<td><strong>SME-Reports</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sincerity</td>
<td>-1.28</td>
<td>460</td>
<td>-.12</td>
<td>[-.30, .06]</td>
</tr>
<tr>
<td>Fairness</td>
<td>-1.38</td>
<td>460</td>
<td>-.13</td>
<td>[-.31, .05]</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-1.72</td>
<td>460</td>
<td>-.16</td>
<td>[-.34, .02]</td>
</tr>
<tr>
<td>Modesty</td>
<td>-.51</td>
<td>460</td>
<td>-.05</td>
<td>[-.23, .14]</td>
</tr>
</tbody>
</table>
Table 12

*Study Two Facet Final Cell Comparisons*

Cell Two (General, Probes) with Cell Three (Specific, No Probes)

<table>
<thead>
<tr>
<th></th>
<th>Sincerity</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>df</td>
<td>d</td>
<td>p</td>
<td></td>
</tr>
<tr>
<td>Self-Reports (HEXACO-200)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2.08</td>
<td>446</td>
<td>-.19 [-.38, -.01]</td>
<td>.04</td>
</tr>
<tr>
<td>Fairness</td>
<td>6.07</td>
<td>446</td>
<td>.56 [.39, .77]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-1.68</td>
<td>446</td>
<td>-.15 [-.35, .03]</td>
<td>.09</td>
</tr>
<tr>
<td>Modesty</td>
<td>-.72</td>
<td>446</td>
<td>-.07 [-.26, .12]</td>
<td>.47</td>
</tr>
<tr>
<td>Self-Reports (RPM)</td>
<td>-6.79</td>
<td>446</td>
<td>-.67 [-.84, -.45]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>7.15</td>
<td>396</td>
<td>.69 [.50, .88]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-1.81</td>
<td>446</td>
<td>-.17 [-.36, .01]</td>
<td>.07</td>
</tr>
<tr>
<td>Modesty</td>
<td>-6.12</td>
<td>403</td>
<td>-.59 [-.78, -.40]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>SME-Reports</td>
<td>2.33</td>
<td>446</td>
<td>.22 [.03, .41]</td>
<td>.02</td>
</tr>
<tr>
<td></td>
<td>.93</td>
<td>446</td>
<td>.09 [-.10, .27]</td>
<td>.35</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>-1.37</td>
<td>446</td>
<td>-.13 [-.32, .06]</td>
<td>.17</td>
</tr>
<tr>
<td>Modesty</td>
<td>.63</td>
<td>446</td>
<td>.06 [-.13, .25]</td>
<td>.53</td>
</tr>
</tbody>
</table>
Figure 1. A model of the process of accurate personality judgement
Appendix A – Study One Research Assistantship Description

Job Description

**Position:** Research Assistant

**Department:** Psychology

**Rate of pay:** $23.50/hour – for 30 hours of work in total

**Total hours of work:** 30 hours

* Hours are flexible and can be completed when preferred. E.g. hours can be completed after exams or around exam schedule. *

**Required qualifications and skills:**

- Current University of Guelph undergraduate student from any program
- Written communication skills
- Interpersonal and oral communication skills
- Problem-solving and problem sensitivity abilities
- Time management and organizational skills
- Attention to detail
- Ability to work and collaborate with others on a team
- Active listening skills
- Ability to work with minimal supervision

**Duties:**

- Assist with the preparation of project-related reports, manuscripts, and poster presentations.
- Collaborate with other research assistants and students to complete tasks.
- Code data in preparation for computer entry.
- Verify the accuracy and validity of data entered in databases, correcting for errors.
- Prepare tables, graphs, fact sheets, and written reports summarizing research results.
- Conduct internet-based and library research and summarize findings.
- Provide assistance in the design of survey instruments such as questionnaires.
- Interact and communicate with research participants to conduct studies and obtain informed consent.
- Administer standardized tests to research participants or interview them to collect research data.
- Recruit and schedule research participants.
- Track research participants and perform any necessary follow-up tasks.
## Appendix B – Study One Interview Questions and Rating Scale

<table>
<thead>
<tr>
<th>#1</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Can you tell me about a time when you had to work on a large independent project at either work or school?</td>
</tr>
</tbody>
</table>

**Potential Follow-up Probes:**
- What happened next OR what were the next steps?
- How did you feel OR what did you think about X?
- How did you react to X?
- Why did you choose this approach OR to do X?

Is there anything else that you would like to add, that either reflects STAR or that you have not been able to say yet?

<table>
<thead>
<tr>
<th>Evaluation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>#2</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

**Potential Follow-up Probes:**

- What kind of data did you use?
- What kinds of analyses did you conduct?
- Which computer programs did you use?
- What did you find?

Is there anything else that you would like to add, that either reflects STAR or that you have not been able to say yet?

### Evaluation

<table>
<thead>
<tr>
<th>1</th>
<th>Has minimal/no experience with data analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cannot name computer programs used for analyses.</td>
</tr>
<tr>
<td></td>
<td>Cannot articulate the research question.</td>
</tr>
<tr>
<td></td>
<td>Cannot describe what kinds of data are used.</td>
</tr>
<tr>
<td></td>
<td>Reveals confidential information about the data.</td>
</tr>
<tr>
<td></td>
<td>Is unable to answer what they found from the data.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Can name programs used (e.g., SPSS, R), but maybe not mention a specific test/type of analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Research question may not be clearly articulated.</td>
</tr>
<tr>
<td></td>
<td>Can explain some aspects of the data (e.g., types of participants, types of measures, source of archival data).</td>
</tr>
<tr>
<td></td>
<td>Describes 1 type of analysis or technique (e.g., t-test, content analysis, data screening).</td>
</tr>
<tr>
<td></td>
<td>Has some experience with data entry.</td>
</tr>
<tr>
<td></td>
<td>Describes a research question.</td>
</tr>
<tr>
<td></td>
<td>Describes what they found from the data.</td>
</tr>
<tr>
<td></td>
<td>Describes one limitation of research project.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>Clearly articulates a research question.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clearly explains where data came from (e.g. types of participants, types of measures, source of archival data).</td>
</tr>
<tr>
<td></td>
<td>Mentions a statistical analysis package (e.g., lavaan, psych, MBESS).</td>
</tr>
<tr>
<td></td>
<td>Describes more than 1 type of analysis or technique (e.g., t-test, content analysis, data screening).</td>
</tr>
<tr>
<td></td>
<td>Clearly describes more than one limitation of research project.</td>
</tr>
</tbody>
</table>
#3  Question

Describe a situation in which you were able to successfully convince someone (or several people) to see things your way or to do something you wanted them to do.

**Potential Follow-up Probes:**
- What was the situation?
- Who was/were the person(s) involved?
- What did you do?
- What happened next OR what were the next steps?
- How did you feel OR what did you think about X?
- How did you react to X?
- Why did you choose this approach OR to do X?

Is there anything else that you would like to add, that either reflects STAR or that you have not been able to say yet?

## Evaluation

| 1 | Has no experience persuading or convincing others.  
 |   | Presents a situation where the person simply communicated something to someone (without trying to convince him/her).  
 |   | Tried to persuade someone to do something, but the target was not receptive to the attempt and/or the persuasion backfired (e.g., tried to persuade a customer to buy a high-end model instead of an entry-level model and the customer left the store without buying anything). |

| 2 | Tried to persuade someone to do something, the target considered the ideas, but was not fully convinced.  
 |   | Was able to persuade someone to do something that he/she was not willing to do at first.  
 |   | The target did not do what was expected, although the overall outcome was still positive in the end (e.g., tried to persuade a customer to buy a high-end model instead of an entry-level model, the customer evaluated the option but went with his/her original choice).  
 |   | Used negotiation skills and/or well-crafted arguments to convince the target to change his/her opinion.  
 |   | Convinced the target that the change was in his/her own best interest. |

| 3 | Was able to persuade someone or a group of people to do something they were opposed to at first.  
 |   | Used negotiation skills and well-crafted arguments to convince a few opponents to change their opinions.  
 |   | Convinced a few opponents that the change was in their own best interest.  
 |   | Used others as ambassadors to convince the person(s) of interest.  
<p>|   | Everyone finally accepted the change and acted as expected (e.g., was able persuade all coworkers to adopt a new technique by demonstrating the value-added of the technique to a few informal leaders at first, and then by using them as ambassadors to convince the rest of the team). |</p>
<table>
<thead>
<tr>
<th>#4</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tell me about a time when a colleague or subordinate was in trouble and you took upon yourself to help him or her to deal with the situation.</td>
</tr>
</tbody>
</table>

**Potential Follow-up Probes:**
- Who was the person involved?
- What was his or her problem?
- What did you do?
- What happened next OR what were the next steps?
- How did you feel OR what did you think about X?
- How did you react to X?
- Why did you choose this approach OR to do X?

Is there anything else that you would like to add, that either reflects STAR or that you have not been able to say yet?

<table>
<thead>
<tr>
<th>Evaluation</th>
</tr>
</thead>
</table>
| 1 | • *Ignores* colleagues in trouble.  
   *Helps them only if asked by a supervisor* or if this is directly *rewarded*.  
   *The help provided was really *limited* and or was *not really what the colleague needed* (e.g., a coworker was struggling with personal issues and was so upset that he/she made mistakes and typos writing a report, the person corrected those mistakes).* |
| 2 | • Willingly *helped a colleague to solve an important issue* but did it only to *have an edge* over the colleague, or to *expect something in return* later.  
   *Willingly put his/her own tasks aside to help a colleague* solving an issue.  
   *Willingly helps a colleague* because doing so could have *important consequences for the team* or organization.  
   *Demonstrates that helping others was the logical choice for the “greater good” (e.g., a coworker was struggling to finish a project that could make the department lose a large client, the person puts his/her own project “on pause” to help put together a strong project to keep the client).* |
| 4 | • Presents a situation that illustrates that the person is a *good citizen*.  
   *Demonstrates that he/she possesses important *mentoring* skills.  
   *Willingly helps colleagues* dealing with personal or professional issues in the interest of *teamwork*.  
   *Willingly helps colleagues* to guarantee a *positive work climate*.  
   *Shared knowledge* (or helped others develop new skills), allowing colleagues to perform their tasks (e.g., a coworker was struggling to finish a project, the person puts his/her own project “on pause” to give valuable advice on how to finish the project, and makes sure that the colleague is able to implement the advice before going back to his/her task). |
Appendix C

Honesty-Humility Subject Matter Expert Training

This document contains information to help you become a subject matter expert on the trait of Honesty-Humility. Please read all the provided information to prepare you for ratings of Honesty-Humility in videotaped interviews later.

Below are descriptions of the trait Honesty-Humility, one of the six traits in the HEXACO model of personality. The five other traits are Emotionality, Extraversion, Agreeableness, Conscientiousness and Openness to Experience (although they will not be apart of this training). There are also descriptions of the four facets that make up Honesty-Humility.

These have been taken directly from the www.hexaco.org

Domain-Level Scale

Honesty-Humility: Persons with very high scores on the Honesty-Humility scale avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status. Conversely, persons with very low scores on this scale will flatter others to get what they want, are inclined to break rules for personal profit, are motivated by material gain, and feel a strong sense of self-importance.
Facet-Level Scales

**Honesty-Humility**

The *Sincerity* scale assesses a tendency to be genuine in interpersonal relations. Low scorers will flatter others or pretend to like them in order to obtain favors, whereas high scorers are unwilling to manipulate others.

The *Fairness* scale assesses a tendency to avoid fraud and corruption. Low scorers are willing to gain by cheating or stealing, whereas high scorers are unwilling to take advantage of other individuals or of society at large.

The *Greed Avoidance* scale assesses a tendency to be uninterested in possessing lavish wealth, luxury goods, and signs of high social status. Low scorers want to enjoy and to display wealth and privilege, whereas high scorers are not especially motivated by monetary or social-status considerations.

The *Modesty* scale assesses a tendency to be modest and unassuming. Low scorers consider themselves as superior and as entitled to privileges that others do not have, whereas high scorers view themselves as ordinary people without any claim to special treatment.

Below are items that correspond with Honesty-Humility in the HEXACO self-report measure.

**Sincerity**

- I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.
• If I want something from someone, I will laugh at that person's worst jokes. (Reverse-coded)

• I wouldn’t pretend to like someone just to get that person to do favors for me.

Fairness

• If I knew that I could never get caught, I would be willing to steal a million dollars. (Reverse-coded)

• I would never accept a bribe, even if it were very large.

• I’d be tempted to use counterfeit money, if I were sure I could get away with it. (Reverse-coded)

Greed-Avoidance

• Having a lot of money is not especially important to me.

• I would get a lot of pleasure from owning expensive luxury goods. (Reverse-coded)

Modesty

• I think that I am entitled to more respect than the average person is. (Reverse-coded)

• I want people to know that I am an important person of high status. (Reverse-coded)

Please read chapters 1, 2, 3, 4, 5, 9 and 10 of Kibeom Lee and Michael C. Ashton’s book, *The H Factor of Personality*. You can sign a copy out from me, or Dr. Powell.
Once you have read all the above information and feel comfortable, please complete the exercise below and email/hand it back to me. Completing this exercise actually will demonstrate knowledge in the trait of Honesty-Humility and provide concrete evidence that you are ready to complete the rating task.

If you have any questions about Honesty-Humility, or feel you need more information to be prepared, please email me at: mpike01@uoguelph.ca

To show you are comfortable with Honesty-Humility, look at the items from the 100 item HEXACO and indicate which items correspond to Honesty-Humility and specifically which facet of Honesty-Humility. On the line, indicate S (Sincerity), F (Fairness), GA (Greed Avoidance) or M (Modesty) for which facet the item corresponds with if the item fits within the umbrella of the trait Honesty-Humility.

1____ I would be quite bored by a visit to an art gallery.

2____ I clean my office or home quite frequently.

3____ I rarely hold a grudge, even against people who have badly wronged me.

4____ I feel reasonably satisfied with myself overall.

5____ I would feel afraid if I had to travel in bad weather conditions.

6____ If I want something from a person I dislike, I will act very nicely toward that person in order to get it.
7. I'm interested in learning about the history and politics of other countries.

8. When working, I often set ambitious goals for myself.

9. People sometimes tell me that I am too critical of others.

10. I rarely express my opinions in group meetings.

11. I sometimes can't help worrying about little things.

12. If I knew that I could never get caught, I would be willing to steal a million dollars.

13. I would like a job that requires following a routine rather than being creative.

14. I often check my work over repeatedly to find any mistakes.

15. People sometimes tell me that I'm too stubborn.

16. I avoid making "small talk" with people.

17. When I suffer from a painful experience, I need someone to make me feel comfortable.

18. Having a lot of money is not especially important to me.

19. I think that paying attention to radical ideas is a waste of time.

20. I make decisions based on the feeling of the moment rather than on careful thought.

21. People think of me as someone who has a quick temper.

22. I am energetic nearly all the time.

23. I feel like crying when I see other people crying.

24. I am an ordinary person who is no better than others.

26. I plan ahead and organize things, to avoid scrambling at the last minute.

27. My attitude toward people who have treated me badly is "forgive and forget".

28. I think that most people like some aspects of my personality.

29. I don’t mind doing jobs that involve dangerous work.

30. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.

31. I enjoy looking at maps of different places.

32. I often push myself very hard when trying to achieve a goal.

33. I generally accept people’s faults without complaining about them.

34. In social situations, I'm usually the one who makes the first move.

35. I worry a lot less than most people do.

36. I would be tempted to buy stolen property if I were financially tight.

37. I would enjoy creating a work of art, such as a novel, a song, or a painting.

38. When working on something, I don't pay much attention to small details.

39. I am usually quite flexible in my opinions when people disagree with me.

40. I enjoy having lots of people around to talk with.

41. I can handle difficult situations without needing emotional support from anyone else.

42. I would like to live in a very expensive, high-class neighborhood.
43. I like people who have unconventional views.

44. I make a lot of mistakes because I don't think before I act.

45. I rarely feel anger, even when people treat me quite badly.

46. On most days, I feel cheerful and optimistic.

47. When someone I know well is unhappy, I can almost feel that person's pain myself.

48. I wouldn't want people to treat me as though I were superior to them.

49. If I had the opportunity, I would like to attend a classical music concert.

50. People often joke with me about the messiness of my room or desk.

51. If someone has cheated me once, I will always feel suspicious of that person.

52. I feel that I am an unpopular person.

53. When it comes to physical danger, I am very fearful.

54. If I want something from someone, I will laugh at that person's worst jokes.

55. I would be very bored by a book about the history of science and technology.

56. Often when I set a goal, I end up quitting without having reached it.

57. I tend to be lenient in judging other people.

58. When I'm in a group of people, I'm often the one who speaks on behalf of the group.

59. I rarely, if ever, have trouble sleeping due to stress or anxiety.

60. I would never accept a bribe, even if it were very large.
People have often told me that I have a good imagination.

I always try to be accurate in my work, even at the expense of time.

When people tell me that I’m wrong, my first reaction is to argue with them.

I prefer jobs that involve active social interaction to those that involve working alone.

Whenever I feel worried about something, I want to share my concern with another person.

I would like to be seen driving around in a very expensive car.

I think of myself as a somewhat eccentric person.

I don’t allow my impulses to govern my behavior.

Most people tend to get angry more quickly than I do.

People often tell me that I should try to cheer up.

I feel strong emotions when someone close to me is going away for a long time.

I think that I am entitled to more respect than the average person is.

Sometimes I like to just watch the wind as it blows through the trees.

When working, I sometimes have difficulties due to being disorganized.

I find it hard to fully forgive someone who has done something mean to me.

I sometimes feel that I am a worthless person.

Even in an emergency I wouldn't feel like panicking.

I wouldn't pretend to like someone just to get that person to do favors for me.
I’ve never really enjoyed looking through an encyclopedia.

I do only the minimum amount of work needed to get by.

Even when people make a lot of mistakes, I rarely say anything negative.

I tend to feel quite self-conscious when speaking in front of a group of people.

I get very anxious when waiting to hear about an important decision.

I’d be tempted to use counterfeit money, if I were sure I could get away with it.

I don’t think of myself as the artistic or creative type.

People often call me a perfectionist.

I find it hard to compromise with people when I really think I’m right.

The first thing that I always do in a new place is to make friends.

I rarely discuss my problems with other people.

I would get a lot of pleasure from owning expensive luxury goods.

I find it boring to discuss philosophy.

I prefer to do whatever comes to mind, rather than stick to a plan.

I find it hard to keep my temper when people insult me.

Most people are more upbeat and dynamic than I generally am.

I remain unemotional even in situations where most people get very sentimental.

I want people to know that I am an important person of high status.
97___ I have sympathy for people who are less fortunate than I am.

98___ I try to give generously to those in need.

99___ It wouldn’t bother me to harm someone I didn’t like.

100___ People see me as a hard-hearted person.
Appendix D

Honesty-Humility (HH) SME Rating Sheet

Please follow these instructions for rating the interviewees.

You will rate each interviewee on a rating scale. You will rate their overall HH level, as well as their level of each HH facet (Sincerity, Fairness, Greed Avoidance and Modesty). In addition to this you will specify any cues you noticed that led you to select your specific rating for each facet. If you have any cues that you feel are important but do not fit within a specific facet, please look at the training document again. If they still do not seem to fit, but you still see them are important after reviewing the document, please list these cues but not the corresponding facet. The participant number of each interviewee must also be recorded so it can be correlated with the self-report of that interviewee. The participant number is indicated in the video title. All of the ratings will be recorded in a google sheets document that I have created and linked you to.

The rating scales below are visual aids to represent how you will be ranking the interviewees. As specified, 0 indicates someone who is very low on the trait, 50 indicates that they would be average and 100 indicates they appear to be very high on the trait. When you are recording your rating for HH overall and for each facet in the google sheet, imagine you are putting a X on the scale below. Recording in excel sheet simply makes it easier for my data analysis later.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>X</th>
<th></th>
<th></th>
</tr>
</thead>
</table>
0      |       |       |       |       |       |   65|       |       |
50     |       |       |       |       |       |     |       |       |
100    |       |       |       |       |       |     |       |       |

(Very Low) (Average) (Very High)
Below is a visual of what is included in the google sheet for extra clarity.

**Your InitialsParticipant Number:**

**Honesty-Humility Overall Rating**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Very Low)  (Average)  (Very High)

**Sincerity**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Very Low)  (Average)  (Very High)

**Cues:**
**Fairness**

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Very Low) (Average) (Very High)

**Cues:**

•

**Greed-Avoidance**

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>50</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Very Low) (Average) (Very High)

**Cues:**

•
### Modesty

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Very Low)</td>
<td>(Average)</td>
<td>(Very High)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Cues:
Appendix E – M-turk HIT Description

We are conducting an academic survey about personality judgments made in an interview. You will be required to watch five videos and rate the interviewees' personality in these videos. **You will receive $4.00 for this study however, you must pass all of the attention checks to be compensated. There will be five attention checks, each following your ratings of the video.** Select the link below to complete the survey. At the end of the survey, you will receive a code to paste into the box below to receive credit for taking our survey.

**If you have previously completed this survey** (i.e., another HIT/batch titled 'Interviewer Judgements and Video Ratings', by the same requester "Melissa Pike"), **please do not complete it again. You will not be compensated twice.**

**Make sure to leave this window open as you complete the survey.** When you are finished, you will return to this page to paste the code into the box.
Appendix F – Study Two Participants Consent Form

Consent to Participate

You are asked to participate in a research study conducted by Dr. Deborah Powell and Melissa Pike from the Psychology Department at the University of Guelph. Results will contribute to the completion of a psychology thesis project at the University of Guelph. If you have any questions or concerns about the research, please feel free to contact Deborah Powell by email @ dpowell@uoguelph.ca or daytime phone @ (519) 824-4120 ext. 52167.

This research project is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

1. Invitation to Participate
This study is examining the detection of personality in interviews. You are invited to participate in this study because you are a worker on Amazon Mechanical Turk. The purpose of this letter is to give you information, so you can make an informed decision regarding participation in this research. Please print a copy for your records.

2. Purpose of this Study
The purpose of this study is to examine individual’s ability to detect personality in a job interview. This work can help inform organizations of the possibility of interviews as a selection tool based on one’s personality.

3. Inclusion Criteria
Amazon MTurk workers from the United States

4. Study Procedures
If you agree to participate, you will be asked watch videotaped interviews and rate the candidates in those videos on personality traits. This study will take place entirely online and will take approximately 40 minutes to complete.

5. Risks and Benefits
The risks of this study are not any greater than you would encounter in day to day life. You will not directly benefit from participating in this study, but information gathered may benefit organizations in the future.

6. Compensation
You will be compensated with 4.00 for the full completion of this study. You will be compensated 0.80 for each video that you rate, therefore you will only gain the full $4.00 after rating all five videos. Attention checks will occur at the end of each video, so if you do not pass the attention checks you will not continue on to the next video and you will be paid for your participation up to that point.

7. Voluntary Participation
Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no ill effects. To withdraw from the study, you can
simply stop answering the survey questions. Once you have completed the survey and submitted your results you will no longer be able to withdraw your data. Withdrawing will not affect compensation, you will get paid for what you have completed so far. **You must click to the end of the survey to receive compensation. If you simply click out of the browser, you will not be paid.** You do not waive any legal rights by participating.

8. **Confidentiality**
Your identity will be kept confidential. The only personal information that will be collected will be demographic information such as gender, race and age. No identifying information otherwise will be collected with this survey. The anonymized data set will be made available for other researchers to use in accordance with the principles of open science. If the data from this study were to be published, overall trends will be reported without identifiable details. Please note that confidentiality cannot be guaranteed while data are in transit over the internet.

10. **Contacts for Further Information and Publication**
If you require any further information regarding this research project or your participation in the study you may contact the Principal Investigator: Dr. Deborah Powell (519) 824-4120 ext. 52167, email: dpowell@uoguelph.ca, or Melissa Pike, email: mpike01@uoguelph.ca. If you would like to receive a copy of any potential study results, please contact Melissa Pike.

If you have questions regarding your rights and welfare as a research participant in this study (REB#18-08-24), please contact: Director, Research Ethics; University of Guelph; reb@uoguelph.ca; (519) 824-4120 (ext. 56606)

This project has been reviewed by the Research Ethics Board for compliance with federal guidelines for research involving human participant
Appendix G - Honesty-Humility Descriptions and Cues for Study Two

The **Sincerity** scale assesses a tendency to be genuine in interpersonal relations. Low scorers will flatter others or pretend to like them in order to obtain favors, whereas high scorers are unwilling to manipulate others.

Example of cues candidates could emit of Sincerity are:

- Not using manipulative strategies to convince others to do things (High Sincerity) vs Uses manipulative strategies to convince others to do things (Low Sincerity)
- Doing additional work to benefit future employees (High Sincerity) vs Only doing additional work if there is some benefit to themselves (Low Sincerity)

The **Fairness** scale assesses a tendency to avoid fraud and corruption. Low scorers are willing to gain by cheating or stealing, whereas high scorers are unwilling to take advantage of other individuals or of society at large.

Example of cues candidates could emit of Fairness are:

- Follows the rules (High Fairness) vs does not follow the rules (Low Fairness)
- Tries to benefit everyone in situations (High Fairness) vs Tries to benefit themselves in situations (Low Fairness)

The **Greed Avoidance** scale assesses a tendency to be uninterested in possessing lavish wealth, luxury goods, and signs of high social status. Low scorers want to enjoy and to display wealth
and privilege, whereas high scorers are not especially motivated by monetary or social-status considerations.

Example of cues candidates could emit of Greed Avoidance are:

- Focuses on personal growth and the process when completing tasks (High Greed Avoidance) vs Focuses on monetary gain (Low Greed Avoidance)
- Focuses on experiences and self-fulfillment as rewards (High Greed Avoidance) vs Focuses on money as a reward (Low Avoidance)

The Modesty scale assesses a tendency to be modest and unassuming. Low scorers consider themselves as superior and as entitled to privileges that others do not have, whereas high scorers view themselves as ordinary people without any claim to special treatment.

Example of cues candidates could emit of Modesty are:

- Willing to listen to others who disagree with their opinion (High Modesty) vs Thinks that they are always right (Low Modesty)
- Can admit to their mistakes (High Modesty) vs Not able to state that they may be wrong (Low Modesty).

**Composite Trait:** Persons with very high scores on the Trait X scale avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status. Conversely, persons with very
low scores on this scale will flatter others to get what they want, are inclined to break rules for personal profit, are motivated by material gain, and feel a strong sense of self-importance.
Appendix H – Job Description for Study Two

Account Manager for Walmart

POSITION: Onsite Account Manager
COMPANY: Walmart Canada
DEPARTMENT: Sales
LOCATION: Guelph, ON

Job Description:
Walmart Canada is currently seeking an Account Manager to join our team in the newly finished Walmart Supercentre in Guelph, Ontario. Reporting to the CEO, the successful candidate’s primary responsibility will be to establish and maintain relationships between Walmart and partner companies, ensuring sales and service targets are met.

Qualifications:
- High School Diploma / GED
- A completed Undergraduate Degree
- No previous account management experience required

Skills:
- Effective problem solving and critical thinking skills
- Ability to set goals, multi-task, meet/exceed objectives
- Strong written/verbal communication and social skills
- Excellent planning and organizational skills
- Proficient in Microsoft Office

Essential Duties and Tasks of the Position:
- Manage accounts, including contract negotiation and agreements
- Build and maintain strong customer relationships
- Manage customer requests, address complaints and concerns
- Preparation of sales reports and tracking account statuses
- Conduct monthly/quarterly customer review meetings
- Conduct research within the industry to identify new market opportunities and make recommendations to supervisors
- Work through challenges to improve the customer experience
Appendix I – Interview Questions for Study Two Interview Videos

1. Tell me about a time when you had a difference of opinion or conflict with a supervisor/co-worker

    **Probes**
    a) How did you feel during this situation?
    b) What were important things for you to consider in this situation?
    c) Why did you choose your approach?

2. Tell me about a time when you had a difference of opinion or conflict with a supervisor/co-worker and tried to persuade them to see things your way.

    **Probes**
    a) How did you feel during this situation?
    b) What were important things for you to consider in this situation?
    c) Why did you choose your approach?

3. Tell me about a time when you saw the need and opportunity to set up a long-term relationship with other people.

    **Probes**
    a) How did you feel during this situation?
    b) What were important things for you to consider in this situation?
    c) Why did you choose your approach?

4. Tell me about a time when you saw the need and opportunity to set up a long-term relationship with other people because it would produce a personal benefit.

    **Probes**
    a) How did you feel during this situation?
    b) What were important things for you to consider in this situation?
    c) Why did you choose your approach?
Appendix J – Study Two Interview Candidate Consent Forms

Consent to Participate

You are asked to participate in a research study conducted by Dr. Deborah Powell and Melissa Pike from the Psychology Department at the University of Guelph. Results will contribute to the completion of a psychology thesis project at the University of Guelph. If you have any questions or concerns about the research, please feel free to contact Deborah Powell by email @ dpowell@uoguelph.ca or daytime phone @ (519) 824-4120 ext. 52167.

This research is funded by the Social Sciences and Humanities Research Council of Canada (SSHRC).

1. Invitation to Participate

This study is examining the detection of personality in interviews. This study will be used to create video tools for participants to view at a later date. You are invited to participate in this study. The purpose of this letter is to give you information, so you can make an informed decision regarding participation in this research. Please keep a copy for your records.

2. Purpose of this Study

The purpose of this study is to examine individual’s ability to detect personality in a job interview. This work can help inform organizations of the possibility of interviews as a selection tool based on one’s personality.

3. Inclusion Criteria

Any University of Guelph student.
4. **Study Procedures**

If you agree to participate, you will be filmed answering a series of interview questions. Following the interview, you will complete a short online survey. The filming session should take approximately 50 minutes and the survey should not take longer than 10 minutes.

* **If you choose to consent to participate, you will be agreeing for your video to be used in the second phase of this study.** The second phase involves the video filmed today being shown to trained research assistants and participants from an online participant pool in the United States. The videos will be a part of an online survey for the participants in the United States. In addition to this form, a release form will be given to you to sign, indicating further that you agree for your video to be used. The survey data collected will not be shared with the participants in the second phase.

5. **Risks and Benefits**

The risks of this study are not any greater than you would encounter in a typical job interview. You will not directly benefit from participating in this study, but information gathered may benefit organizations in the future.

6. **Compensation**

You will be compensated $20 for participation in this study.

7. **Voluntary Participation**

Participation in this study is voluntary. You may refuse to participate, refuse to answer any questions or withdraw from the study at any time with no ill effects. To withdraw from the study, you can inform the interviewer that you would like to stop filming or simply stop answering the survey questions. If you do decide to withdraw, but you have already submitted survey responses, please email Melissa Pike (mpike01@uoguelph.ca) so she can remove your data. You can withdraw your
data up until two weeks after today’s date. Withdrawing will not affect compensation. You do not waive any legal rights by participating.

8. Confidentiality

Your videos will be shown to research assistants and participants who use the Amazon Mechanical Turk forum. However, no identifying details about you, other than your answers and your face will be shared with these individuals. Separate from this, you will be assigned a participant number that will be used to match your video with the survey responses. This data will contain no identifiable information. The anonymized data from the survey will be made available for other researchers to use in accordance with the principles of open science. The videos however will not be made available to others through open science. If the data, from the survey, were to be published, overall trends will be reported without identifiable details. Please note that confidentiality cannot be guaranteed while data are in transit over the internet.

9. Contacts for Further Information and Publication

If you require any further information regarding this research project or your participation in the study you may contact the Principal Investigator: Dr. Deborah Powell (519) 824-4120 ext. 52167, email: dpowell@uoguelph.ca, or Melissa Pike, email: mpike01@uoguelph.ca. If you would like to receive a copy of any potential study results, please contact Melissa Pike.

If you have questions regarding your rights and welfare as a research participant in this study (REB#18-08-24), please contact: Director, Research Ethics; University of Guelph; reb@uoguelph.ca; (519) 824-4120 (ext. 56606)
10. Contacts for Co-op and Career Services

If you experience any discomfort with your interview performance during this mock interview and want more information about successful interview performance, you may contact Co-op Education and Career Services, (519) 824-4120 ext. 52323, email: recruit@uoguelph.ca.

This project has been reviewed by the Research Ethics Board for compliance with federal guidelines for research involving human participant
Appendix K – Study Two Interview Candidates Media Release Form

Photograph, video and/or voice recording release form

Project Title: The Detection of Personality Traits in Employment Interviews

Name of Principal Investigator: Dr. Deborah Powell

Name of Participant: __________________________________________

As part of the Project, the Project researchers, which includes the Principal Investigator, will:

☐ make video recording(s) of you while you participate in the Project

With this form, you are being asked for your permission to share these video recording(s) (the “Records”) with the Research Assistants and with Amazon Mechanical Turk participants for this research project. Please indicate below (by initialing) what uses of the Records you consent to. You are under no obligation - what uses you consent to is completely up to you. The Records will be used only in the ways you consent to. Your name will not be included in any use of the Records.

1. The Records can be used, displayed, and distributed on the internet through a survey to the participants in phase two.

   Audio recording(s) _____ Video recording(s) _____

I have read this form and given my consent to the use of the Records as indicated above. I acknowledge and agree that: (a) the University of Guelph owns the Records including all rights in the Records; (b) while, in any use of the Records, your name will not be included there is no assurance or guarantee you will not be identifiable; and, (c) I will not have the opportunity to inspect or approve any finished or unfinished
material in any media in which the Records appear unless specific arrangements are negotiated with
the Principal Investigator prior to disclosure.

Participant First and Last Name (Print) ____________________________________________________
____________________________________________________

____________________________________________________
Signature of Participant Date
Interview Judgements and Appraisals

Invitation to Participate
The University of Guelph Department of Psychology is currently seeking participants for a research study examining the detection of personality in interviews. This portion of the study will allow participants to be involved in a videotaped mock interview and completion of a short questionnaire.

Who is eligible to participate?
- Any University of Guelph student

What will be asked of me?
You will participate in a mock job interview and complete a short questionnaire. The interview will be videotaped. The interviews generated from this study will be used in a follow-up studying regarding the detection of personality in interviews. We ask that you come dressed
in appropriate interview attire. Answers to the questionnaire and any identifying information, outside the videotaped interview, will remain completely confidential. Questions will be general interview questions and not of a sensitive nature.

As compensation for participating in this study participants will receive $20.

If you have any questions about this study or would like to participate please contact:

Melissa Pike, Masters Student, Department of Psychology, University of Guelph
Email: mpike01@uoguelph.ca
Appendix M – Material Creation Methods Section

**Participants.** Participants were recruited through flyers placed around the University of Guelph Campus as well as ads in University of Guelph student Facebook groups, on Recruitment Guelph and on Psychology and Business School Courselink homepages. Participants were compensated $20 for their participation. Seventeen participants were recruited.

Of the participants, 71% identified as female and 29% as male. Ages ranged from 18 to 25. The participants indicated they identified as the following: eleven as Caucasian (65%), one as Indigenous and Black/African/Caribbean (6%), one as Arab (6%), one as South Asian (6%), two as Southeast Asian (12%), and one as West Asian (6%).

**SMEs.** Subject matter experts consisted of three SMEs from Study One (including myself), plus an additional three research assistants recruited by Dr. Powell. All SMEs were female. SMEs that were not included in Study One were provided training on the trait HH (as specified in Study One). SMEs provided ratings of HH to use one form of estimated true score to compare participants ratings as well as watched the videos to provide cues of HH.

**Interviewer.** Each candidate was interviewed by myself using a standardized interview guide.

**Procedure.** Participants came into Dr. Powell’s lab and participated in a mock job interview. Participants answered questions as if they were applying to an Accounts Manager job. This is a job that students would have the qualifications for and therefore could envision themselves applying. Participants answered a total of four interview questions. They answered two general personality-tailored behavioural questions and two behavioural questions tapping specifically into HH. The general questions were asked to the candidates and were then followed
by the specific questions. This process enabled the candidates to go from a broad question to a more specific question, to increase the ease of answering the question. The candidates were informed that because the questions were similar but will be split into separate videos (for Study Two), that they should treat each question as if the interviewer did not hear the answer to the question before it. This process allowed candidates to use the same answer to the general and specific questions if they saw fit, but also ensured that the full story was provided when answering both questions. Following each question, participants were probed to provide additional information to each interview question. Standard probes were used and asked to every candidate. These probes were used to obtain the most complete answer to the question as possible.

Following the interview, the participants provided self-report ratings of HH. They also filled out an RPM scale to rate their own HH levels. The participants were provided definitions of each of the facets and HH to be able to rate themselves. HH was labelled as ‘Composite Trait’ so participants did not focus on the concept of honesty when rating their overall HH level. After they completed their ratings they received compensation.

**Cues.** Analogous to Study One, along with rating the HH level of the candidates in the videos, SMEs provided cues for each facet of HH. These cues will be used to gain an understanding of HH cues that are elicited in an interview, as Study One did not produce any SME accuracy correlations above .5.

**Measures.**

**Personality.** Personality of candidates in the videos was measured using the Honesty-Humility scale from the 200-item HEXACO (Lee & Ashton, 2004). Internal consistency reliability for this scale is .90 (Lee & Ashton, 2004). The reliabilities are .75, .81, .85, and .78 for
the facets of Sincerity, Fairness, Greed-Avoidance and Modesty respectively. The items are presented on a 5-point Likert scale (1 = *strongly disagree*, to 5 = *strongly agree*).

**Relative Percentile Method (RPM) Scale.** The scale used by participants to rate their own level of HH was an RPM scale, similar to the one used in Study One. This scale ranges from 0-100 (0 = very low, 50 = average, 100 = very high). Participants rated themselves by deciding whether they thought they were very low, average, or very high on HH and each of its facets. SMEs also used the RPM scale to rate the candidates in the videos. The RPM scale type has been judged as advantageous in past research (Goffin, Jelley, Powell, & Johnston, 2009).

**Selection of Videos.** Five videos were selected from the 17 videos filmed for use in Study Two. These videos were selected based on the following criteria: a large range in HH scores, videos that were easiest to rate, as well as a mix of males and females. Videos were first considered based on the range of self-reported HH scores. In the general population scores on HH would range from low to high, and therefore the videos were selected to try to represent that range. SMEs also watched these videos and provided ratings of the interviewees on an RPM scale. These ratings were then used to determine which videos were easiest to rate (by determining for which videos SMEs provided observer reports that approximated the candidates’ self-reports). In addition, the rank order created with interviewees’ self-report scores, was compared to the rank order created with SME observer reports. Videos that were easier to rank order should be easier to rate and therefore videos were selected based on this criterion as well. It should be noted that although six SMEs rated the candidates, only five of their scores were considered when selecting videos. One SME was eliminated because, after running a reliability analysis on the SMEs, it was determined that reliability of the SMEs was highest with one
specific SME eliminated. Finally, gender was considered when selecting the videos, to ensure that there was a mix of both genders. Based on these four criteria, five videos were selected.

These videos consisted of two males and three females, with four identifying as Caucasian and one identifying as Arab. The interviews had a range of HH scores from 2.00 to 3.91 (on a scale from 1 to 5). Both males had self-report scores that were lower than the three females. Although this difference may be viewed as a potential issue, to have the range in HH scores, both females and males needed to be included in the selected videos. Research demonstrates that in the normal population males tend to have lower HH scores than females (Lee & Ashton, 2004), indicating that these videos are more representative of the population with males scoring on the lower end of the HH scale.

Once the videos were chosen, they were split into conditions based on question type and whether probes were asked to the candidate. To do this, each interview video was first edited/split into two videos, based on question type. Then a copy of each question type video was edited to create a copy without probes. For the no-probe condition it is assumed that if a candidate were to be asked a question and not probed, the original answer they gave would remain the same, and therefore, removing any answers they had given following the probes would not affect what their original answer would have been. To summarize, based on the editing, four videos were created for each interviewee, one for each condition.
Appendix N - Study Two HH Differences between Accuracy Correlations

<table>
<thead>
<tr>
<th></th>
<th>No Probes</th>
<th>Probes</th>
<th>z</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Self-Reports (HEXACO-200)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>.03 (.45)</td>
<td>.31 (.48)</td>
<td>3.10*</td>
<td></td>
</tr>
<tr>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.21 (.48)</td>
<td>.06 (.46)</td>
<td>-1.60</td>
<td></td>
</tr>
<tr>
<td>(n = 249)</td>
<td>(n = 213)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>-2.07</td>
<td>2.49*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Self-Reports (RPM)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>.07 (.45)</td>
<td>.33 (.46)</td>
<td>2.91*</td>
<td></td>
</tr>
<tr>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.19 (.50)</td>
<td>.08 (.47)</td>
<td>-1.17</td>
<td></td>
</tr>
<tr>
<td>(n = 249)</td>
<td>(n = 213)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>-2.07</td>
<td>2.62*</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SME Reports</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General</td>
<td>.11 (.48)</td>
<td>.31 (.48)</td>
<td>2.25*</td>
<td></td>
</tr>
<tr>
<td>(n = 272)</td>
<td>(n = 199)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific</td>
<td>.22 (.49)</td>
<td>.11 (.47)</td>
<td>-1.20</td>
<td></td>
</tr>
<tr>
<td>(n = 249)</td>
<td>(n = 213)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>-1.29</td>
<td>0.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05
Appendix O – Study Two Cues for Use for Detection of HH in the Interview (By Facet)

<table>
<thead>
<tr>
<th>Facet</th>
<th>Cues</th>
</tr>
</thead>
</table>
| Sincerity | • Trying to put themselves in the other person’s shoes  
  • Viewing that it is important to understand how the other person felt  
  • Seeing the importance of making friends, but also not having it affect your work  
  • Getting a research job that has influence in different parts of the university because of its impact on being re-hired and used as a reference  
  • Frustrated when supervisors do not stick up for them  
  • Making it clear they are a people person by stating that friendships make it easier to work with and that having these friends makes the job easier to enjoy  
  • Doesn’t just think about how their behaviour benefits you but also how it benefits others  
  • Stating that relationships are important in the workplace to have someone to relate to and talk to and keep you happy and motivated to work  
  • Stating that it is good to have someone in corner in case anything was to go wrong  
  • Viewing long-term friendships as a way to promote gain in your job  
  • Engaging in behaviours because it is important for the kids  
  • Stating that what they were was asking for was reasonable  
  • Stating that they don’t want to be to forward when asking for a donation  
  • Taking advantage of the fact that someone had kids to convince them  
  • Feeling bad for leaving a job because they don’t want to leave the supervisor that got them the job  
  • Honest about feeling nervous dealing with patrons because they can complain to you or your supervisor  
  • Honest about not coming across the situation the question is asking about.  
  • Doing coop and interested in vet program. Discussed excitement about vet school and being taken under vet’s wing.  
  • Discussing how they enjoyed talking to co-workers. Being nice and interested because the co-worker was engaging  
  • Discussing co-worker as a nice and dependable guy and that they loved to help him out and take as many shifts as needed  
  • Persuaded supervisor that they had already worked 8 hours and it was their decision on whether they would stay or go  
  • Focusing on their own best interest when making decisions  
  • Making sure they are on good terms with superior, so they will be hired again or get a good reference  
  • Establishing and feeling mutual trust between them and the management team.  
  • Caring deeply about relationship with friends because it is better for life in general to have someone who understands you and that you can count on. |
• Wanting to help others because they had a co-worker who took the time to support them.
• Focusing on having a good relationship with co-workers and helping them out.
• Making sure to have good relationships with co-workers. It provided a benefit when they needed help with customers.
• Wanting to be genuine in interactions with co-workers.
• Want to make sure not to hurt other’s feelings and make sure they are heard.
• Used the benefit for students to persuade others to keep their job.
• Seeing the importance in establishing trust between them and parents so they understand that they are navigating their kids in a good direction.
• Seeing the importance of honestly in relationships.
• Viewing promotion as important and establishing a good relationship with one’s manager to do that. Also important for references.
• Sees the importance of being considerate of other’s opinions.
• Thinks critically about how to persuade others to agree with them.
• Viewing relationship with friends as meaningful and understanding others as important. Good to have relationships that are mutually beneficial.
• When thinking about establishing relationships with co-workers, first thinking about if the relationship will benefit them and the other.
• Keeping relationships for professional value.
• Becoming close with professors because they can help you in the future.
• Trying to persuade others by showing a benefit you can give them for helping you out.
• Establishing a relationship because of the connections the other person had to help you find a job.
• Relieved to have someone who shares their experiences and that they can learn from.
• Focus on investing in others lives and building positive relationships.
• Builds relationships with customers to benefit the company.
• Focusing on honesty when establishing relationships.
• Sees the importance in how others feel and being respectful.
• Interested in fostering good work environment
• Sees the need for relationship to be more than just a personal benefit.

| Greed-Avoidance | • Enjoys instant gratification
• Leaving job because they will get more from the other job
• Focused on themselves as opposed to others.
• Focused on others being able to help them.
• Taking work shifts not just for money.
• Focussed more on themselves instead of their work
• Wanting to give back to others
• Emphasized the importance of getting a promotion
• Focussed primarily on relationships with others |
<table>
<thead>
<tr>
<th>Modesty</th>
<th>Important of being respectful.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Acknowledging that the work sounds mundane</td>
</tr>
<tr>
<td></td>
<td>Acknowledging others points of view</td>
</tr>
<tr>
<td></td>
<td>Acknowledging when they don’t have a lot of experience</td>
</tr>
<tr>
<td></td>
<td>Focussing on wanting respect</td>
</tr>
<tr>
<td></td>
<td>Taking responsibility for things they did wrong.</td>
</tr>
<tr>
<td></td>
<td>Doesn’t want to talk down to others</td>
</tr>
<tr>
<td></td>
<td>Being polite and respectful to others</td>
</tr>
<tr>
<td></td>
<td>Empowered and confident enough to tell co-worker what they were doing was the right way.</td>
</tr>
<tr>
<td></td>
<td>Considering how they might be wrong</td>
</tr>
<tr>
<td></td>
<td>Honest about doing what was most convenient as to what was necessarily right</td>
</tr>
<tr>
<td></td>
<td>Acknowledges that others may have better opinions and they might not always be right</td>
</tr>
<tr>
<td></td>
<td>Said that they thought their opinion is best</td>
</tr>
<tr>
<td></td>
<td>Reflecting back on choices and saying they weren’t the best</td>
</tr>
<tr>
<td></td>
<td>Discussed how they felt so overwhelmed first on the job</td>
</tr>
<tr>
<td></td>
<td>Admitted to being anxious when dealing with their supervisor. Wants to ensure they do not come off as condescending.</td>
</tr>
<tr>
<td></td>
<td>Empathized that no one’s opinion is better than others and that people should reach a compromise</td>
</tr>
<tr>
<td></td>
<td>Admitted to feeling hesitant.</td>
</tr>
<tr>
<td></td>
<td>Empathized being accommodating</td>
</tr>
<tr>
<td></td>
<td>Discussing that her manager knows what’s best for the company and knows better than them.</td>
</tr>
<tr>
<td></td>
<td>Upset when they got things wrong because they felt that they were right.</td>
</tr>
<tr>
<td></td>
<td>Empathized being humble.</td>
</tr>
<tr>
<td></td>
<td>Discussed that it is important to make sure everyone is taken care of, even if that means asking for help</td>
</tr>
<tr>
<td></td>
<td>Empathized treating others how they want to be treated.</td>
</tr>
<tr>
<td></td>
<td>Working for money, but also for the job of helping others.</td>
</tr>
<tr>
<td></td>
<td>Making money important, especially for student</td>
</tr>
<tr>
<td></td>
<td>Focused on harmonious outcomes for all</td>
</tr>
<tr>
<td></td>
<td>Passionate about the company and want them to do well.</td>
</tr>
<tr>
<td></td>
<td>Largely focused on working with children</td>
</tr>
<tr>
<td></td>
<td>Doesn’t want friendship to hurt their work and focussed on listening to manager.</td>
</tr>
<tr>
<td></td>
<td>Building relationships to get a job</td>
</tr>
<tr>
<td></td>
<td>Focussed on comfort in relationships</td>
</tr>
<tr>
<td></td>
<td>Work comes before personal issues.</td>
</tr>
<tr>
<td></td>
<td>Happy to come up with a solution that helps client.</td>
</tr>
<tr>
<td></td>
<td>Focussed on friendships just to make things easier</td>
</tr>
</tbody>
</table>
| • Importantly mindful of what others are doing and understanding their role in the context of greater organization.  
• Annoyed when others have power over them but cannot doing anything because of this power  
• Honest about mistakes on the job  
• Talked about jobs being about learning  
• Admitted using someone else's approach because it was better  
• Talked about the need to improve their interpersonal skills and about needing mentors' help  
• Stated that their approach is best  
• Stated that you can never assume that your way is the right way |
Appendix P - Proposal

In employee selection, organizations can select employees based on a variety of criteria. One of these criteria, and an important predictor of job performance, is personality (Barrick & Mount, 1991). Research has found that traits such as Extraversion are important for specific jobs, whereas traits like Conscientiousness are important predictors of performance in all jobs (Barrick & Mount, 1991). One trait that has been overlooked, but is also particularly important to consider in selection, is the trait of Honesty-Humility (HH). HH is a trait within the six-factor HEXACO model of personality and it defines an individual’s level of sincerity, fairness, greed-avoidance and modesty (Lee & Ashton, 2004). HH is important in selection for two reasons. First, HH can be a predictor of performance in specific jobs. Particularly, high levels of HH can predict success in jobs that involve caregiving (Johnson, Rowatt, & Petrini, 2011). Second, individuals with low levels of HH engage in behaviours such as theft, absenteeism, workplace delinquency (Lee, Ashton, & de Vries, 2005) and unethical business decisions (Lee, Ashton, Morrison, Cordery, & Dunlop, 2008). Therefore, organizations may want to be able to detect a candidate’s level of HH for a specific job, but more often it is useful for organizations to be able to detect HH to decrease the negative behaviours associated with low levels of this trait. To be able to accomplish either of these goals, organizations need to be able to accurately assess HH.

Common ways to assess constructs such as HH, and those similar to HH, are self-report personality measures and integrity tests. It should be noted that while integrity is not synonymous with HH there is a correlation of .48-.63 between integrity tests and HH, depending on the type of test used (Marcus, Lee & Ashton, 2007). Therefore, integrity tests do tap into some aspects similar to HH. Personality and integrity tests have previously been viewed as effective and efficient ways to assess HH and integrity, respectively. However, Morgeson et
al. (2007) noted a number of problems with these measures when they are used within the selection context. These problems include negative applicant reactions (Rosse, Miller, & Stecher, 1994) and fake-ability, which involves candidates’ ability to present themselves in an ideal light on a measure (Berry, Sackett, & Wiemann, 2007; Topping, & O’Gorman, 1997). Because of these issues, Morgeson et al. (2007) suggested that alternatives to self-reports should be investigated for selection purposes.

A possible alternative to self-report measures for detecting the HH level of a candidate is to assess HH during the employment interview. Interviews could be a more ideal method for detecting personality because they are widely used and accepted in candidate selection, and applicants tend to view interviews more positively than personality tests (Steiner & Gilliland, 1996). Also, research has found that faking in general occurs to a lesser extent in interviews than on personality tests (Van Iddekinge, Raymark, & Roth, 2005). Less response distortion in interviews may be a result of interviews separating the descriptions of behaviours from the evaluation of those behaviours (Levashina, Hartwell, Morgeson, & Campion, 2014). This separation likely occurs because interviewers, as opposed to the individuals themselves, assign the ratings. Therefore, applicants have less control over their scores when an interviewer is responsible for assigning them. Meta-analyses have also been conducted that show that observer ratings of personality are more predictive of job performance than self-reports (Connelly & Ones, 2010; Oh, Wang, & Mount, 2011). Therefore, in the context of selection, observer ratings may have more merit as long as those ratings are made accurately. Interviews may be a more expensive and time-intensive method; however, given their overall benefits over self-reports, interviews could be a viable alternative for assessing personality in selection.
For interviews to be a useful tool in terms of personality detection in selection, they must also be effective enough to merit the extra time and potential cost. Fortunately, research conducted by Powell and Goffin (2009), as well as Powell and Bourdage (2016), has shown promise for the detection of personality in interviews. The problem, however, is that researchers have yet to examine the detection of HH in interviews. Therefore, this research will be examining the overarching question of: can HH be detected in the interview?

This proposal will begin by explaining the models that are crucial for HH and its detection, and then explain how the overarching research question will be addressed. This research will look at the cues of HH that are elicited in the interview as well as the questions interviewers can ask to elicit HH cues.

**The HEXACO Model of Personality**

Personality is often thought of in terms of five traits within the Big Five model of personality (Goldberg, 1990; Goldberg, 1992). However, research has recently discovered that personality is more robustly measured by six traits (Lee and Ashton, 2008). These six traits are encompassed in the HEXACO model of personality and are categorized as the following: Honesty-Humility, Emotionality, Extraversion, Agreeableness, Conscientiousness, and Openness to Experience. The latter five traits are encompassed within the Big Five, whereas Honesty-Humility is unique to the HEXACO. Each trait is described in more detail in Table 1. Honesty-Humility is broken into the facets of modesty, sincerity, fairness and greed-avoidance. Therefore, Honesty-Humility captured elements of personality which had yet to be properly encompassed within the Big Five. These new elements within the HEXACO have also enabled this personality model to gain higher predictive validity than the Big Five in some organizational concepts. For example, the HEXACO accounts for more variance in workplace delinquency, likelihood
to sexually harass (Ashton & Lee, 2007), unethical business decisions (Ashton & Lee, 2008), and job performance in specific jobs, such as those that involve caregiving (Johnson et al., 2011). As a result, benefit can be seen from using the HEXACO personality model in the workplace and examining the trait of HH.

Table 1

*Descriptions of HEXACO Personality Traits*

<table>
<thead>
<tr>
<th>HEXACO Traits</th>
<th>Low Levels</th>
<th>High Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honesty-Humility</td>
<td>Greedy, corrupt, cunning and vindictive</td>
<td>Sincere, loyal, modest and fair</td>
</tr>
<tr>
<td>Emotional Stability</td>
<td>Anxious, vulnerable and emotionally sensitive</td>
<td>Self-assured, brave, independent</td>
</tr>
<tr>
<td>Extraversion</td>
<td>Reserved, shy and withdrawn</td>
<td>Cheerful, talkative and lively</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>Irritable, stubborn and impatient</td>
<td>Calm, patient and friendly</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>Disobedient, frivolous and careless</td>
<td>Meticulous, self-disciplined and organized</td>
</tr>
<tr>
<td>Openness to Experience</td>
<td>Narrow-minded, conventional and conservative</td>
<td>Creative, inventive, imaginative</td>
</tr>
</tbody>
</table>

**True Scores of Personality**
In order to measure HH, as encompassed within the HEXACO model of personality, the true scores of HH must be considered. According to the Classical Test Theory of measurement, whenever you are attempting to measure any construct, in that measurement you will have a true score, an observed score and error (Novick, 1966). For example, every individual has what is classified as their true score of each personality trait. This true score indicates an individual’s true/actual level of that trait. If personality were to be perfectly measured by a scale, this score is what would be measured. However, when measuring any construct, there is error that causes the observed score to differ to some degree from the true score (Novick, 1966). Although scales try to attain a high reliability, there is always some random error that occurs in measurement each time it is used (Carmines & Zeller, 1979). Also, measures aim to attain the highest validity possible, in terms of accurately measuring the true score of a construct, however this is never fully achieved (Carmines & Zeller, 1979). Therefore, although getting an exact measure of one’s true score is unlikely to occur, measures of personality try to tap into an individual’s true score for personality with as much accuracy as possible.

When trying to measure an individual’s true score of personality, common methods used are observer reports of personality as well as self-reports of personality. Neither is a perfect form of measurement. Both self-reports and observer reports can fall prey to various response biases that can affect their ability to detect an individual’s true score (Dodorico McDonald, 2008). However, within the personality judgement literature, self-reports are often used as a measure of one’s true score. Schmid Mast, Bangerter, Bulliard, and Aerni (2011) classify personality judgement accuracy in terms of the correlations between self-report and other/judges’ reports of personality, with this accuracy using self-reports as a central measure of “true score”. Although what Schmid Mast et al. (2011) call accuracy could be thought of
as convergent validity (i.e., the correlation between two different ways of measuring one’s personality), in the person-perception literature the correlation between self-reports (or another form of ‘true score’) and observer ratings is typically referred to as personality judgement accuracy. Research has shown that self–other agreement is an accurate criterion when examining judgement accuracy (Funder & Dobroth, 1987). In addition to this, self-reports are believed to be an accurate assessment of one’s true score of personality because individual’s themselves should have the most information about themselves in terms of behaviours they engage in and the motivations behind those behaviours (Paulhus & Vazire, 2007). This literature provides support for the use of self-reports as a measure of “true score”. Following this, self-reports will be a key method of true score measurement for personality in this research.

Although self-reports do have merit as a “true score” measure, it is beneficial to have more than one method of “true score” measurement (Dodorico McDonald, 2008). Because of this benefit, true scores will also be measured in this research using expert observer ratings. HH subject matter experts serve as a more expert judgement of HH and therefore they will be used as an interview specific observer report for comparison of interviewer judgements as well. Accuracy in this sense would be determined by comparing the expert observer reports with ratings made by the interviewer. Using both types of measures of “true score” will help provide a more thorough judgement of accuracy in HH judgement.

**Accuracy of Personality Judge Ratings**

Self-reports are often compared to a judge’s ratings as a form of accuracy (Schmid Mast et al., 2011). The goal is to determine how similar a judge’s rating is to an individual’s true score of personality, which is here dictated by that individual’s self-report and expert observer reports. For a judge to make accurate appraisals of an individual’s personality the judge needs to
view personality cues elicited from the individual. These cues could be based on behaviour that
the person is engaging in or based on things that the person says that lend information to
the judge about the individual’s personality traits. The process in which these cues are elicited
and detected, and consequently lead to judge rating accuracy, is described by Funder’s (1995)
Realistic Accuracy Model.

**Funder’s Realistic Accuracy Model**

For a judge to detect the personality traits of others, such as the trait of Honesty-Humility, that judge must go through a variety of steps. As mentioned above, one model that describes the steps involved in this process is the four-stage model purposed by Funder (1995), the Realistic Accuracy Model (RAM). According to the RAM, for a personality trait to be detected in an interview, an interviewee must emit cues of that trait (the relevance stage) in front of the interviewer (the availability stage), and the interviewer must pay attention and notice these cues (the detection stage), and must know that the cues are related to the specific personality trait and use the cues to make that specific trait attribution (the utilization stage). This process is depicted in Figure 1.

![Figure 1. A model of the process of accurate personality judgement](image)

Funder (1995) states that whether a trait is easy or difficult to detect depends on four components. The first component is visibility. If a trait is highly visible and it has cues that are presented easily and often, it is easier to detect. An example of a highly visible trait is
Extraversion, as this trait can be easily detected through simple interactions with others. A trait that is not as visible is the trait of Openness to Experience (Connelly & Ones, 2010). This trait encompasses factors such as inquisitiveness and being imaginative, and these factors are often not inherently visible. The second component is frequency. As cues are presented with higher frequencies, they become easier to detect. This effect is due to increased relevance and availability of the trait’s cues. In other words, when cues are being presented at higher frequencies, they are more readily elicited and therefore, made more readily available for detection by the judge. The third component is whether the trait is presented ambiguously. If cues of a trait are ambiguous, that trait can be difficult to detect because judges may be unsure if the cues an individual is eliciting are related to that specific trait or not. For example, if someone is not very talkative, they might be introverted, or the situation may have made them feel nervous. The fourth and final component is whether people are evaluated based on that specific trait. This component relates to how positively or negatively levels of a trait are judged by others. If an individual is going to be evaluated negatively because they have a certain level of a trait, they will try to avoid displaying cues of that trait. On the other hand, if they know they are going to be evaluated positively by displaying cues of a trait, they are more likely to do so. For example, if people are low on agreeableness, they often won’t be regarded positively in normal day-to-day interactions. Therefore, they may try to appear more pleasant and easy to get along with. To summarize, each of these four components are important for the detection of personality traits and the more of each component that is present, the easier a trait will be to detect.

All of the factors specified above affect the ability for others to detect HH. The problem with the personality trait of HH is that it is what Funder would characterize as a ‘difficult trait’; in other
words, HH is hard to detect. For the trait of HH, some facets, such as one’s sincerity, are not always inherently visible to others and can be quite ambiguous (it is often hard to tell if someone is truly being sincere). In addition, behaviours associated with HH are not always presented frequently, i.e. displays of modesty or lack thereof. Finally, people tend to be evaluated negatively if they are greedy, insincere and unfair, and therefore, individuals generally avoid openly indicating if they are low on HH. These factors combined make the trait of HH hard to detect.

Research provides support for HH being a ‘difficult trait’. Without ample time to get to know someone it is difficult to judge someone’s HH level (Ashton & Lee, 2010). The correlation between acquaintances’ self- and other-reports of HH is only .22, and it typically isn’t until individuals are friends, family members or in romantic relationships (self-other correlations of .30, .49 and .60, respectively) that they can judge another’s level of HH with some degree of accuracy (de Vries, Lee, & Ashton, 2008). Therefore, when individuals have not spent a great deal of time getting to know each other, HH is a difficult personality trait to judge. With time, people are able to judge HH with greater accuracy due to increased exposure to that individual in different situations, and consequently, increased chances to view HH cues emitted from that individual overtime. A problem arises, therefore, when one is considering the detection of HH in the interview context, because time with a candidate is short and confined to a specific situation. Therefore, interviewers need to ensure candidates are emitting cues of HH in the short timeframe of the interview so that these cues are available to be detected.

To fully examine the issue of the detection of HH in the interview context, I will conduct two studies: one to examine cues of HH in real interviews, and another to examine the questions
asked in the interview and their ability to elicit cues of HH. Each of these studies, and further explanation behind their importance, is explicated below.

**Study One**

**Interviewers Judgements of Honesty-Humility**

To detect personality in an interview, interviewers must make judgements of the candidate’s level of a trait and they do this based on cues the candidate emits. Cues of a candidate’s personality could come from many different things. Interviewers may make judgements based on what the candidate says during or before/after the interview, how the candidate says these things, and/or the candidate’s non-verbal communication. In this context of this research, cues of personality will focus on information the interviewee verbally provides in their answers to interview questions. Ideally, the interviewer would notice the cues provided to them by this information and then make a rating of a personality trait that approximates the candidate’s “true score”, here the candidate’s “true score” of HH. However, in a more realistic situation, an interviewer’s detection will be imperfect. This imperfect rating makes it important to know the actual level of accuracy that is achieved by interviewers. If adequate accuracy of HH rating is achieved by trained interviewers, that would demonstrate that detection of HH in the interview is feasible. Therefore, the first research question for this study is:

R1: Can trained interviewers judge HH with some degree of accuracy?

**Between Rater Accuracy**

Although individual interviewer accuracy is important, it is also important for interviewers judging HH to be similar in accuracy to one another, in other words, that the judgements are reliable. Often in a selection system more than one interviewer will be used
to conduct interviews. If this is the case for a personality/HH-based interview, the interviewers must be able to produce similarly accurate scores. If the scores are not similar in accuracy, then candidates will be selected at different rates simply due to the interviewer who conducted the interview. To avoid this problem, research needs to investigate how much variance there is between individual interviewers in HH detection and the reliability of their scores. If multiple interviewers can conclude an interview and give a candidate a similarly accurate score on HH, then candidates would be selected at similar rates and it could be said that the interviewers provide reliable scores regardless of who is conducting the interview. Therefore, the second research question of this study is:

R2: Can trained interviewers judge HH with some degree of reliability?

**Using Interviewer Accuracy and Cues of HH**

To determine interviewers’ personality rating accuracy, their ratings of candidates’ levels of HH must be compared to the candidates’ “true scores”. This accuracy information can be useful in more ways than one. First, this information demonstrates how accurate interviewers are at detecting HH, which in of itself is important.

Finally, this information can be used to help determine cues of HH. If interviewers are accurate at detecting HH, then the cues they used to detect HH may be valuable cues for interviewers in general to use to detect HH in the interview. Therefore, information obtained about the accuracy of interviewers will help to determine useful cues of HH.

**Cues of Honesty-Humility**
The Realistic Accuracy Model’s stages revolve around the elicitation and detection of personality trait cues. Without these cues, personality detection cannot occur. In addition, for an interviewer to be able to detect a difficult trait like HH in the short time frame of an interview, the interviewer must understand what cues to look for. A problem then arises, as research has yet to fully identify good cues of HH. Lee and Ashton (2012) have suggested some valid signs of HH such as those who try to beat the system, engage in instrumental ingratiation, gamble, are sexually unfaithful, partake in conspicuous consumption, have an above the law mentality, and have contempt for other groups. The problem with many of these cues is they are hard to detect in an employment interview. For example, it is very unlikely that someone is going to discuss sexual infidelity or a gambling problem in an interview. In addition, if someone tries to beat the system, they likely are not going to make that fact explicit in a job interview. Therefore, not only do more cues of HH need to be identified, interview specific cues needed to be identified.

The integrity literature can help us with possible cues of HH because integrity tests seek to discover individuals who are dishonest and lack integrity. Trying to tap into these traits is often done through methods such as inquiring about concepts such as punitiveness and projectiveness. These relate to honesty, because those who are more honest tend to be more punitive and those who are dishonest tend to think others are close to themselves in terms of levels of dishonesty, in other words project their honesty level onto others (Cunningham, Wong & Barbee, 1994). Although these cues may be helpful to detect HH, it is important to remember that integrity and HH do not overlap completely, as stated earlier, and therefore cues of HH may somewhat differ from cues of integrity.

Because the knowledge we have of interview specific HH cues is scarce it is important to find cues of HH in the interview. These cues could be utilized by an interviewer to help them
accurately detect HH in a candidate and the more possible cues for detection, the better the accuracy of an interviewer can be. Therefore, the third research question for this study therefore is:

R3: What are interview specific cues of HH that can be elicited in naturalistic employment interviews?

**Generalizability Theory**

To be able to look at the cues in the interview, the variance accounted for by the interviewer must be examined with Generalizability Theory. As specified by Shavelson and Webb (1991), generalizability (G) theory relates to how dependable behavioural measurements are. Its main goal is to determine the variance accounted for by various things i.e., items, raters, etc. G theory looks at variance accounted for to determine how many items or raters are needed to obtain dependable or generalizable scores. Classical test theory, mentioned earlier, is a branch of G theory, with G theory being able to account for more than one source of error/variance in one analysis. For example, in the research explained here, there are various sources of variability. These sources include variability due to the interviewees in their levels of HH (which is classified as the object of measurement), interviewers’ variability between each other in rating an interviewee’s level of HH, and the combination of interviewer and interviewee variability and error. G theory can calculate each individual variability (not including interviewer x interviewee variability which is inseparable from error). Ideally the main source of variability for this specific study would come from differences in the interviewees’ levels of HH and not from differences in the interviewers’ ratings of each interviewee’s HH. G theory can help examine if this is the case and therefore answer this study’s final research questions of:
R4: In interviews, what percentage of variance is due to differences in interviewees’ scores of HH and what percentage is captured by differences in interviewers’ ratings of interviewees’ HH levels? In other words, can interviewers reliably detect HH?

**Study One**

**Proposed/Partially Completed Methodology**

**Participants.** Interviews were videotaped prior to this study and the participants in the videos, the interviewees, consisted of University of Guelph students applying for a research assistant job in Dr. Powell’s lab. There was a total of 82 candidates from various programs around campus. 83% present were female and 17% were male. Ages ranged from 18 to 41. 43 were Caucasian (52%), two were Aboriginal/First Nations/Metis (2%), five were Arab (6%), five were Black/Africa/Caribbean (6%), 12 were South Asian (15%), five were Southeast Asian (6%), 10 classified their ethnicity as other (13%). This sample provides a unique opportunity for research as they provide such a large quantity of videotaped, real interviews.

**Interviewers.** The interviewers consisted of four research assistants in Dr. Powell’s lab. They were trained to administer the interview questions, ask probes, and score the candidates on behaviourally anchored rating scales. These scales were only used for hiring purposes and were not examined for the purpose of this research. All of the interviewers were female.

**Raters.** The SMEs that watched the videos for this study included myself and three research assistants hired by Dr. Powell. All of the raters received training to become experts on the trait of Honesty-Humidity. All raters were female.

**Procedure.** The participants were videotaped as they completed an interview for a research assistantship with Dr. Powell. They answered four questions which are detailed in Appendix
A. They were also probed to provide further information, if deemed necessary by the interviewer. Following the interview, the participants completed a self-report of their personality. This data was originally collected for another study by Dr. Powell. Other items were also collected for Dr. Powell’s study that will not be used for this study’s purposes; specifically the Honest Impression Management Scale (HIMS), the Interview Faking Behaviour – Short scale (Bourdage, Roulin & Tarraf, 2017), the Measure of Anxiety in Selection Interviews (MASI; McCarthy & Goffin, 2004), STAR (Powell, Roulin, & Bourdage, 2018), and the Machiavellian Personality Scale (MPS; Dahling, Levy & Whitaker, 2009).

Following the collection of the data specified above, SMEs were provided training to become experts in the trait of HH. After training, the SMEs watched the videotaped interviews and rated each of the candidates on a Relative Percentile Method (RPM) scale. They rated each of the candidates individually. Each candidate was rated on their overall level of HH as well as each facet of HH. These facets are sincerity, fairness, modesty and greed-avoidance. In addition to rating the candidate, SMEs recorded any cues of HH they thought the interviewee elicited in the interview and indicated which facet they thought each cue corresponded with.

The SMEs’ ratings were analyzed to differentiate variance due to interviewees’ levels of HH, differences between SMEs’ ratings of each interviewee on HH, and error. This analysis occurred at the trait as well as facet level (see Table 2 for the preliminary results). (Please note that the methods described up to this point have already been carried out). In addition to this, self-ratings and the averaged SMEs’ ratings will be correlated, and separately, self-ratings and individual SME ratings will be correlated to determine each SMEs accuracy in detecting the interviewees’ HH. The cues provided by
SMEs will be examined if, after analyzing the ratings, the SMEs provided relatively reliable and accurate HH ratings.

**Materials.**

*Training.* Training for the SMEs consisted of having them read descriptions of HH and its facets, view self-report items that correspond with HH, as well as read chapters of Lee and Ashton’s (2012) book on HH. They also completed a test to demonstrate their knowledge of HH. This test consisted of a list of the HEXACO-100 items. The SMEs had to indicate which items they thought corresponded with HH and each of its facets. Any mistakes made on this test were identified and corrected by explaining to the SMEs which items should have been selected and why. The training took SMEs approximately 2 hours. The training document provided to SMEs provided in Appendix B.

*Relative Percentile Method (RPM).* The scale used by SMEs to rate each candidate’s level of HH was an RPM scale. This scale ranges from 0-100 (0 = very low, 50 = average, 100 = very high). SMEs rated each candidate by deciding whether they thought the candidate was very low, average or very high on HH when compared to others in the general population. A depiction of the scale as well as instructions given to the SMEs for them to complete their ratings are provided in Appendix C. The RPM scale type has been judged as advantageous in past research (Goffin, Jelley, Powell, & Johnston, 2009).

**Measures.**

*Personality.* Personality was measured by the self-report 60-item HEXACO (Ashton & Lee, 2009). This measure has an internal consistency reliability ranging from .73 to .80 for the scales of each of the personality traits (Ashton & Lee, 2009). The internal consistency reliability
of the HH scale specifically is .76 (Ashton & Lee, 2009). The items of the HEXACO are presented on a 5-point Likert scale (1 = strongly disagree, to 5 = strongly agree). This measure was used to obtain self-reports of HH from the interviewees and is used in this study to reflect an individual’s true score of personality. This measure was filled out by interviewees’ after their interview, in a separate room. They were informed that the results from this survey would not affect their chances of getting the position.

**Proposed Analyses**

**G-Study.** In generalizability theory, analyses are conducted for a generalizability study (G-study). This G-study examines the sources of measurement error within a study design. As this is a one-facet design, meaning the universe is defined by one source of measurement error, the variance examined was among interviewees, raters, the interaction of interviewees and raters and random error. In the G-study, ANOVA was used to partition and examine each of the variances. It examined the variance provided by the interviewees (o), the interviewers (i), and the interaction (o x i) combined with the error/variability due to randomness.
Table 2

*Variance accounted for by Interviewers vs Interviewees*

<table>
<thead>
<tr>
<th></th>
<th>Variance Accounted For</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
<td>Interviewee</td>
</tr>
<tr>
<td></td>
<td>28.5%</td>
</tr>
<tr>
<td>Sincerity</td>
<td>28.5%</td>
</tr>
<tr>
<td>Fairness</td>
<td>18.0%</td>
</tr>
<tr>
<td>Greed-Avoidance</td>
<td>28.6%</td>
</tr>
<tr>
<td>Modesty</td>
<td>25.3%</td>
</tr>
</tbody>
</table>

**Reliability Analysis.** To determine the reliability of the SMEs intraclass correlations were calculated. These correlations demonstrate the reliability present when only a single SME rated the candidates’ HH and each HH facet, as well as the reliability when all of the SMEs rated the candidates’ (see Table 3 for preliminary results).

Table 3

*Reliability of One versus the Four Raters*

<table>
<thead>
<tr>
<th>Intraclass Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>HH</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
SME Accuracy. The accuracy of HH detection by the SMEs was assessed in terms of specific personality trait accuracy as specified by Davis and Kraus (1997). Schmid Mast et al. (2011) termed this type of accuracy as the correlation between the judges’ ratings and interviewees’ self-reports of a single personality trait across all of targets in the employment interviews. This type of accuracy examines how well the judge detected variation between the interviewees on a specific personality trait, here the trait being HH. The same correlations were also completed for each of the facets individually. See Table 4 for preliminary results.

Table 4

*HH Overall and Facet Score Correlations*

<table>
<thead>
<tr>
<th>Specific Personality Trait Accuracy</th>
<th>SME 1</th>
<th>SME 2</th>
<th>SME 3</th>
<th>SME 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Honesty-Humility</td>
<td>.01</td>
<td>.02</td>
<td>.29</td>
<td>.22</td>
</tr>
<tr>
<td>Sincerity</td>
<td>.00</td>
<td>-.07</td>
<td>.05</td>
<td>.03</td>
</tr>
<tr>
<td>Fairness</td>
<td>.01</td>
<td>.11</td>
<td>.15</td>
<td>-.11</td>
</tr>
</tbody>
</table>
Cues. More variance in the scores from the G-study came from the interviewee than the interviewer and together the raters demonstrated adequate reliability. This information demonstrates that the cues provided by the SMEs could be useful to detect HH in the interview. However, the SMEs have accuracy correlations were not greater than .5 with the candidates’ self-reports. This information demonstrates that there is more error than true scores present. Then the cues will be sorted and categorized based on facet to demonstrate cues that could be useful in an interview to detect HH and its facets. A list of these cues will be provided in the results for use in future research and application.

Study Two

Study One is important because it helps to answer the question of: can HH be detected in the interview? Study Two however, will focus on assessing conditions that improve accurate detection in the interview for a broader group of participants. For the detection of HH to increase, the interview questions must elicit HH cues, thus increasing the chances of the judge to detect and utilize those cues for HH judgement accuracy. Therefore, it is important to know which questions can best achieve this goal. Levashina et al. (2014) in a review of the literature on employment interviews discussed that research needs to be done on the design of personality-based interviews and one of the areas that needs to be examined specifically is the questions asked. The questions need to be examined to determine what kinds of questions can be used to best elicit trait relevant information and information consistent with how an individual

<table>
<thead>
<tr>
<th></th>
<th>G-avoidance</th>
<th>Avoidance</th>
<th>.08</th>
<th>-.05</th>
<th>.08</th>
<th>.19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modesty</td>
<td>.12</td>
<td>.18</td>
<td>.28</td>
<td>.06</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
would act day-to-day, so the interviewer can accurately detect the candidate’s personality. Following this, the research question that will be addressed in this study is: what questions best elicit cues of HH and thereby lead to the greatest degree of HH detection accuracy?

**Situation Strength**

One important factor to consider in designing questions to elicit HH cues is the concept of situation strength. Situation strength is defined as external cues in a situation regarding the desirability of certain behaviours; these cues cause psychological pressure on individuals to act in a similar way as others, regardless of how they themselves would typically act (Meyer, Dalal & Hermida, 2010). By this definition, if a situation elicits many cues of behavioural desirability, the situation will be strong, and personality will not play a large role in behaviour. Instead, individuals will behave in a way that the cues of the situation tell them to behave. An example of this is an extravert in a quiet office. If they are silent, it is not because they typically behave this way based on their personality. This behaviour is likely because the situation signals to them that it is not appropriate to be loud and outgoing. If situation strength were to be applied to the trait of HH, and a strong situation signalled to an individual that high HH is most appropriate, then that individual would feel as though they needed to appear sincere, fair, greed-avoidant and/or modest, regardless of if they are high on these facets and consequently high on HH. The situation therefore constrains what HH cues an individual will elicit, based on what that individual thinks is appropriate and/or desired.

Because of how situation strength affects behaviour, it could affect the availability of HH cues in an interview. If the questions asked create a situation in which the candidate is restricted in the possible answers they can give, it may affect the cues the candidate can emit in response to
those questions. In other words, the questions may signal to the candidate to how they should respond and therefore genuine HH cues are not elicited. Based on this idea, situation strength seems important for a personality interview. Despite this importance, research has not examined how the questions asked in an interview could affect situation strength and therefore personality cue availability and the overall detection of specific traits. My research will look at varying the situation strength in interview questions. My aim is to demonstrate the possible effect of situation strength on HH detection.

**Structured Interviews vs Unstructured Interviews**

To deal with the issue of situation strength in interview questions, a specific interview type must be selected, as the type of interview affects the types of questions asked. The two main types of interviews are structured interviews and unstructured interviews. Based on the concept of situation strength, it should follow that unstructured interviews are beneficial for personality, and research supports this idea. Blackman (2002) found that unstructured interviews lead to better judgement of personality when compared to more structured interviews. With this in mind, it might seem best to look at question types in the unstructured interview. However, unstructured interviews are less reliable and less valid tools when compared to structured interviews (Levashina et al., 2014). Therefore, it is valuable to investigate the questions asked in structured interviews to try to increase the ability for personality traits such as HH to be elicited in structured interviews. This research will try to accomplish this by decreasing the situation strength of the questions (with regards to Honesty-Humility) within the structured interview.

**Structured Interview Questions**
Within structured interviews there are many different types of questions that can be asked. The questions asked are typically broken into the categories of behavioural questions, situational questions, and technical job knowledge questions. Job knowledge questions ask about aspects of the job that candidates are expected to know to properly perform the job, behavioural questions ask candidates about previous behaviour on the job, and situational questions ask how a candidate would behave in a hypothetical job situation (Levashina et al., 2014). Behavioural and situational questions can be created to tap into a variety of knowledge, skills and abilities, including one’s personality.

For the detection of personality, behavioural questions may have some overall benefits. Some research has found that behavioural questions measure personality traits (Krajewski, Goffin, McCarthy, Rothstein & Johnston, 2006), and that they are structured like personality inventories because they are past-oriented (Levashina et al., 2014). In addition, behavioural questions have slightly higher criterion validity than situational questions (Levashina et al., 2014). Therefore, it may be beneficial to use behavioural questions in a personality interview. Following this, behavioural questions will be the focus of this proposal from this point forward.

**General vs Specific Personality Questions**

When creating questions to elicit personality cues, questions are likely to be created to tap into a specific trait (e.g. Van Iddekinge et al). A problem with this approach, however, is that interview questions created to tap into specific personality traits may create a strong situation. This strong situation may result because the question provides cues to candidates as to how they should answer. For example, if an interviewer is trying to elicit cues of Extraversion from a candidate and they ask: “Tell me about a time when you had a work environment you
enjoyed. Describe your interactions with your co-workers.” This question will likely to only elicit a small range of answers, and depending on the organization, the candidate may feel that they need to appear extraverted or introverted to get the job. However, if the question was created to be more general, therefore decreasing situation strength, sincere personality cues may be elicited more readily. If a candidate was asked “Tell me about a time when you had a work environment you enjoyed. What about this environment made it enjoyable?” the question could elicit cues of Extraversion but could also elicit cues of other traits such as Conscientiousness. Simply by eliminating one constraint, the focus on co-workers, the question becomes broader and more readily able to elicit a variety of personality cues. For the trait of HH, if questions evidently tap into HH’s facets, candidates will likely feel as though they need to answer in a certain way. However, if the question is general enough that HH cues be expressed more naturally, then the interviewer should be able to make a more accurate judgement of HH. In more general terms, questions created to tap into multiple personality traits increase the range of possible answers the candidate can give and, therefore, the number of genuine personality cues that can be elicited.

The general personality questions described above can be compared to those asked in unstructured interviews. These types of questions create an environment where the interviewee is able to speak more freely because they feel less constraints on how they should answer. Tying this back to the RAM (Funder, 1995), general personality-tailored questions should elicit more cues relevant to the individual’s “true” level of specific personality traits and make these cues available to the interviewer for them to judge. Therefore, when using behavioural questions as specified earlier, general questions (versus specific) should lead to better HH detection overall.
H1: Interviewers will have greater accuracy in detecting HH in response to general personality-tailored behavioural interview questions as compared to behavioural interview questions that tap specifically into HH.

**Probing**

Decreasing the strength of the situation in an interview question is one way to help elicit personality information from a candidate. Another way to elicit personality-related cues could be through probing. Probes are a type of follow-up question used to obtain a more elaborate response from an interviewee if their answer originally seemed vague or incomplete (Levashina et al., 2014). Research has found that probing may help to increase the accuracy of information gathered in an interview (Motowidlo et al., 1992; Schmidt & Conaway, 1999). There is some support that this increased accuracy also applies to personality interviews. Blackman (2002) found that in interviews that included more probing, interviewers made more accurate personality judgements. In Blackman’s research the interviews that included more probing were unstructured interviews, however, this finding may also apply to structured interview questions. By asking probing questions, interviewers may obtain more accurate information about the overall behaviour the interviewee engaged in and their reasons behind their behaviour. In other words, the probes may help the interviewer to elicit more trait relevant cues. Specifically, this research is going to focus on asking probes that inquire about an applicant’s reasons behind their behaviours and feelings they had when engaging in specific behaviours. By asking these types of questions candidates must reflect on why they did certain things which may help to pull out cues of HH. For example, if a candidate is asked why they chose a certain approach, they may make a comment along the lines of ‘I did this because I wanted to help my teammates.’ This could tap into the fairness facet of HH and therefore lead to
higher, and more accurate ratings of that interviewee’s HH level. If the probes elicit more
cues from interviewees, then the interviewer should be able to make a more accurate judgement
of the interviewee’s HH. Also, this research will focus on adding probes, but keeping probes
consistent which each question, which will allow for the structured interview environment to be
upheld, while still giving increased opportunities for information to be elicited and more free
speech to occur like in unstructured interviews.

H2: The use of probes (versus no probes) will lead to increased accuracy in HH detection.

Probing may be important for eliciting more personality information and cues, however it may
be especially important for general personality-tailored questions. For these questions, candidates
may provide an array of different information, tapping into a variety of traits, because the
questions are quite broad. Although this is the goal of asking this type of question, the
information the candidate discusses may stray away from information that will lead to the
elicitation of cues that the interviewer desires. If this happens, it could be difficult for the
interviewer to detect the trait they were hoping to elicit. By adding probes, the
interviewer gives the candidate more opportunities to emit relevant personality cues. Also,
probes may help to delve deeper into what the candidate has said to ensure the interviewer is
not making an inference to what the candidate means by their statements. For example, if a
candidate were to say they were influencing their friend to do something, an interviewer may
assume that the candidate is lower on HH. However, by asking a probe, the interviewer may find
out that the candidate is trying to influence their friend to do something because the
candidate knows it will benefit their friend in the long run. Therefore, the interviewer may not
judge the candidate as low on HH simply because a probe was asked, and the interviewer was
able to make a better judgement as a result. In sum, probes may help the interviewer obtain
more accurate personality cue information from general personality questions specifically and therefore make the interviewer better at detecting HH.

H3: General personality-tailored behavioural questions in combination with probing will lead to the most accurate HH detection as compared to any other condition.

Study Two

The data collected and analyzed in study two will be used to answer the hypotheses of study two, however the videos in Study Two will also be analyzed for HH cues to help provide a wider variety of cues to answer the Study One research question of: what are cues of HH presented in the interview?

Proposed Methodology

Phase 1.

Participants. Participants for Phase 1 will be recruited by placing flyers around the University of Guelph Campus as well as ads in University of Guelph student Facebook groups. Participants will be compensated $20 for their participation. For this study I aim to obtain 30 participants. This number will hopefully allow me to obtain variability in scores on HH while also providing me with the time to carefully review the videos and pick the best ones for Phase 2.

SMEs. Subject matter experts for this study will hopefully be the same SMEs as those from Study One. If this does not occur, more RAs will be recruited.

Procedure. Participants will come into Dr. Powell’s lab and participate in a mock job interview. This interview will be filmed for use in Phase 2. Participants will be asked to answer questions as if they were applying to an Accounts Manager job. This is a job that students would have the
qualifications for and therefore could envision themselves applying for in real life (see Appendix D for the job ad that will be given to participants). Participants will answer a total of four interview questions. They will answer two questions corresponding to general personality-tailored behavioural questions and two behavioural questions tapping specifically into HH (see Appendix E for some of the questions being considered). The general personality questions will aim to tap into more than one personality trait at a time, to reduce the situational strength of the question. The general questions will be asked first, followed by the specific questions. This process will enable the candidates to go from a broad question to a more specific question, hopefully making it easier for them to answer the more specific question. The candidates will be informed that because the questions are so similar but will be split into separate videos that they should treat each question as the interviewer did not hear the answer to the question before it. This process will allow candidates to use the same answer to the general and specific questions if they see fit. Following each question, for both general and specific questions, participants will be probed to provide addition information to each interview question. Standard probes will be provided to the interviewer (see Appendix E) and will asked to every candidate. The standard probes will be used to obtain the most complete answer to the question as possible while also being mindful of the length of the interview for participants in Phase 2. Following the interview, the participants will provide self-report ratings of HH to serve as one “true score” for HH judgement accuracy. They will also fill out a RPM scale to indicate where they think they fit the scale in terms of HH, and each of the facets, to help with comparisons across similar scales (participants in Phase Two will be ratings these interviewees on an RPM scale). The participants will be provided definitions of each of the facets and the trait to be able to rate themselves. HH will be labelled as trait X so that participants do not focus on
the idea of honesty versus dishonesty. After they have completed their ratings they will receive compensation.

**Selection of Videos.** Between Phase 1 and Phase 2, videos will be selected from the Phase 1 for use in Phase 2. These videos will be selected based on the participants’ self-reports to ensure the a range of HH scores in the videos for Phase 2. In addition, the videos will be watched to see which candidates are the easiest to rate. SMEs will watch the videos and rate each candidate’s HH level on an RPM scale. The videos in which the SMEs were most accurate in rating HH (when compared to the candidate’s self-report) will be chosen. Five videos will be chosen for Phase 2 in total. Of the videos chosen, they will be split into condition videos based on question type and whether or not probing was provided. To do this, each interview video will be first edited/split into two videos, each based on question type. Then a copy of each question type video will be edited to create a copy that does not contain the answers following any of the probes. For the no-probe condition it is assumed that if a candidate were to be asked a question and not probed the original answer they gave would remain the same, and therefore, removing any answers they give following the probes does not affect what their original answer would have been. Therefore, for each interviewee four videos will be created, one for each condition.

**Cues.** Analogous to Study One, along with rating HH of the candidates in the videos, SMEs will also provide cues for each facet of HH. These cues will be used alongside cues from Study One to gain a more comprehensive idea of HH cues in interviews.

**Measures.**

**Personality.** Personality of candidates in Phase 1 will be measured using the Honesty-Humility scale from the 200-item HEXACO. Internal consistency reliability for this scale is .90. The
Reliabilities are .75, .81, .85, and .78 for the facets of Sincerity, Fairness, Greed-avoidance and Modesty respectively. The items are presented on a 5-point Likert scale (1 = *strongly disagree*, to 5 = *strongly agree*).

*Relative Percentile Method (RPM).* The scale used by participants to rate their own level of HH was an RPM scale, similar to the one used in Study One. This scale ranges from 0-100 (0 = very low, 50 = average, 100 = very high). Participants rated themselves by deciding whether they thought they were very low, average, or very high on HH and each of its facets. The RPM scale type has been judged as advantageous in past research (Goffin, Jelley, Powell, & Johnston, 2009).

**Phase 2.**

*Participants.* Participants will be recruited for Phase 2 through M-Turk. Based on a power analysis, 788 participants will be needed (based on a small Cohen’s d). Participants will be rewarded $4.00 for their compensation.

*Procedure.* Participants will be split into four conditions based on a 2x2 study design. The conditions will be split based on question type (general vs specific) and probing (probing vs no probing). When participants first open the survey, they will complete two audio checks. This will involve the participant listening to two short, couple-second sound clips to ensure their sound is working. They will have to answer a question corresponding with each clip as to what they heard. If they are unable to answer these correctly then it is assumed that they will not be able to adequately hear the videos and therefore they will not continue. Compensation will not be rewarded to those who do not pass the audio checks.
Once participants have passed the audio checks, they will complete an online consent form. Once they have provided their consent, participants will be given a background of the facets of HH as well as be provided some cues for each of the facets (as collected from Study One). They will then be provided a definition of the trait that the facets compose into, however it will be labelled as Trait X so that raters are not focused on Honesty/Dishonesty. This strategy also should help to ensure that participants are focused on rating based on the definitions provided to them. See Appendix F an example of what will be provided to participants. The participants will watch the five Phase 1 videos that correspond with their condition and rate each interviewee’s level of HH by providing a general HH score and score for each of HH’s facets (sincerity, fairness, modesty and greed-avoidance). The ratings will be done on the same RPM scale as used by the participants in Phase 1, however they will be rating the candidate in the video, and not themselves. The rating scale for each of the facet will be present in a random order that will be different between participants. The definitions of the trait and its facets will be provided above each scale to ensure that the participants do not simply forget the definition, therefore affecting their ability to rate the videos. There will also be attention checks that correspond with each video to ensure that participants are paying enough attention to keep their data. These questions will be about something the interviewee said in the video and the participant will have to have been paying attention to get the question correct. If an attention check is not answered correctly, the participant will be paid for their time up to that point and will not continue on. Participants that do not pass all of the attention checks will not have their data included in analyses. Following the ratings, participants will be asked if they believed their data should be included in research analyses based on the attention they paid to the videos. Following this,
participants will receive compensation. The entire procedure should take participants approximately 40 minutes.

**Materials.**

*RPM.* The scale used by participants to rate the candidates’ level of HH and each facet was an RPM scale. This scale ranges from 0-100 (0 = very low, 50 = average, 100 = very high).

**Data Cleaning**

Participant data will be deleted based on two rules. First, if participants do not pass all the attention checks their data will not be analysed. Second, if any participants stated that their data should not be used because they did not pay attention, their data will be deleted.

**Proposed Analyses**

**Condition Comparison.** Study two will use ANOVA to distinguish the differences between the four conditions in the 2x2 study design (general vs specific and probing vs no probing). The cells to be compared are depicted in Table 5. To test $H_1$ (the effect of general vs specific questions), two cell comparisons will occur. First, cells one and three will be compared, and then cells two and four will be compared separately. For the hypothesis to receive complete support, cell one must have a significantly higher correlation compared to cell three, and cell two must be significantly higher than cell four. To test $H_2$ (probes vs no probes), two cell comparisons will also occur. These comparisons will be between cell one and two, and then cell three and four. For hypothesis two to receive full support, cell two must have a significantly higher correlation compared to cell one, and cell four must be significantly higher than cell three. To
test $H_3$ (general specific probes are best), the interaction of question type and probes will be tested. For this hypothesis to be supported, cell one would need to have a significantly higher correlation when compared to each other cell individually.

Table 5

*Cell comparisons in ANOVA*

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Probes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Probe</td>
</tr>
<tr>
<td>General</td>
<td>1. General, No Probes</td>
</tr>
<tr>
<td>Specific</td>
<td>3. Specific, No Probes</td>
</tr>
</tbody>
</table>

**SME Accuracy.** Accuracy of HH detection will be the same as accuracy of SMEs was assessed in Study 1. Therefore, SME accuracy will be assessed in terms of specific personality trait accuracy as specified by Davis and Kraus (1997). Schmid Mast et al. (2011) termed this type of accuracy as the correlation between the judges’ ratings and interviewees’ self-reports of a single personality trait across all of targets in the employment interviews. This type of accuracy examines how well the judge detected variation between the interviewees on a specific personality trait, here the trait being HH. The same correlations will also be completed for each of the facets individually. The SME accuracy will be used to select videos.

**Participant Accuracy.** Participant accuracy also will be calculated based on the Schmid Mast et al. (2011) definition for accuracy. Specific trait correlations will occur between interviewee’s self-reports and the participants’ RPM rating for HH as well as each of the facets. In addition to
this, SME ratings of the interviewee’s HH will also be correlated, in the same fashion, with participants’ RPM ratings of HH and each of its facets individually. By comparing SME and judge ratings we get an accuracy correlation that is specific to the interview context, as opposed to comparing to an interviewee’s self-report which is based on the interviewee’s knowledge of their behaviour over their entire lifetime. Also, by looking at the facets in the specific trait correlations, as well as HH overall, we can assess whether certain traits are harder to judge than others.

**Cues.** Similar to Study One, if the SMEs in Study Two have accuracy correlations of greater than .5 with the candidates’ self-reports, then the cues will be sorted and categorized based on facet to demonstrate cues that could be useful in an interview to detect HH and its facets. These cues will be added and compared to the list in Study One. Comparing the cues from those in Study One will help to see any possible similarities between the cues elicited by the candidates in each study’s videos.

**Conclusion**

By way of study one and two, my research aims to: examine possible cues of HH in employment interviews, assess the reliability and validity of Subject Matter Experts’ ratings of HH, investigate if general personality-tailored behavioural questions can more readily elicit personality cues than HH specific questions, and evaluate if probing increases personality detection accuracy.

**Limitations**

**Study One**
One of the most important limitations of Study 1 is that I am not able to directly detect what
the best cues of HH are. I have to infer that SMEs who make accurate judgements will have
accurate cues. Unfortunately, this is the best information I have in study one. Second, given the
nature of the interview and the questions asked, I could only examine cues that were elicited in
response to the specific questions asked. It is impossible to say what all the possible cues of HH
in the interview are, and this study will only provide a small selection of possible
cues. Examining cues presented in Study Two as well will help to address this limitation,
however not eliminate it completely. Third, the majority of the participants in the videos in Study
One were female. This gender distribution may affect the cues elicited overall and may
demonstrate more of the cues that females elicit that correspond with HH instead of cues of HH
more generally. Research has demonstrated that females have higher levels of HH compared to
males (Lee & Ashton, 2004) and therefore this may relate to the cues they elicit in the
interview. However, it was beneficial to use this sample to demonstrate some of the cues elicited
in more of a high stakes interview. Finally, all the SMEs were female as a result of who was
willing to volunteer. Cues that females specify and notice related to HH may be different than
those detected by males. However, because general HH training was provided, hopefully this is
not the case.

**Study Two**

The interviews used for this study are mock interviews and therefore, the answers may not be
quite the same as what are provided in real interviews. However, given the low stakes
situation this interview creates, the results found from this study will likely demonstrate the best
that can be expected from interviewers in detecting HH as the candidates are likely the most
honest in this type of interview situation. In addition, given that the participants are on M-Turk,
they may not pay as close attention as an actual interviewer would. However, attention checks will be included to try to ensure that participants are paying close enough attention to notice specific details in the video and therefore also paying attention to HH cues. By including this measure, I will hopefully be able to eliminate participants that were not paying close enough attention to accurately rate HH.

**Implications**

Despite possible limitations, this research has a variety of implications. Study two provides theoretical contributions by offering support for the RAM. If the types of questions asked in interviews do affect the detection of personality, then support is provided for Funder’s stages of relevance and availability. There are also practical implications of this research. This research could demonstrate the types of questions that should be asked in interviews to best elicit personality and indicate if probing should be used to increase HH detection accuracy. It will also help reveal some of the cues that interviewers should pay attention to in order to detect HH.
References


Appendix A – Study 1 Interview Questions

1. Can you tell me about a time when you had to work on a large independent project at either work or school?

2. Can you describe a situation where you used data to answer a research question?

3. Describe a situation in which you were able to successfully convince someone (or several people) to see things your way or to do something you wanted them to do.

4. Tell me about a time when a colleague or subordinate was in trouble and you took it upon yourself to help him or her to deal with the situation.

Appendix B

**Honesty-Humility Subject Matter Expert Training**

*This document contains information to help you become a subject matter expert on the trait of Honesty-Humility. Please read all the provided information to prepare you for ratings of Honesty-Humility in videotaped interviews later.*

Below are descriptions of the trait Honesty-Humility, one of the six traits in the HEXACO model of personality. The five other traits are Emotionality, Extraversion, Agreeableness, Conscientiousness and Openness to Experience (although they will not be apart of this training). There are also descriptions of the four facets that make up Honesty-Humility.
These have been taken directly from the www.hexaco.org

Domain-Level Scale

**Honesty-Humility**: Persons with very high scores on the Honesty-Humility scale avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status. Conversely, persons with very low scores on this scale will flatter others to get what they want, are inclined to break rules for personal profit, are motivated by material gain, and feel a strong sense of self-importance.

Facet-Level Scales

**Honesty-Humility**

The **Sincerity** scale assesses a tendency to be genuine in interpersonal relations. Low scorers will flatter others or pretend to like them in order to obtain favors, whereas high scorers are unwilling to manipulate others.

The **Fairness** scale assesses a tendency to avoid fraud and corruption. Low scorers are willing to gain by cheating or stealing, whereas high scorers are unwilling to take advantage of other individuals or of society at large.

The **Greed Avoidance** scale assesses a tendency to be uninterested in possessing lavish wealth, luxury goods, and signs of high social status. Low scorers want to enjoy and to display wealth and privilege, whereas high scorers are not especially motivated by monetary or social-status considerations.
The Modesty scale assesses a tendency to be modest and unassuming. Low scorers consider themselves as superior and as entitled to privileges that others do not have, whereas high scorers view themselves as ordinary people without any claim to special treatment.

Below are items that correspond with Honesty-Humility in the HEXACO self-report measure.

Sincerity

- I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.

- If I want something from someone, I will laugh at that person's worst jokes. (Reverse-coded)

- I wouldn’t pretend to like someone just to get that person to do favors for me.

Fairness

- If I knew that I could never get caught, I would be willing to steal a million dollars. (Reverse-coded)

- I would never accept a bribe, even if it were very large.

- I’d be tempted to use counterfeit money, if I were sure I could get away with it. (Reverse-coded)

Greed-Avoidance

- Having a lot of money is not especially important to me.
• I would get a lot of pleasure from owning expensive luxury goods. (Reverse-coded)

Modesty

• I think that I am entitled to more respect than the average person is. (Reverse-coded)

• I want people to know that I am an important person of high status. (Reverse-coded)

Please read chapters 1, 2, 3, 4, 5, 9 and 10 of Kibeom Lee and Michael C. Ashton’s book, The H Factor of Personality. You can sign a copy out from me, or Dr. Powell.

Once you have read all the above information and feel comfortable, please complete the exercise below and email/hand it back to me. Completing this exercise actually will demonstrate knowledge in the trait of Honesty-Humility and provide concrete evidence that you are ready to complete the rating task.

If you have any questions about Honesty-Humility, or feel you need more information to be prepared, please email me at: mpike01@uoguelph.ca
To show you are comfortable with Honesty-Humility, look at the items from the 100 item HEXACO and indicate which items correspond to Honesty-Humility and specifically which facet of Honesty-Humility. On the line, indicate S (Sincerity), F (Fairness), GA (Greed Avoidance) or M (Modesty) for which facet the item corresponds with if the item fits within the umbrella of the trait Honesty-Humility.

1____ I would be quite bored by a visit to an art gallery.

2____ I clean my office or home quite frequently.

3____ I rarely hold a grudge, even against people who have badly wronged me.

4____ I feel reasonably satisfied with myself overall.

5____ I would feel afraid if I had to travel in bad weather conditions.

6____ If I want something from a person I dislike, I will act very nicely toward that person in order to get it.

7____ I'm interested in learning about the history and politics of other countries.

8____ When working, I often set ambitious goals for myself.

9____ People sometimes tell me that I am too critical of others.

10____ I rarely express my opinions in group meetings.

11____ I sometimes can't help worrying about little things.

12____ If I knew that I could never get caught, I would be willing to steal a million dollars.
13. I would like a job that requires following a routine rather than being creative.

14. I often check my work over repeatedly to find any mistakes.

15. People sometimes tell me that I'm too stubborn.

16. I avoid making "small talk" with people.

17. When I suffer from a painful experience, I need someone to make me feel comfortable.

18. Having a lot of money is not especially important to me.

19. I think that paying attention to radical ideas is a waste of time.

20. I make decisions based on the feeling of the moment rather than on careful thought.

21. People think of me as someone who has a quick temper.

22. I am energetic nearly all the time.

23. I feel like crying when I see other people crying.

24. I am an ordinary person who is no better than others.


26. I plan ahead and organize things, to avoid scrambling at the last minute.

27. My attitude toward people who have treated me badly is "forgive and forget".

28. I think that most people like some aspects of my personality.

29. I don’t mind doing jobs that involve dangerous work.

30. I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed.
31. I enjoy looking at maps of different places.

32. I often push myself very hard when trying to achieve a goal.

33. I generally accept people’s faults without complaining about them.

34. In social situations, I'm usually the one who makes the first move.

35. I worry a lot less than most people do.

36. I would be tempted to buy stolen property if I were financially tight.

37. I would enjoy creating a work of art, such as a novel, a song, or a painting.

38. When working on something, I don't pay much attention to small details.

39. I am usually quite flexible in my opinions when people disagree with me.

40. I enjoy having lots of people around to talk with.

41. I can handle difficult situations without needing emotional support from anyone else.

42. I would like to live in a very expensive, high-class neighborhood.

43. I like people who have unconventional views.

44. I make a lot of mistakes because I don't think before I act.

45. I rarely feel anger, even when people treat me quite badly.

46. On most days, I feel cheerful and optimistic.

47. When someone I know well is unhappy, I can almost feel that person’s pain myself.

48. I wouldn’t want people to treat me as though I were superior to them.
49. If I had the opportunity, I would like to attend a classical music concert.

50. People often joke with me about the messiness of my room or desk.

51. If someone has cheated me once, I will always feel suspicious of that person.

52. I feel that I am an unpopular person.

53. When it comes to physical danger, I am very fearful.

54. If I want something from someone, I will laugh at that person's worst jokes.

55. I would be very bored by a book about the history of science and technology.

56. Often when I set a goal, I end up quitting without having reached it.

57. I tend to be lenient in judging other people.

58. When I'm in a group of people, I'm often the one who speaks on behalf of the group.

59. I rarely, if ever, have trouble sleeping due to stress or anxiety.

60. I would never accept a bribe, even if it were very large.

61. People have often told me that I have a good imagination.

62. I always try to be accurate in my work, even at the expense of time.

63. When people tell me that I'm wrong, my first reaction is to argue with them.

64. I prefer jobs that involve active social interaction to those that involve working alone.

65. Whenever I feel worried about something, I want to share my concern with another person.

66. I would like to be seen driving around in a very expensive car.
I think of myself as a somewhat eccentric person.

I don’t allow my impulses to govern my behavior.

Most people tend to get angry more quickly than I do.

People often tell me that I should try to cheer up.

I feel strong emotions when someone close to me is going away for a long time.

I think that I am entitled to more respect than the average person is.

Sometimes I like to just watch the wind as it blows through the trees.

When working, I sometimes have difficulties due to being disorganized.

I find it hard to fully forgive someone who has done something mean to me.

I sometimes feel that I am a worthless person.

Even in an emergency I wouldn't feel like panicking.

I wouldn't pretend to like someone just to get that person to do favors for me.

I've never really enjoyed looking through an encyclopedia.

I do only the minimum amount of work needed to get by.

Even when people make a lot of mistakes, I rarely say anything negative.

I tend to feel quite self-conscious when speaking in front of a group of people.

I get very anxious when waiting to hear about an important decision.

I’d be tempted to use counterfeit money, if I were sure I could get away with it.
85. I don't think of myself as the artistic or creative type.

86. People often call me a perfectionist.

87. I find it hard to compromise with people when I really think I’m right.

88. The first thing that I always do in a new place is to make friends.

89. I rarely discuss my problems with other people.

90. I would get a lot of pleasure from owning expensive luxury goods.

91. I find it boring to discuss philosophy.

92. I prefer to do whatever comes to mind, rather than stick to a plan.

93. I find it hard to keep my temper when people insult me.

94. Most people are more upbeat and dynamic than I generally am.

95. I remain unemotional even in situations where most people get very sentimental.

96. I want people to know that I am an important person of high status.

97. I have sympathy for people who are less fortunate than I am.

98. I try to give generously to those in need.

99. It wouldn’t bother me to harm someone I didn’t like.

100. People see me as a hard-hearted person.
Appendix C

Honesty-Humility (HH) SME Rating Sheet

Please follow these instructions for rating the interviewees.

You will rate each interviewee on a rating scale. You will rate their overall HH level, as well as their level of each HH facet (Sincerity, Fairness, Greed Avoidance and Modesty). In addition to this you will specify any cues you noticed that led you to select your specific rating for each facet. If you have any cues that you feel are important but do not fit within a specific facet, please look at the training document again. If they still do not seem to fit, but you still see them are important after reviewing the document, please list these cues but not the corresponding facet. The participant number of each interviewee must also be recorded so it can be correlated with the self-report of that interviewee. The participant number is indicated in the video title. All of the ratings will be recorded in a google sheets document that I have created and linked you to.

The rating scales below are visual aids to represent how you will be ranking the interviewees. As specified, 0 indicates someone who is very low on the trait, 50 indicates that they would be average and 100 indicates they appear to be very high on the trait. When you are recording your rating for HH overall and for each facet in the google sheet, imagine you are putting a X on the scale below. Recording in excel sheet simply makes it easier for my data analysis later.
Below is a visual of what is included in the google sheet for extra clarity.

Your Initials_Participant Number:

**Honesty-Humility Overall Rating**

**Sincerity**
Cues:

•

Fairness

Cues:

•

Greed-Avoidance
Cues:

- Modesty

Cues:

Page Break
Account Manager for Walmart

POSITION: Onsite Account Manager

COMPANY: Walmart Canada

DEPARTMENT: Sales

LOCATION: Guelph, ON

Job Description:

Walmart Canada is currently seeking an Account Manager to join our team in the newly finished Walmart Supercentre in Guelph, Ontario. Reporting to the CEO, the successful candidate’s primary responsibility will be to establish and maintain relationships between Walmart and partner companies, ensuring sales and service targets are met.

Qualifications:

- High School Diploma / GED
- A completed Undergraduate Degree
- No previous account management experience required
Skills:

- Effective problem solving and critical thinking skills
- Ability to set goals, multi-task, meet/exceed objectives
- Strong written/verbal communication and social skills
- Excellent planning and organizational skills
- Proficient in Microsoft Office

Essential Duties and Tasks of the Position:

- Manage accounts, including contract negotiation and agreements
- Build and maintain strong customer relationships
- Manage customer requests, address complaints and concerns
- Preparation of sales reports and tracking account statuses
- Conduct monthly/quarterly customer review meetings
- Conduct research within the industry to identify new market opportunities and make recommendations to supervisors
- Work through challenges to improve the customer experience

Appendix E – Interview Questions for Phase 1
1. Tell me about a time when you had a difference of opinion or conflict with a supervisor/co-worker

_Probes_

a. How did you feel during this situation?

b. What were important things for you to consider in this situation?

c. Why did you choose your approach?

Examples of possible cues:

- Sincerity - how they interacted with their supervisor/co-worker i.e. were they open and honest about why they were upset?

- Fairness - did they come up with a compromise or did they want to simply get their way?

2. Tell me about a time when you had a difference of opinion or conflict with a supervisor/co-worker and tried to persuade them to see things your way.

_Probes_
a. How did you feel during this situation?

b. What were important things for you to consider in this situation?

c. Why did you choose your approach?

Examples of possible cues:

- Similar cues to above may be elicited, however the focus may be primarily on influencing which may lead to more negative sincerity scores

3. Tell me about a time when you saw the need and opportunity to set up a long-term relationship with other people.

Probes

a. How did you feel during this situation?

b. What were important things for you to consider in this situation?

c. Why did you choose your approach?
Examples of possible cues

- Fairness – did they try to benefit both parties? Did they try to seek benefits on just for themselves?

- Sincerity – were they sincere in interactions with those individuals or did they put on a façade to influence others to gain the benefit?

- Greed avoidant – if money is mentioned in terms of a gain

4. Tell me about a time when you saw the need and opportunity to set up a long-term relationship with other people because it would produce a personal benefit.

Probes

a. How did you feel during this situation?

b. What were important things for you to consider in this situation?

c. Why did you choose your approach?

Examples of possible cues:

- Could be similar to above, however one could get a lower score because the focus is on personal benefit, so they may answer with a focus based on that.
Modesty could come out in terms of answers to all of these questions, depending on the extent of bragging that occurs in their interview answers.

Appendix F – Honesty-Humility Information for Participants in Study Two, Phase Two

The Sincerity scale assesses a tendency to be genuine in interpersonal relations. Low scorers will flatter others or pretend to like them in order to obtain favors, whereas high scorers are unwilling to manipulate others.

Cues could include:

•

The Fairness scale assesses a tendency to avoid fraud and corruption. Low scorers are willing to gain by cheating or stealing, whereas high scorers are unwilling to take advantage of other individuals or of society at large.

Cues could include:

•

The Greed Avoidance scale assesses a tendency to be uninterested in possessing lavish wealth, luxury goods, and signs of high social status. Low scorers want to enjoy and to display wealth and privilege, whereas high scorers are not especially motivated by monetary or social-status considerations.

Cues could include:
The Modesty scale assesses a tendency to be modest and unassuming. Low scorers consider themselves as superior and as entitled to privileges that others do not have, whereas high scorers view themselves as ordinary people without any claim to special treatment.

Cues could include:

Trait X: Persons with very high scores on the Trait X scale avoid manipulating others for personal gain, feel little temptation to break rules, are uninterested in lavish wealth and luxuries, and feel no special entitlement to elevated social status. Conversely, persons with very low scores on this scale will flatter others to get what they want, are inclined to break rules for personal profit, are motivated by material gain, and feel a strong sense of self-importance.