UNDERSTANDING THE RELATIONSHIP BETWEEN INTERVIEWERS' DISPOSITIONAL REASONING AND JUDGEMENT ACCURACY OF DECEPTIVE IMPRESSION MANAGEMENT IN INTERVIEWS: DOES THE ACCURACY MEASURE MATTER?

Yoliswa Mzuzile Magangane

(MGNYOL001)

A dissertation submitted in partial fulfilment of the requirements for the award of the degree of Master of Social Science in Organisational Psychology

Faculty of Humanities

University of Cape Town

2021

Supervisor: Professor Francois S. De Kock

Compulsory Declaration:

This work has not been previously submitted in whole, or in part, for the award of any degree in this or any other university. It is my own work. Each significant contribution to, and quotation in, this dissertation from the work or works of other people has been attributed and has been cited and referenced.

Signature: *Date:*

Abstract

Deceptive impression management (DIM) refers to faking to create a positive image of oneself in an interview. The use of DIM poses a potential risk to organisations and threatens the validity of interviews. The risks of hiring an employee who does not meet the required performance standards and unnecessary turnover emphasise the importance to detect DIM in interviews. Previous research suggested that judges with higher dispositional reasoning ability are better at accurately judging DIM in interviews. However, recent research suggests that accuracy measures that distinguish between normative and distinctive elements may shed light on unique aspects of accuracy. For example, *normative profile accuracy* refers to the degree to which a judge can consistently judge a target in line with the expected trait profile average, whereas *distinctive profile accuracy* refers to the extent to which a judge can differentiate trait levels across targets. This secondary research study sought to understand the relationship between raters' dispositional reasoning and their DIM detection accuracy, operationalised as both normative *and* distinctive accuracy. To this end, primary data from a previous study were re-analysed, but using fresh operationalisations of accuracy that distinguished between normative and distinctive elements. The primary study, conducted in a sample of South African university students (N = 516), required students to rate the levels of DIM in interview transcripts written to depict different levels and types of impression management. Results show that judges who were higher in dispositional reasoning ability were able to accurately judge DIM in interviews, irrespective of how accuracy was operationalised. As criterion validity was consistent across normative and distinctive profile accuracy measures, the results of the study suggest the choice of accuracy measure in the study was not a study artefact.

Keywords: deceptive impression management, dispositional reasoning, individual differences, normative profile accuracy, distinctive profile accuracy, interviews.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Judgement accuracy has been a topic of interest in Human Resources Management (HRM) for several decades, dating back as far as 1933 (Hall et al., 2018). This topic is of particular importance to organisations for selection, training evaluation and appraisal purposes (De Kock, 2015; De Kock, et al, 2018) as many of these processes rely on ratings as the basis for decision-making. The level of judgement accuracy applied during selection or appraisal may ultimately have implications for the selection or promotion decisions made by an organisation and could mean a hire that will bring either value or mediocrity to an organisation (Keller & Meaney, 2017). Given that part of an organisation's competitive edge lies within the talent or employees that an organisation brings on board, accurate ratings in selection and appraisal processes are critical. It is therefore important that those providing human resource management ratings are 'good judges'.

1.1 What Makes a Good Judge: Individual Differences and Cognitive Process?

Given the crucial role of judgement accuracy for organisations, the understanding of what makes an individual judge better than another is of great importance in HRM. Individual differences or what makes a 'good judge' have been of particular interest within this field, as illustrated by the various studies previously conducted (Christiansen, et al., 2005; De Kock et al., 2015; Funder, 1995; Roulin et al., 2015). The Model of Individual Differences in Judgement and Rating Accuracy (De Kock et al., 2020) depicts examples of such individual difference constructs that are of particular interest to this study; these include dispositional reasoning and personality. *Dispositional reasoning* refers to the knowledge of traits, behaviours and how situations influence the expression of traits (Christiansen et al., 2005; De Kock et al., 2020). It comprises three components, namely trait induction, trait extrapolation and trait contextualisation (De Kock et al., 2015). *Trait induction* refers to the knowledge of how traits are represented in behaviours; *trait*

extrapolation refers to the understanding of how traits and their manifest behaviours naturally co-exist; while *trait contextualisation* refers to the knowledge of how certain situations relate to certain traits and their manifest behaviours. *Personality* represents a collection of dispositional characteristics or traits considered to be stable (Funder & Colvin, 1997). The makeup of this collection of characteristics may differ from person to person, differentiated by the levels of individual characteristics within the collection (Goldberg, 1992). Individual difference constructs are important to understand because they are considered predictors of the social cognitive process involved in making an accurate judgement (De Kock, et al., 2020).

A judgement results from a four-step interdependent process explained by the Realistic Accuracy Model (RAM; De Kock et al., 2020; Funder, 1995; Funder, 1995 2012; Letzring, 2008). This process involves a (1) *relevant* behaviour emitted by a target, such as an interviewee, in a way that forms (2) *available* information to the observer. The judge must then (3) *detect* and (4) *utilise* this information to form a view about a candidate interview dimension (De Kock, et al., 2020). Understanding how individual difference constructs relate to judgement accuracy may assist organisations to select judges who are more likely to detect the required levels of important target dimensions during recruitment and selection processes. These target dimensions may include both job-relevant dimensions (e.g., verbal communication) as well as other dimensions which may speak to aspects of applicant behaviour.

One such dimension of applicant behaviour which is receiving increasing research attention is deceptive impression management (DIM), which refers to faking or the conscious use of dishonest impression management (IM) tactics in an effort to make a good impression (Kristof-Brown, et al., 2002; Lievens & Peeters, 2008). Honest impression management (HIM) refers to an applicant's effort to create a positive image of themselves in interviews (Roulin et al., 2015). An

example of DIM occurs when a candidate exaggerates existing positive job-related competencies and past outcomes or claims to have competencies that he/she actually does not possess (Melchers et al., 2020; Roulin, 2016; Roulin et al., 2014; Roulin et al., 2015; Roulin & Powell, 2018). The prevalence of DIM use, which research indicates as high, poses a threat to the validity of interviews as organisations may make poor selection decisions based on inaccurate interview ratings (Roulin et al., 2015).

Previous research indicates a stronger and weaker relationship between (IM) and interviewer ratings and job performance respectively, suggesting that in cases where less suitable applicants are better at IM use than those more qualified, organisations are at risk of making poorer hiring decisions (Kristoff-Brown, et al., 2002; Roulin et al., 2015; Roulin, 2016). This is worse so in cases of use of DIM as a deceptive candidate is more likely to exhibit undesirable, counterproductive behaviours such as poor work performance, absenteeism, or theft (Roulin et al., 2015; Roulin, 2016). Organisations may therefore suffer monetary and non-monetary costs from a poor hire given impacts on the direct team as well as potential requirement to replace such a hire (Roulin, 2016; Keller & Meaney, 2017). The ability of judges to detect DIM use in interviews is thus of particular interest to the current study.

1.2 Judges' Ability to detect DIM

The ability of interviewers to 'spot faking' in interviews has recently been receiving increased research attention. There is disagreement in the field regarding the individual differences of a good judge that may predict an interviewers' accuracy to judge DIM (Roulin, 2016). To assess individual differences, a study by Roulin et al. (2015) made use of five experimental studies in which real-time video coding of IM was viewed by both professional and novice interviewers. Their results suggested that there are no individual differences between judges in their ability to

accurately judge IM. They also found that interviewers were most often unsuccessful when attempting to detect candidate IM and that it was easier for interviewers to detect HIM than deceptive IM.

In a later study, Roulin (2016) assessed the degree to which an individual interviewer's constructs of cognitions and social sensitivity predicted IM detection. His findings suggested that these individual differences were not related to IM detection. While Roulin (2016) and his colleagues (Roulin et al., 2015) have found no individual differences in interviewers' ability to accurately judge the use of IM, findings from other research suggest that there are interviewer constructs that may play a role in increasing the ability to judge accurately in interviews.

One such construct relates to a judge's personality. It is argued that the possession of certain personality traits by a judge may be related to a judge being better able to recognise behavioural characteristics as manifestations of certain traits in targets and to use such information to make judgements about them. A meta-analysis conducted by De Kock, et al. (2020) reviewed over 80 years of findings regarding the relationship between individual differences and judgement accuracy. Some findings include a positive relationship between judgement accuracy and the personality trait of openness to experience ($\bar{r}$ = 0.10) and agreeableness ($\bar{r}$ = 0.09). Another individual construct that has shown a stronger relationship than the abovementioned personality traits with interview judgement accuracy is dispositional reasoning, defined above as the knowledge of traits, behaviours and the influence of situations on the expression of traits. A small positive relationship was found between dispositional reasoning and judgement accuracy ($\bar{r}$ = 0.31; De Kock et al., 2020).

In addition to earlier findings that dispositional reasoning predicts interviewers' ability to judge personality traits (Christiansen et al., 2005), more recent research also showed that the

components of dispositional reasoning may explain interviewers' ability to judge interview dimensions (De Kock et al., 2015). The research was conducted through a study that made use of 146 managers (ranging from lower-level supervisor to senior manager) who rated interviewees in video-recorded interview segments on interview dimensions. The external validity of the study was ensured by inviting participants who were employed (at least five years' experience) and had managerial experience. Contrary to the abovementioned findings that there is no evidence to suggest that there are individual differences in judgement accuracy, these results suggest that such a relationship exists, and that dispositional reasoning may help to explain why interviewers may differ in their ability to DIM (De Kock et al., 2015).

To address this question, Pieterse (2016) conducted a laboratory study comprised of a sample of 676 university students from the University of Cape Town. The study participants were each tasked to rate three interview candidates depicted through vignettes on the two competencies, namely, Organisational Skill and Perseverance, as well as the level of DIM used by an interviewee. Accuracy scores were derived for each participant by computing within-person profile correlations between a participant's score and a true score (established by the researcher) of the target traits (interview competencies and DIM); an overall score was computed by averaging correlations across dimensions using Fisher's r-to-z transformed scores (De Kock, 2015). The dispositional reasoning and personality traits of each participant was assessed for each participant using a shortened version of the Revised Interpersonal Judgement Inventory (RIJI) and Big Five Inventory (BFI) respectively. The study found a positive relationship ($r = .10$; $p < .05$) between participants' dispositional reasoning and the judgement accuracy of applicant DIM. A positive relationship was also found for one of the dispositional reasoning components, namely trait induction ($r = .12$; $p < .05$). Only a small significant relationship was found between conscientiousness and overall and

performance dimension accuracy respectively (r = .10, p < .05; = .12, p < .01). The above findings suggest that a judge who has higher dispositional reasoning ability is better able to accurately judge DIM in interviews.

Although the study by Pieterse (2016) was useful in enhancing understanding about individual differences constructs and judgement accuracy in interviews, it must be noted that the research made use of profile accuracy to evaluate the degree to which judges were accurate in rating candidates on DIM and other interview dimensions. In other words, in order to measure accuracy, the study made use of only one of the available types of accuracy measures to compute accuracy scores, namely profile accuracy (Sulsky & Balzer, 1988). Profile accuracy is one of two dominant approaches used for measuring judgement accuracy and can be divided into normative and distinctive accuracy (Hall et al., 2018). The former relates to the degree to which a judge can consistently judge a target in line with the expected profile trait average; and the latter refers to the extent to which a judge can differentiate trait levels across targets (Hall et al., 2018). Profile accuracy scores were operationalised as Borman's Differential Accuracy (DA) and derived through correlating a judge's score with true scores of target traits (DIM and other interview dimensions). Borman's DA represents a type of accuracy measure that computes a score by correlating a judge's score with a true score.

The manner in which such an accuracy measure collapses potentially relevant information into a single index has been criticised as a potential shortcoming given that a single index may risk the loss of information that may be relevant for understanding different aspects of accuracy (Sulsky & Balzer, 1988). Profile normativeness is another risk posed by such an accuracy measure as it may bring about ambiguity where there is high similarity between an average profile and a judge's profile (Furr, 2008; Hall et al., 2018).

The proposed solution to the above risks of potential loss of information and profile normativeness may be to break down the correlation of a judge's accuracy scores and corresponding criterion scores into four separately measurable components know as Cronbach's Accuracy Component Scores (Cronbach, 1955; Hall et al., 2018; Kenny & Albright, 1987; Sulsky & Balzer, 1988). To elaborate on these issues, judgement accuracy measures are discussed in more detail below.

1.3 Measuring Judges' Degree of Accuracy

A variety of different methods to determine accuracy scores exist following the evolution of measurement accuracy. This evolution was driven partly through the identification of the limitations of earlier accuracy measurement techniques (e.g., a global single discrepancy score to represent accuracy) that could adversely impact the interpretation of results (Cronbach, 1955; Furr, 2008; Hall et al., 2018; Kenny & Albright, 1987; Sulsky & Balzer, 1988). An example of this evolution was seen in the development of Cronbach's Accuracy Component Scores in response to the identified shortcomings of accuracy measures such as Borman's DA.

Borman's DA represents a single index that correlates a judge's ratings for each dimension with corresponding true scores across targets which yields a single score per dimension. A higher score indicates greater accuracy (De Kock, 2015; Sulsky & Balzer, 1988). On the contrary, a lower score indicates higher accuracy in the case of Cronbach's accuracy component scores (De Kock et al., 2015; Powell, 2007). These component scores are referred to as Elevation (E), Differential Elevation (DE), Stereotype Accuracy (SA), and Differential Accuracy (DA). A detailed definition of each score is discussed in section 2.2. Cronbach's accuracy component scores break down a judge's score into four components which represent different strategies for decomposing the

judge's rating and true score discrepancy (Roch et al., 2015). In this way, Cronbach's framework yields a result that preserves information that may relate to different aspects of aspects of accuracy.

While different accuracy measures all seek to measure a judge's degree of accuracy, previous research has found that different accuracy measures correlate slightly with each other ($\bar{r}$ = .19; Hall et al., 2018; Sulsky & Balzer, 1988). These earlier studies agree that low correlations suggest that different operational definitions of accuracy have different conceptual bases and may therefore measure a different aspect of accuracy. This is demonstrated in the findings by Sulsky and Balzer (1988) who suggest the following:

> Although all of the operational definitions of performance rating accuracy include a comparison of the rater's ratings with the true scores of the ratee performance, a variety of different methods exist for comparing the two sets of ratings. A review of these methods clearly shows, however, that they do not all share a common conceptual base. (p. 498)

In the case of Cronbach's four accuracy component measures, the distance between a judge's score and the true score is fragmented into distinct accuracy component scores. Both a correlational and a variance component score can be derived for the latter three component measures (DE, SA, and DA). Given that conceptually, the psychometric definition of judgement accuracy concerns the distance between judge and true scores, the correlational components of these measures do not align to this definition and therefore rest on a different conceptual base (Sulsky & Balzer, 1988). Similarly, Borman's DA stems from a different conceptual base to Cronbach's variance component scores that focus on distance between judge and true scores. A possible implication of measures with different conceptual bases may be that the same study may produce different accuracy results depending on the accuracy measure used (Sulsky & Balzer,

1988). This brings into question whether study findings can generalise across different accuracy measures such as distinctive and normative profile accuracy used in the study by Pieterse's (2016).

Pieterse's (2016) research was important in establishing the relationship between dispositional reasoning and the normative profile accuracy of DIM in interviews; however, it is also important to explore the question regarding the relationship between dispositional reasoning and the distinctive profile accuracy of DIM in interviews in order to see the type of results produced by an alternative accuracy measure. This will also contribute towards understanding of whether or not accuracy scores can be influenced by more than just individual difference constructs (Sulsky & Balzer, 1988). By so doing, further insights will be gained on the individual interview constructs indicated in the model by De Kock et al. (2020), particularly in the context of DIM judgement accuracy. To the best of the researcher's knowledge, there is limited research on the comparison of results of different accuracy measures that assess the relationship between dispositional reasoning and judgement accuracy of DIM (Roulin et al., 2015; Sulsky & Balzer, 1988). To investigate the above, this study thus seeks to answer the question:

What is the relationship between dispositional reasoning and judgement accuracy of DIM when using both normative and distinctive judgement accuracy measures?

1.4 Study Objectives

Judgement accuracy of DIM in interviews is crucial for the selection process, not only to ensure that selected candidates are those who meet the required standard of interview dimensions or competencies to ensure adequate future performance in the job, but also for safeguarding against the risks posed by DIM in interviews (Kristoff-Brown, et al., 2002; Roulin et al., 2015; Roulin,

2016). The understanding of judges' ability to detect DIM and the predictors related to DIM judgement quality are thus important. Of equal importance is understanding whether the DIM judgement accuracy results are generalisable across different accuracy measures or if measures may be assessing a different aspect of judgement accuracy. The purpose of this research is to investigate whether different accuracy measures yield different study findings. It particularly aims to empirically investigate the relationship between dispositional reasoning and DIM detection in interviews when different accuracy measures are used.

1.5 Study Overview

The present study sets out to answer the research question posed firstly, through a review of relevant literature. This includes the theoretical background relating to some of the individual interviewer constructs considered to contribute to what makes a good judge, as well as the cognitive process that underlies judgement accuracy. An overview of DIM use and judges' ability to detect DIM in interviews is then discussed before moving into the review regarding different judgment accuracy measures and how they relate to each other. The study then conceptually distinguishes the individual difference constructs discussed above before leading into the study hypotheses. The methods chapter describes the research design, sample, stimulus material construction, measurement scales, data collection and statistical analyses. The results chapter details the psychometric properties of measurement scales used as well as descriptive statistics, and hypotheses results. The final chapter discusses the study findings and their theoretical and practical implications to the field of HRM and organisations respectively. The study is brought to a close by concluding remarks.

Chapter 2: Literature Review

This chapter begins with a discussion of the judgement accuracy of DIM in interviews and the process that enables an interviewer to detect DIM during interviews. It is followed by a discussion of the individual interviewer constructs that predict DIM judgement accuracy, namely dispositional reasoning and personality. A discussion of judgement accuracy then follows that looks at conceptual and operational definitions, as well as relationships between different accuracy measures. The first and second study hypotheses are presented within the abovementioned discussion. The chapter concludes with a discussion of the incremental validity of dispositional reasoning followed by the third proposed study hypothesis.

2.1 Deceptive Impression Management Detection in Interviews and Interviewer Differences that Predict Deceptive Impression Management Judgement Accuracy

2.1.1 Deceptive Impression Management in Interviews

DIM refers to the effort of a candidate to paint him or herself in a positive light using dishonest means (Roulin et al., 2014). DIM tactics can be divided into three categories each comprising certain techniques (Melchers et al., 2020). *Deceptive ingratiation* involves falsely expressing beliefs or values held by the interviewer or company as one's own. Assertive techniques comprise *slight image creation,* which refers to over-exaggerating one's competencies and enhancing experiences. *Extensive image creation* refers to the inventing or borrowing of experiences and achievements. Defensive tactics include *image protection,* which includes omission of unfavourable information (Roulin et al., 2014, 2015).

There is evidence to suggest that honest IM (HIM) is made use of by candidates during interviews (Kristoff-Brown et al., 2002; Lievens & Peeters, 2008). The use of IM is positively or negatively related to the perceptions of interviewers (Kristoff-Brown et al., 2008; Roulin et al.,

2014, 2015). Kristoff-Brown et al. (2008) found that different types of IM tactics were positively related to interviewer perceptions of job-fit and interviewer-candidate similarity and would relate to better interview ratings.

In a different study, Roulin et al. (2015) found that perceptions of DIM were related to negative interviewer ratings, while perceptions of HIM were related to positive interviewer ratings. The authors did however find that the detection of DIM by interviewers was low. It can be argued that interviewers could have given higher ratings to candidates who made use of DIM because they did not detect it as they perceived the behaviour to be HIM.

DIM could put an organisation at risk of appointing an incumbent who is not well-suited or underqualified and who will be unable to meet the performance requirements of a job (Roulin et al., 2014, 2015; Schneider, Powell, & Roulin, 2015). It also places an organisation at risk of wasted opportunities where suitable candidates are overlooked as a result of the influence of DIM by less suitable candidates. These missed opportunities could cost an organisation in terms of time and money as they may need to go through another interview process after a certain period when they need to let go of a candidate due to poor performance and search for a more suitable candidate. Another risk to an organisation is the time spent addressing the counterproductive behaviours that may be displayed by a deceptive candidate (Roulin, 2015; Roulin et al., 2016). The inability to detect DIM could therefore render the interview an inefficient personnel selection tool and thus affect the validity of the tool.

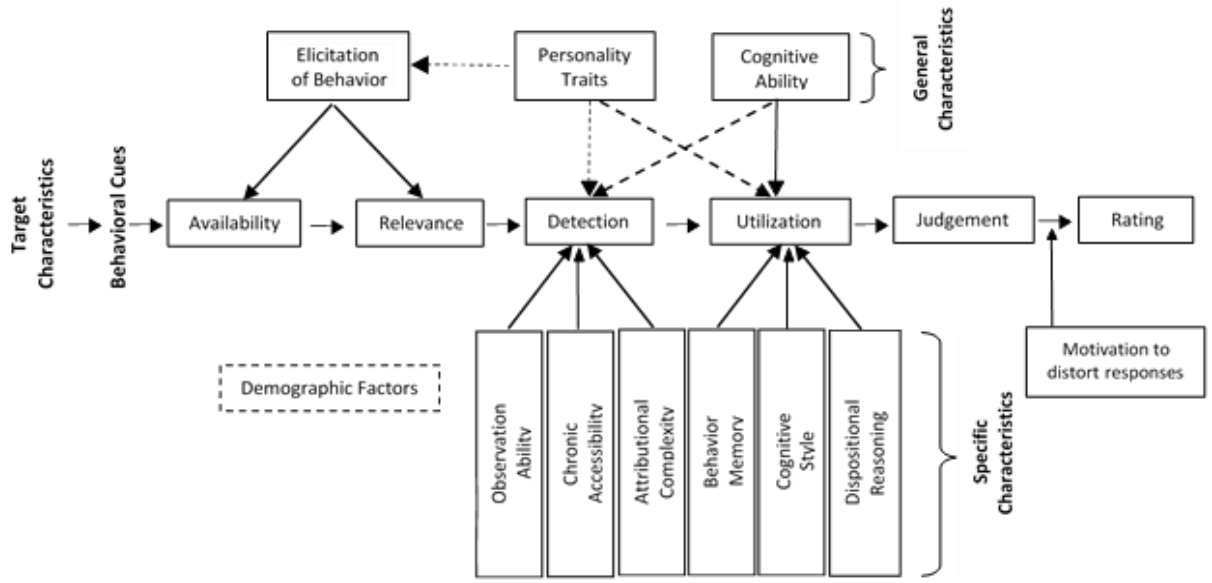
2.1.2 How Deceptive Impression Management is Detected in Interviews

The model established by De Kock et al. (2020) through their review of the individual differences that make a good judge, provides a framework that explains both how an interviewer makes judgements (cognitive processes involved) as well as the individual interviewer constructs

that aid key judgement accuracy processes, namely cue detection and utilisation (see Figure 1; De Kock et al., 2020).

Figure 1

A Model of Individual Differences in Judgement and Rating Accuracy



The model incorporates Funder's Realistic Accuracy Model (RAM; 1995, 2012) as well as the context model of judgement and rating by Murphy and Cleveland (1995). This model will be used to explain the detection of DIM in interviews as well as the individual differences that may explain why some judges are better than others at detecting DIM in interviews. The model identifies judgements as the private evaluations made by a judge and ratings as the public statements made about a target's performance; it thus distinguishes between a good judge and a good rater.

An introductory feature of the model is the moderator between judgements and ratings, termed motivation to distort, which explains what makes an accurate judge versus an accurate rater. The motivation to distort is explained by factors such as the perceived consequences of

ratings, political pressures and personal goals to advance interests. While the model emphasises the distinction between judgement and ratings in performance appraisal contexts more than selection contexts, it is thought appropriate for explaining judgement accuracy in interviews.

Detection of DIM in interviews can be explained by the RAM (De Kock, 2015; Funder, 1995, 2012; Letzring, 2008). This model represents a framework that facilitates judgement accuracy through the completion of four interdependent steps. In the first step, a candidate's use of DIM tactics provides behavioural cues that are *relevant* to the trait being judged. Second, the behavioural cues must be displayed in such a way that they are *available* to the judge or interviewer; in other words, they must occur in the presence of the interviewer. The judge or interviewer will have the opportunity to *detect* or identify the cues, thus completing the third stage of the model. Lastly, once detected, the interviewer must then *utilise* this information effectively to form a judgement about the use of DIM by a candidate (De Kock, 2015; Funder, 1995, 2012; Letzring, 2008).

The third and fourth stages of RAM imply that a judge would need to know a behaviour to be able to identify it and to also have knowledge about how behaviours are influenced to make a judgement about detected behaviour within a certain context. This proposition can be explained by individual interviewer constructs which are proposed to predict judgement accuracy (Christiansen et al., 2005; De Kock et al., 2015; De Kock et al., 2020).

2.1.3 Individual Differences: Interviewer Constructs

The abovementioned model encompasses the judgement accuracy moderator of the 'good judge' and suggests that some judges are better than others at judgement accuracy (Funder, 2012). It is proposed that some individual differences or constructs differentiate judges on their ability to accurately judge targets (Christiansen et al., 2012; De Kock et al., 2015; Letzring, 2008).

Previous research suggests that some judges are better than others at judging personality and interview dimensions due to certain personality traits or higher cognitive ability or a combination of both (Christiansen et al., 2012; De Kock et al., 2015; Letzring, 2008). Findings by Letzring (2008) suggest that judges who are agreeable were more accurate at judging personality. The study by Christiansen et al. (2005) found that judges who had the personality trait of openness and higher general mental ability (GMA) made more accurate judgements through the traits' influence on dispositional intelligence or reasoning.

Findings by De Kock et al. (2015) show that individual differences in dispositional reasoning, which were found to meet the criteria of intelligence, explained differences in interviewer accuracy (De Kock et al., 2015). Previous findings illustrate that dispositional reasoning and personality may be predictors of judgement accuracy and may thus play a role in judgement accuracy in interviews. These interviewer constructs form part of the *specific characteristics* in the model compiled by De Kock et al. (2020). They are discussed in more detail below.

2.1.3.1 Personality. Personality represents dispositional characteristics considered to be stable (Funder & Colvin, 1997). It is thought to regulate behaviour in the workplace as well as aspects of interpersonal judgement and rating quality (Tziner, et al., 2008). This notion of rating quality or accuracy can be explained through the suggested conceptual link between certain personality traits and the information processing stages – detection and utilisation – of RAM (De Kock et al., 2020).

These characteristics facilitate better knowledge and understanding of traits and behaviours, thereby contributing to the key judgement processes of RAM (De Kock et al., 2020). For example, the trait of openness is characterised by inquiry, imagination and seeking an

understanding of abstract concepts and theories (Christiansen et al., 2005; Goldberg, 1992). Thus, openness is considered to be related to better judgement accuracy of observed behaviours. Agreeable individuals generally show concern for others' feelings and it is thought that this sympathetic characteristic encourages such people to be more socially and, therefore, behaviourally aware of those around them. Knowledge of behaviour is increased in this way and facilitates better judgement accuracy (De Kock et al., 2020).

Extraverted individuals tend to seek out social interactions and it is thought that social exposure provides practice and feedback to refine their interpersonal judgements. Conscientiousness is manifested through detail orientation, making such individuals likely to be more attentive and consistent in the stages of cue detection and utilisation, respectively (De Kock et al., 2020). The above conceptual arguments can be grouped into two categories.

The first is proximal in nature, such that personality may enhance accuracy by making people more inclined to enjoy studying others socially or value understanding others' intentions. The second category represents a distal explanation which suggests that personality may influence preferences for social interactions and the importance attached to interpersonal skills; these factors provide the opportunity to develop accuracy faster (De Kock, 2015).

The inconsistent nature of the findings of prior studies is indicative of the complexity of understanding how personality relates to judgement accuracy. While the empirical support reflecting this relationship is relatively low, certain studies have found a correlation between certain personality traits and accuracy (De Kock et al., 2020; Pieterse, 2016). A significant relationship was found between conscientiousness and judgement accuracy of interview performance dimensions as well as overall accuracy in the study by Pieterse (2016). As illustrated in the meta-analytic findings by De Kock and his colleagues (De Kock et al., 2020), only openness

($\bar{r}$ = 0.10) and agreeableness ($\bar{r}$ = 0.09) showed to predict rating quality. Both individual constructs showed a 90% CI that did not include zero.

2.1.3.2 Dispositional reasoning. As mentioned earlier, dispositional reasoning refers to the knowledge of traits, behaviours and how situations influence the expression of traits. It consists of three components. First, *trait induction,* which refers to the knowledge of how traits are represented in behaviours; second, *trait extrapolation* which refers to the understanding of how traits and their manifest behaviours naturally co-exist; and third, *trait contextualisation* which refers to the knowledge of how certain situations relate to certain traits and their manifest behaviours (De Kock et al., 2015, 2017).

As discussed above, the last two stages of RAM require a judge to *detect* behavioural cues relevant to the criterion being judged and *utilise* information received about behavioural cues in a specific context to make a judgement about the witnessed behaviour against the respective criterion. In the same way, dispositional reasoning utilises behavioural cues detected in a particular context to make a judgement about the witnessed behaviour based on the knowledge of how certain traits, behaviours and situations influence the expression of certain manifest behaviours (Christiansen et al., 2005; De Kock et al., 2015; Funder, 1995).

Dispositional reasoning shows the ability to predict the judgement accuracy of personality and interview dimensions (Christiansen et al., 2005; De Kock et al., 2015). A study by Christiansen et al. (2005) found a moderate relationship (Cohen, 1988) between dispositional reasoning and interview accuracy (r = .42, p < .01) as well as acquaintance accuracy (r = .41). Findings by De Kock et al. (2015) showed a moderate relationship between interview accuracy and dispositional reasoning components for which significant correlations were found, namely, trait extrapolation (r = .33, p < .001) and trait contextualisation (r = .26; p = .002). In the judgement of both

personality and interview dimensions, the judge makes use of information regarding identified behavioural cues and appropriately evaluates the information to form a judgement.

The DIM tactics discussed above can manifest through verbal and non-verbal expressions made by a candidate (Roulin et al., 2014, 2015; Schneider et al., 2015). The candidate can make use of words to make false claims about possessing values held by the interviewer or to relay past experiences that are exaggerated or newly created (Roulin et al., 2014, 2015). The above illustrates the behavioural cues of DIM tactics and provides information that can be used to make a judgement. For example, when a candidate says "I share your sentiments on the issue of gender equality" to ingratiate him or herself, or "In my previous role I single-handedly saved the company large amounts of money" to exaggerate a past experience, interviewers who can detect these DIM behavioural cues would be better able to distinguish DIM cues from job-relevant cues that illustrate the potential of performance in respective interview dimensions, e.g., organisational skills.

In contrast, interviewers who are unable to detect DIM behavioural cues would be unable to distinguish them and may see an interviewer-candidate's similarity in the case of ingratiation or may feel that a candidate can exhibit job-related cues, therefore creating a positive perception of a candidate and inaccurately awarding a candidate higher ratings for a specific interview dimension.

If dispositional reasoning enables a judge to make judgements through the detection and utilisation of behavioural cues, it can be argued that a judge would be able to judge the use of DIM by candidates through the detection and effective utilisation of behavioural cues related to DIM. Findings by Pieterse (2016) illustrate a positive relationship between dispositional reasoning and the judgement accuracy of DIM. In addition, the dispositional reasoning components of trait induction and trait extrapolation were found to be positively related to the judgement accuracy of

DIM, performance dimensions and overall accuracy. The present study will focus on overall dispositional reasoning in relation to accuracy.

2.2 Defining Judgement Accuracy

2.2.1 Conceptual Definitions

Judgement accuracy refers to the degree to which a judge can correctly deduce the traits or characteristics of a target based on information gathered through certain processes, for example, interviews (De Kock, 2015; Sulsky & Balzer, 1988). The variety of perspectives of judgement accuracy that exist is illustrated by the different terms used to describe it. Some examples of these terms include rating accuracy (Sulsky & Balzer, 1988), inferential accuracy (Jackson, 1972), realistic accuracy (Funder, 1995) and interpersonal sensitivity (Hall et al., 2009).

The types of judgements made by an assessor can be divided into the categories of *predictive* and *diagnostic* judgements. The former refers to inferences or predictions made from measures about a target's (e.g., candidate's) future behaviour, while the latter refers to inferences made through assessment of the degree to which a target possesses certain traits (De Kock, 2015).

Diagnostic judgements can be further divided by depth into *behavioural* (shallow), which refers to the judgement of the occurrence of a behaviour (or lack thereof) and involves the observation of behaviours, and *dispositional* (deep), which refers to the judgement of an underlying trait that may explain the exhibition of certain behaviour and involves the evaluation of behaviours.

The focus of this study is diagnostic judgements in the personnel selection field. The accuracy of dispositional inferences is known as *inferential accuracy* and describes a judge's ability to accurately infer the qualities or traits of a target based on the assessment of behavioural information collected from selection tools, e.g., interviews. An important factor regarding

accuracy is that it is not the opposite of rating error, as illustrated by the low correlations (r = - 0.5) found between the indices of these two variables in Murphy and Balzar's (1989) meta-analytic study.

The measurement of judgement accuracy is facilitated by comparing a judge's score with a corresponding measure considered to be the accepted comparison standard, termed a true score. In order to qualify as an accuracy measure, an index of accuracy must align to the psychometric conceptual definition of accuracy, which concerns the distance between a judge's rating and true score (Sulsky & Balzer). While some accuracy measures may have this definition in common, they may differ in operationalisation as discussed below.

2.2.2 Operational Definitions

Accuracy measures indicate the strength and kind of relation between one set of measures and a corresponding set of true scores (Sulsky & Balzer, 1988). Various operationalisations of accuracy have been established for the measurement of accuracy as illustrated by the number of different accuracy measures used in previous research (Hall et al, 2018; Murphy & Balzer, 1989; Sulsky & Balzer, 1988). These accuracy measures include, for instance, Borman's DA, Cronbach's Component Accuracy Scores, Distance Accuracy, Leniency Measures, Halo-Type Accuracy and Trait Accuracy.

Early developments of accuracy measures involved different scores or indices of correlation between a judge's target trait judgements and the target's actual profile (De Kock, 2015). It was later argued that collapsing potentially relevant information into a single index and not taking the direction of deviations from true scores into account is a shortcoming of the earlier measures of accuracy (De Kock, 2015; Kenny & Albright, 1987). This potentially risks the loss of

information that may be relevant for understanding different aspects of accuracy (Sulsky & Balzer, 1988).

One proposed solution to this was to break down the correlation of a judge's accuracy scores and corresponding criterion scores into four separately measurable components known as Cronbach Accuracy Component Scores (Cronbach, 1955; Hall et al., 2018; Kenny & Albright, 1987; Sulsky & Balzer, 1988). Different accuracy measures may support varying approaches to determining judgement accuracy. These approaches include trait and profile accuracy.

2.2.2.1 Trait and Profile Accuracy. There are two dominant approaches to assessing judgement accuracy, namely *trait accuracy* and *profile accuracy*. The former (also known as the trait-centred approach) describes the degree to which a judge can accurately rank targets on a given trait or interview dimension (Allik et al., 2015; Hall et al., 2018; Krzyzaniak et al., 2019). The latter (also known as the person-centred approach) describes a judge's ability to accurately distinguish between relative levels of traits or interview dimensions within a single target (Allik et al., 2015; Hall et al., 2018; Krzyzaniak et al., 2019).

An example of where trait accuracy may be relevant is during a job interview of a salesperson or accountant. A recruiter's ability to accurately judge and compare applicants on extraversion or conscientiousness, respectively, may better ensure a successful hire (Hall et al., 2018). Trait accuracy is computed across targets for a specific trait or dimension by comparing a judge's rating with a criterion rating. A judge's rating accuracy of a trait is determined through item-level correlations of the abovementioned set of ratings (Allik et al., 2015; Krzyzaniak et al., 2019). The trait accuracy score allows a researcher to determine the extent to which a judge can evaluate which interview applicant holds a more or less relevant interview performance dimension (e.g., organisational skills).

Profile accuracy assesses a judge's ability to detect the relative levels of traits for a single target (Allik et al., 2015; Hall et al., 2018; Krzyzaniak et al., 2019). For example, a judge may discern that a candidate has a higher level of the competency organisational skills than perseverance. The profile accuracy method which is centred around the degree of similarity of a judge's profile to a criterion profile (rank ordering of a set of traits) is computed by comparing a judge's assessment of a candidate profile to the corresponding criterion profile (Allik et al., 2015; Hall et al., 2018).

Larger profile accuracy correlations are indicative of a higher similarity between profiles (Allik et al., 2015). The profile accuracy approach can provide researchers with two possible insights: first, the extent to which a judge can assess whether a target is higher on one interview dimension than another, for example, whether a candidate shows to possess more organisational skills than perseverance; and second, which judges are better at judging the profile of interview candidates (Hall et al., 2018; Krzyzaniak et al., 2019).

Profile-centred accuracy may prove useful in the context of selection as it allows for an understanding of how an individual's traits fit together, thereby providing insights into the strengths and development areas of a job applicant. This will allow a recruiter to assess the degree to which an individual meets the requirements of a job, e.g., a job may require an employee higher in analytical skill than interpersonal skill for a specialist role (Hall et al., 2018). Profile accuracy can be divided into two elements, namely, normative and distinctive, discussed below (Allik et al., 2015; Biesanz & Human, 2010; Furr, 2008; Hall et al., 2018; Krzyzaniak et al., 2019).

2.2.2.1.1 Normative and distinctive profile accuracy. Profile normativeness refers to the extent to which a target profile (an individual's profile of interview dimensions) displays an average profile (a group's profile of normative trait ratings). It thus refers to the degree to which a

judge can consistently judge a target in line with the expected trait profile average. The accuracy measure is computed by correlating a judge's average profile with the corresponding normative average target profile (Hall et al., 2018). A higher score is indicative of a good level of knowledge about the average trait profile of others, for example, knowing that the average person has a higher level of a particular trait (extraversion) than another (neuroticism; Furr, 2008; Hall et al., 2018). The present study has implemented operationalised normative profile accuracy using Borman's DA in order to obtain normative profile accuracy scores.

2.2.2.1.1.2 Borman's Differential Accuracy. Borman's DA is an index that correlates a judge's ratings for each dimension with corresponding true scores across targets which yields a DA score for each dimension. An overall score is obtained by averaging the correlations across dimensions using Fisher's r-to-z transformed scores (De Kock, 2015; Sulsky & Balzer, 1988). A higher score indicates greater accuracy. The mathematical formula for Borman's DA is illustrated in Table 1.

Table 1

Mathematical Formula of Borman's Differential Accuracy Measure

Accuracy Measure	Formula
Borman's Differential Accuracy	DA: $DA = 1/d \sum_{j=1}^{d} (T^{*}_{rt})$

Note. d = number of dimensions; T^{*}_{rt} = correlation between ratings and true scores for a particular dimension transformed to a Z score.

A highlighted limitation of this measure is that it does not consider the distance between ratings and true scores because it only provides correlational information, making it conceptually more similar to rating validity (Sulsky & Balzer, 1994). A reason for this challenge is that ratings

may correlate highly even though 'true' ratings (the difference between the judge's rating and the expert rating) are different to a judge's rating, i.e., a judge's rating is not close to the criterion rating.

An example of how this may occur in a performance rating scenario is if a judge consistently assigns ratings two points higher than the corresponding expert rating; the ratings of these two judges may be perfectly correlated and thus this judge may receive a high 'accuracy' score, even though the judge is not providing the same ratings as the expert (Roch et al., 2012). A high correlation therefore does not necessarily reflect a small distance between a judge's rating and the corresponding true or expert rating (Smither et al., 1989).

It is proposed that Borman's DA should rather be used as an indicator of preliminary information about accuracy (De Kock, 2015; Sulsky & Balzer, 1988). Despite its limitations, Borman's DA is commonly used in research due to its ease of interpretation (Smither et al., 1989) as well as its focus on how well a judge can differentiate between targets across traits or dimensions (Roch et al., 2012). The latter was considered by Borman as the reason that makes Cronbach's DA (discussed below) the most conceptually appropriate component score of accuracy (Smither et al., 1989; Sulsky & Balzer, 1988). While there may be this similarity, the correlational nature of Borman's DA means that it is not equivalent to Cronbach's DA given that it does not take distance into consideration. Cronbach's DA is one of four component scores intended to address the limitations posed by Borman's DA. Cronbach's component accuracy scores form distinctive profile accuracy which accounts for normativeness (Furr, 2008).

2.2.2.1.3 Cronbach's Accuracy Component Scores. Profile distinctiveness refers to how an individual's profile diverges from the normative profile – the extent to which the individual profile is above or below the average profile. Distinctive profile accuracy refers to the extent to

which a judge can differentiate trait levels across targets (Allik et al., 2015; Biesanz & Human, 2010; Furr, 2008; Hall et al., 2018; Krzyzaniak et al., 2019). Distinctive profile accuracy reflects how accurately a judge can differentiate a target's profile from the normative profile (typical profile of a specific group; Hall et al., 2018).

Normativeness is mathematically accounted for by Cronbach's accuracy framework. Cronbach's framework assesses accuracy by breaking down a judge's score into four components which represent different strategies for decomposing the judge's rating and true score discrepancy (Roch et al., 2012). In addition to addressing the issues associated with normativeness, Cronbach's component scores seek to ensure that certain aspects of accuracy are not lost, a risk posed by indices that square rating deviations (Cronbach, 1955; De Kock, 2015; Sulsky & Balzer, 1988).

Sulsky and Balzer (1988) explain that these four component scores rest on different conceptual bases. Cronbach's framework is comprised of: (a) *elevation* (EL), which reflects a judge's tendency to rate too high (lenient) or too low (strict) and it indicates accuracy over both traits and targets; (b) d*ifferential elevation (DE),* which reflects a judge's ability to differentiate between or rank targets across all traits despite his or her rating tendency; (c) *stereotype accuracy (SA),* which refers to normativeness and a judge's ability to differentiate between averaged traits for a group of targets and it reflects the extent to which a judge can accurately indicate the trait levels for a group of targets; (d) *differential accuracy (DA),* which reflects a judge's ability to differentiate between relative levels of traits in a single target (Cronbach 1955; De Kock, 2015; Powell, 2007; Roch et al., 2012). The mathematical formulas for the individual component measures are illustrated in Table 2.

Table 2

Mathematical Formula of Cronbach's Accuracy Component Measures

Accuracy Components	Formula
Elevation (EL)	$E^2 = (OGM - TGM)^2$
Differential Elevation (DE)	$DE^2 = 1/n \sum_i [(OMi - OGM) - (TM_i - TGM)]^2$
Stereotype Accuracy (SE)	$SA^2 = 1/k \sum_j [(OM_j - OGM) - (TM_j - TGM)]^2$
Differential Accuracy (DA)	$DA^2 = 1/kn \sum_i \sum_j [(OM_{ij} - OM_i - OM_j + OGM) - (TM_{ij} - TM_i - TM_j + TGM)]^2$

Note. OGM = observed grand mean (mean rating over all targets and all traits); TGM = true grand mean (experts' mean rating over all targets and all traits); OM_i and TM_i = mean rating and mean true score for target i; OM_j and TM_j – mean rating and mean true score for trait j; OM_{ij} and TM_{ij} = rating and true score for target i on trait j.

In addition to Cronbach's DA being considered as corresponding most closely with the conceptual notion of accuracy (Borman, 1977; Powell, 2008), it is also the most commonly used component of Cronbach's framework (Suslky & Balzer, 1988). It is considered a good indicator of target strengths and development areas and has shown to be an effective operationalisation of distinctive accuracy, particularly in the case where a judge is required to rate multiple targets (Furr, 2008; Roch et al., 2015).

Cronbach's DA is computed by mathematically removing a judge's overall average, along with his or her average rating levels for all targets and traits (Powell, 2007). DA can be computed by normatively adjusting a profile accuracy score (Furr, 2008; Furr & Wood, 2013; Hall et al., 2018). The outcome score reflects the degree to which a judge can accurately perceive how different a target is from the average of a target group (Biesanz & Human, 2010).

In summary, previous research evaluating the good judge commonly employed profile accuracy measures (Hall et al., 2018). The ease of interpretation offered by Borman's DA and the meaningful improvement offered by Cronbach's DA to traditional profile accuracy measures, make them appropriate to investigate DIM judgement accuracy (Furr, 2008). Given the above and that, to the best of the researcher's knowledge, no further research evaluating the relationship between dispositional reasoning and DIM judgement accuracy when using different accuracy measures has been conducted, the accuracy measures discussed are appropriate for the study. The above measures will further enable the study to investigate whether previous findings (Pieterse, 2016) of interviewer individual differences and the detection of DIM can be replicated across different accuracy measure. This study thus hypothesises:

H1: Interviewers' dispositional reasoning is positively related to normative profile accuracy of DIM in interviews.

H2: Interviewers' dispositional reasoning is positively related to distinctive profile accuracy of DIM in interviews.

2.3 Incremental Validity of Dispositional Reasoning

The original study results (Pieterse, 2016) successfully illustrated that dispositional reasoning, which characterises intelligence, shares no common conceptual qualities with personality and that these two constructs are empirically unrelated (De Kock, 2015). This is reasonably expected given the behavioural and cognitive nature of the conceptual definitions of personality and dispositional reasoning, respectively.

Personality is described as the stable characteristics that relate to the pattern of expression of behaviours (Funder & Colvin, 1997). Dispositional reasoning on the other hand is concerned with the application of cognitive tools that enable information processing about traits, behaviours and situations (De Kock, 2015). Given that personality is not expected to relate to dispositional reasoning, it is expected that dispositional reasoning should increment personality in predicting accuracy. Additionally, personality and dispositional reasoning should be empirically distinct (Christiansen et al., 2005). Given that the original study demonstrated the above for normative profile accuracy, this study proposes that the same should be seen for distinctive profile accuracy. This study thus proposes:

> *H3: Interviewers' dispositional reasoning will explain additional variance in*
> *in* the *3(a) normative and 3(b) distinctive profile accuracy of DIM over interviewers' personality.*

2.4 Conclusion

This chapter discussed the variables of the present study through the presentation and review of the existing literature. These study variables include DIM, dispositional reasoning, and normative and distinctive profile accuracy. The chapter also discussed the three proposed hypotheses for investigation in the present study.

Chapter 3: Method

This chapter provides details of the research design, sampling procedure and participants of the study. It then discusses the development of the DIM stimulus material, before discussing the instruments used to measure the individual constructs of dispositional reasoning and personality. An explanation of the data collection is then provided before concluding with an outline of the statistical analysis executed for the present study.

3.1 Research Design

A secondary analysis of archival data was used for the present study. The data analysed for this study formed primary data in Pieterse's (2016) study, which made use of a fully crossed within-person research design (Aguinis & Bradley, 2014). For this chapter, the previous study will be referred to as the original study. Secondary analysis is considered appropriate for this study as it would facilitate the investigation to answer the research question that seeks to determine the relationship between dispositional reasoning and judgement accuracy of DIM when different accuracy measures used. Other benefits of this type of design include cost-effectiveness and time efficiency (Neuman, 2000).

3.2 Sampling and Participants

Non-probability sampling, specifically convenience sampling, was used in the original study to ultimately obtain a sample comprised of 676 registered University of Cape Town (UCT) students. Given that participants are not randomly selected but included as part of the study sample due to their accessibility and proximity to the researcher, there is a risk to the ability to generalise the results to the population. While this sampling technique has this challenge, it was considered appropriate for the original study as it is cost-effective and quick to implement (Battaglia, 2011;

Etikan, Musa, & Alkassim, 2016). This was important for the original study given the time constraints in which the study had to take place (a one-year Master's programme).

The sample size was obtained by implementing the criterion of 100% complete cases; thus only participants who completed all items of the online survey were included in the sample. The study made use of the survey tool Qualtrics to obtain data. As an incentive, the top three ratings were awarded a cash prize of R1250, R1000 and R750, respectively. Additionally, two lucky draws of R500 were awarded. An incentive is considered appropriate as it better ensures a good response rate and that participants make a conscious effort when completing the survey (Ryu, Couper, & Marans, 2006).

Further to the above criterion, for the present study, participants with responses deemed careless were eliminated from the sample size. The criterion used was based on duration (less than 20 minutes and more than 70 minutes). Applying the above criterion, this study eliminated $\boldsymbol{n} = 46$ responses that were below 20 minutes and $n = 114$ responses above 70 minutes to eventually get to a sample size of $N = 516$ (76% of the initial sample size based on complete ratings).

The application of the above criteria resulted in a final sample comprising a larger proportion of females (60.6%, n = 312) than males, (39.1%, n = 202), with 0.4% (n = 2) of participants identifying as 'other'. The age of participants ranged from 18 to 55 ($M = 22.35$, $SD = 4.95$). A large portion of the sample identified as Black (37.4%, n = 193). The remaining participants identified as White (30.4%, n = 157), Coloured (12%, n = 62), Indian (10.5%, n = 54), Asian (1%, n = 5), Other (1.4%, n = 7) and 7.4% (n = 38) preferred not to say. A large portion of participants (77%, n = 401) reported to have had some form of work experience. Further description of the sample demographics is provided in Table 3.

Table 3

Sample Demographics

Category	Sub-category	No. of participants	Percentage (%)
Faculty	Commerce	126	24.4
	Humanities	112	21.7
	Engineering	98	19.0
	Health Sciences	92	17.8
	Science	69	13.4
	Law	19	3.7
Year of study	First year	122	23.6
	Second year	121	23.4
	Third year	116	22.5
	Masters	62	12.0
	Fourth year	58	11.2
	Honours	37	7.2
Frequency of subjective evaluations of other people (e.g., lecturers, tutors, peers)	Twice a year	250	48.4
	Four times a year	118	22.9
	Monthly	63	12.2
	Once a year	33	6.4
	Never	31	6.0
	Weekly	19	3.7
	Daily	2	0.4
Device used to participate in study	Laptop	279	54.1
	Cell phone	141	27.3
	Desktop Computer	53	10.3
	Tablets	42	8.1

Note. $N = 516$.

3.3 Deceptive Impression Management Stimulus Material Development

The original study made use of Experimental Vignette Methodology (EVM) to develop the DIM stimulus material that would be conveyed through three interview transcripts of an entry-level management trainee position. All three transcripts were viewed by all participants. EVM is the use of short descriptions of meticulously constructed, realistic situations or people (vignettes) to elicit participant judgements, choices or decisions. This is to enable the assessment of a study's dependent variable(s), for example, attitudes, behaviours and, in the case of this study, judgement accuracy (Aguinis & Bradley, 2014). EVM not only enhances experimental realism through the construction techniques mentioned above, but it also allows for control and manipulation of independent variables, thus enhancing both external and internal validity (Aguinis & Bradley, 2014).

In the context of EVM, the chosen research design (within-person) for the original study is considered appropriate given that it allows for the assessment of manipulation effects within a participant as well as understanding his or her judgement processes (Aguinis & Bradley, 2014). In addition, the viewing of the same interview transcripts by all participants allows for comparison across respondents for the same vignette. The type of EVM used for the original study is considered a 'paper people' study given the short written descriptions of people and scenarios (Aguinis & Bradley, 2014). This EVM type assesses explicit processes and outcomes, i.e., participants' decisions or judgements regarding their observations of the 'paper people' are evaluated.

In addition to the abovementioned advantages, another benefit of the EVM type employed is that it requires lower cognitive demands and selective attention from participants as compared to other types of observations such as videotapes or live events (Bradbury-Jones et al., 2014;

Kinicki et al., 1995). Participants are better guarded against contradictory visual and non-visual cues that may occur in a videotaped vignette. For example, a participant may be willing to provide a good lecturer rating for a written description of a lecture, but less so for a video representation of the same lecturer who is dressed in clothing perceived to be unfavourable clothing (old or untidy) presenting the same lecture. The above characteristics of EVM better ensure that study results are related to an isolated stimulus (Aguinis & Bradley, 2014; Kinicki et al., 1995).

While an experimental design would offer a higher level of confidence regarding internal validity, EVM addresses the risk of lower external validity brought about by this design type through its opportunity for enhanced realism within written scenarios. In addition, EVM requires less time, effort and funding compared to experimental designs. This was an important factor given the time and cost constraints of the original study's author (Aguinis & Bradley, 2014; Erfanian et al., 2020). While EVM has not been extensively used in the field of management, it has been relied upon by various researchers, as evidenced in major journals within the field (Aguinis & Bradley, 2014).

In light of Pieterse's (2016) abovementioned time and cost constraints as well as the benefits of EVM, it is considered an appropriate method to employ to sufficiently control the dependent variable stimulus – DIM – as well as to collect data for the original study. The use of the original study's primary data for secondary analysis is appropriate given that the data allowed for the investigation of the present study's research question and hypotheses. In terms of quality, the evaluation of the data indicated that it was appropriate for the required analyses of the present study (Hox & Boeije, 2005). Furthermore, the benefit of secondary analysis is that it avoids the financial and logistical challenges associated with the collection of primary data (Miller & Brewer, 2003).

3.3.1 Situational Interview Construction

The original study's author (Pieterse, 2016) employed best practice, as set out by Aguinis and Bradley (2014), for the construction of the 'paper people' (Aguinis & Bradley, 2014; Bradbury-Jones et al., 2014). Three entry-level situational interviews for a graduate management trainee role were created. A situational interview consists of questions that ask a candidate how they would behave in a hypothetical job-related scenario (Lievens & Peeters, 2008).

The situational structured interview is considered appropriate for this study given findings that suggest that this type of interview triggers the use of IM tactics (Lievens & Peeters, 2008; Peeters & Lievens, 2006; Roulin et al., 2015). It is suggested that it may be easier to employ DIM tactics in situational interviews given that the interview questions are hypothetical and harder to verify (Roulin et al., 2015). Furthermore, situational interview questions are considered to contribute towards question sophistication which is an aspect of interview structure considered to be related to interviewer rating quality (Melchers, Lienhardt, Von Aarburg, & Kleinmann, 2011).

In addition, an entry-level graduate management trainee interview is considered relatable to students who may either have had the experience of such interviews or possess the knowledge required for such an interview (Aguinis & Bradley, 2014). While a large portion of participants indicated that they have had some form of work experience, it can be reasonably expected that most of them have not had extensive work experience and are likely to be more eligible to apply for entry-level positions, making the situational interview more relatable and appropriate (Huffcutt et al., 2001).

The type of interview scenario employed in the original study allows participants to realistically immerse themselves in the described scenarios – the inevitable next step for most participants once they have completed their respective programmes is to become part of selection

processes to secure employment. This adds to the realism of the vignette and external validity. Interview transcripts of three candidates were presented to all participants. The presentation of the same interview transcripts to all participants allows for the comparison of the ability to detect DIM across all participants (Aguinis & Bradley, 2014).

The two interview dimensions assessed in the situational interviews included competencies, perseverance and organisational skills (organisation behaviour). The choice of competencies by Pieterse (2016) is considered appropriate given the sample (tertiary students) to which they were being presented, i.e., these competencies were simple enough to be understood by participants who have minimal interviewing experience. These competencies have also been found to be related to future job performance (Peeters & Lievens, 2006; Wade & Parent, 2001).

Each interview consisted of one question and one prompt per competency. This choice of approach is consistent with previous IM studies and considered appropriate for the study's data collection given its potential contribution towards limiting irrelevant information (Ellis et al., 2002; Peeters & Lievens, 2006). All three interviews posed the same interview question per competency to the candidates (prompts were specific to each candidate). This meets the interview structure requirement of question consistency that relates to rating quality (Melchers et al., 2011).

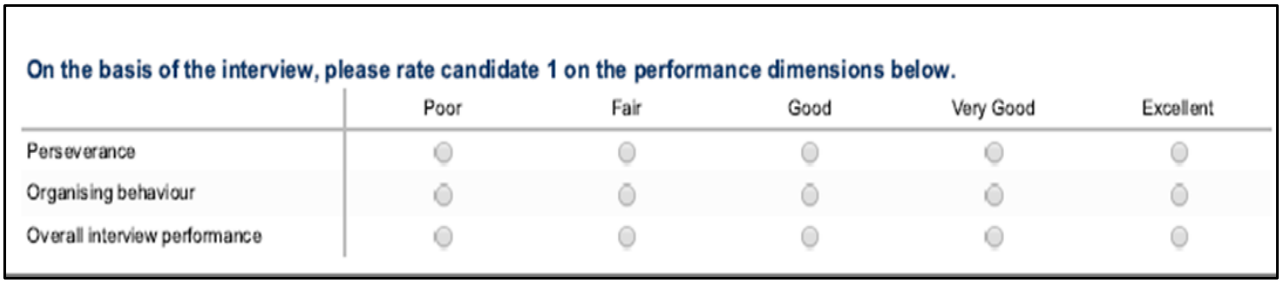
Candidate responses were based on a 5-point behaviourally anchored rating scale (BARS) and represented what *poor* (1), *average* (3) and *very good* (5) behaviours look like with respect to the competencies. Candidate performance varied on the two competencies both within interviews and across the three interviews. For example, Candidate 1 displayed *very good* (5) perseverance and organisation skill (5), while Candidate 2 showed *poor* (1) perseverance and *average* (3) organisational skill. Candidate 3 showed an *average* (3) perseverance and *good* (4) organisational skill performance. Behaviourally anchored rating scales achieve evaluation standardisation,

another aspect of the interview structure that has been found to improve rating quality (Melchers et al., 2011). Please see Appendix B for an example of the vignettes.

Participants were provided with a description of each competency before being asked to rate a candidate's performance on each competency as well as on overall interview performance on a 5-point scale; three interview dimension scores were requested per candidate (perseverance, organisation skills and overall interview performance). An example of the survey question can be seen in Figure 2.

Figure 2

Interview Dimensions Rating Scale



3.3.2 Deceptive Impression Management Cues.

The second part of the stimulus material consisted of the four types of DIM behavioural cues (deceptive ingratiation, slight image creation, extensive image creation and image protection). The DIM cues developed were based on the findings by Schneider and his colleagues regarding which behavioural cues significantly related to DIM (Schneider et al., 2015). The choice of the original study's author is appropriate for the study given that using this source type (research findings) in the construction of vignettes is considered a contributing factor towards internal validity (Erfanian et al., 2020).

DIM cues can occur at a micro-level (verbal, e.g., speaking and non-verbal, e.g., smiling cues) and a macro-level (e.g., anxiousness and attentiveness; Schneider et al., 2015). Pieterse (2016) identified significant relationships between these different behavioural cues as well as types of DIM tactics that were then placed within candidate responses. Given the positive relationship between DIM and interview performance, interview dimensions displaying poor performance contained fewer DIM cues (Schneider et al., 2015).

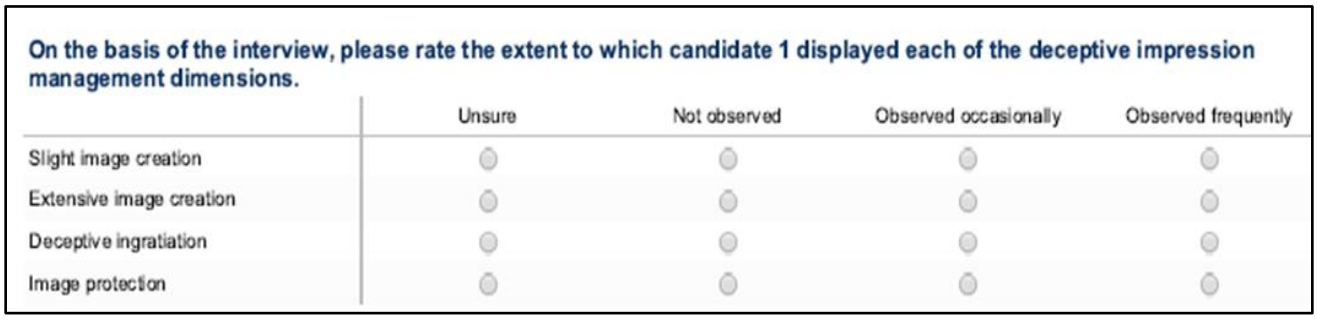
Two behavioural cues per type of DIM were 'coded' into the interview transcripts, thus resulting in varying levels and types of DIM tactics per interview transcript. Pieterse (2016) assigned a DIM 'true score' to each of the vignettes. A true score can be described as the required standard of a specific dimension rating based on expert knowledge and input. It is calculated by averaging scores across experts for a specific dimension (Sulsky & Balzer, 1988).

In the case of the original study, the author applied expert knowledge of DIM behavioural cues based on learnings from Schneider et al. (2015) to establish DIM true scores. The approach used to code DIM into transcripts and apply true scores is considered appropriate given the use of vignettes for the study stimulus. The author could have possibly benefited from obtaining the views of colleagues who are also experts in DIM cues to assess interrater-agreement on DIM tactics, constructed and refined where required (Peeter & Lievens, 2006). However, given the study time and resource constraints, the implementation of this approach is considered suitable.

The DIM cues were described to participants before they were asked to rate the frequency observed on a four-point scale, the response points of which ranged from 'not at all observed' to 'very frequently observed'; the option for 'unsure' was also included. An example can be seen in Figure 3.

Figure 3

DIM Tactics Rating Scale



On the basis of the interview, please rate the extent to which candidate 1 displayed each of the deceptive impression management dimensions.

	Unsure	Not observed	Observed occasionally	Observed frequently
Slight image creation	○	○	○	○
Extensive image creation	○	○	○	○
Deceptive ingratiation	○	○	○	○
Image protection	○	○	○	○

3.3.3 Accuracy Measures and Scoring

Accuracy scores represent the study's dependent variable. An accuracy score was computed for each participant consistent with previous studies (Christiansen et al., 2005; De Kock et al., 2015; Letzring, 2008; Powell & Goffin, 2009). In addition to Borman's DA measure, the present study made use of Cronbach's DA component measures.

3.3.3.1 Borman's Differential Accuracy. Participants' accuracy scores were determined by computing within-person profile correlations. The correlations were calculated using participants' scores of the candidates' levels of DIM and the candidates' true scores of levels of DIM in the vignettes (i.e., between the level of DIM inferred by the participant and the accuracy criterion level of DIM of the candidate vignette) at the dimension level. The participants' correlations were subsequently transformed using Fisher's r-to-z transformation to account for non-normality in Pearson correlations (Furr, 2008; Furr & Wood, 2013). This method allows for the assessment of congruence between the complete set of judgements made by the participant and the true scores (Funder & Colvin, 1997).

3.3.3.2 Cronbach's Differential Accuracy. Cronbach's DA (CDA) measure was also used to determine the participants' accuracy scores. Computation of this accuracy measure entails

the removal of a participant's overall average and their average rating levels for targets and dimensions (Powell, 2007). A higher CDA score indicates lower accuracy as it suggests a greater discrepancy between a judge's scores and the corresponding true scores (De Kock et al., 2015; Powell, 2007).

3.4 Realism

Participants were asked to rate the degree of realism of the interview transcripts on a scale of 1 to 10. Results show that on average, participants found the transcripts moderately realistic (*M* = 7.2, *SD* = 1.6), while other participants at Master's level found them highly realistic (*M = 8.7*, *SD = 1.2*). Realism refers to the degree of similarity between an experimental and natural setting. It is of particular importance for the effectiveness of vignettes (increased observed effects) and the external validity of EVM in research (Aguinis & Bradley, 2014; Erfanian et al., 2020).

3.5 Predictor Measures

3.5.1 Dispositional Reasoning

A shortened Revised Interpersonal Judgement Inventory (RIJI) was used for the original study (see Appendix C for an example question). This measure assesses an individual's general reasoning ability about traits, behaviours and situations. Items were selected based on the confirmatory factor analysis (CFA) loadings (highest) found in the study by De Kock et al. (2015). The measure consists of three subset measures, namely *trait extrapolation, trait induction* and *trait contextualisation* described in Chapter One. The three subset measures showed acceptable construct reliability derived using CFA (induction = .77; extrapolation = .81; contextualisation = .76; De Kock et al., 2015).

3.5.1.1. Trait induction. This subset measures behaviour-trait inferences (De Kock et al., 2015). The measure first describes Big Five personality traits and then presents a list of adjectives

from Goldberg's (1992) factor markers. Participants are required to identify the trait (e.g., conscientiousness) that best matches the marker adjective (e.g., thorough). Ten items were selected, an example of which can be found in Appendix C.

3.5.1.2. Trait extrapolation. This subset measure assesses the respondent's understanding of how traits co-occur (De Kock et al., 2015). Items provide a description of a fictional person and require the respondent to select one of four descriptions that most (or least) likely to be true of the fictional person. Four items were selected, an example of which can be found in Appendix C.

3.5.1.3. Trait contextualisation. The last subset aims to measure respondents' understanding of trait-situation relevance (De Kock et al., 2015). It consists of two subsets measuring the two directions of inferences. The first subset presents a trait description by listing examples of behaviours associated with high and low scores of the trait and then requires the respondent to choose which of five situations would most likely elicit the relevant behaviour. The second subset describes a situation and respondents have to identify the trait most likely to be observed in trait-relevant behaviour. Four items were selected, one of which and an example item can be found in Appendix C.

3.5.2 Personality Factors

The original study made use of a shortened version of the Big Five Inventory (BFI) to assess the personality traits of participants (John, Donahue, & Kentle, 1991; Rammstedt & John, 2007). The researcher ultimately made use of a 20-point scale where each personality trait was measured by four items. A shortened version of the 44-item original scale was deemed appropriate given that it would be time efficient and 70% of the full scale can be predicted by the shortened scale. In addition, the researcher found decent levels of test-retest stability and structural and construct validity (Rammstedt & John, 2007).

Items are short phrases assessing the most prototypical traits associated with the Big Five dimensions (Benet-Martinez & John, 1998; John, 1990). Participants rated each of the items on a 5-point Likert-type scale, ranging from 1 (‘strongly disagree’) to 5 (‘strongly agree’). Scale scores were computed as the participant’s mean item response by adding all the items on the factor subscale and dividing the result by four. Table 4 provides a summary of items used per personality trait/factor.

Table 4

Personality: Adapted Big Five Inventory (BFI)

		Indicator	Item Description: “I see myself as someone who …”
1	BFI_1	Extraversion 1	… is reserved
2	BFI_2	Agreeableness 1	… is generally trusting
3	BFI_3	Conscientiousness 1	… tends to be lazy
4	BFI_4	Neuroticism 1	… is relaxed, handles stress well
5	BFI_5	Openness 1	… has few artistic interests
6	BFI_6	Extraversion 2	… is outgoing, sociable
7	BFI_7	Agreeableness 2	… tends to find fault with others
8	BFI_8	Conscientiousness 2	… does a thorough job
9	BFI_9	Neuroticism 2	… gets nervous easily
10	BFI_10	Openness 2	… has an active imagination

(continued)

Note. Extraversion: 1(R), 6, 15, 20. Agreeableness: 2, 7(R), 12, 19(R). Conscientiousness: 3(R), 8, 13, 16(R). Openness: 5(R), 10, 14, 17. Neuroticism: 4(R), 9, 11, 18 (where R = reverse scored).

Table 4 (continued)

Personality: Adapted Big Five Inventory (BFI)

	Indicator		Item Description: "I see myself as someone who …"
11	BFI_11	Neuroticism 3	… can be tense
12	BFI_12	Agreeableness 3	… is considerate and kind to almost everyone
13	BFI_13	Conscientiousness 3	… perseveres until the task is finished
14	BFI_14	Openness 3	… is original, comes up with new ideas
15	BFI_15	Extraversion 3	… generates a lot of enthusiasm
16	BFI_16	Conscientiousness 4	… tends to be disorganised
17	BFI_17	Openness 4	… is inventive
18	BFI_18	Neuroticism 4	… worries a lot
19	BFI_19	Agreeableness 4	… is sometimes rude to others
20	BFI_20	Extraversion 4	… is talkative

Note. Extraversion: 1(R), 6, 15, 20. Agreeableness: 2, 7(R), 12, 19(R). Conscientiousness: 3(R), 8, 13, 16(R). Openness: 5(R), 10, 14, 17. Neuroticism: 4(R), 9, 11, 18 (where R = reverse scored).

3.6 Data Collection Procedure

Ethics approval was obtained for both the original study as well as the present study from the Commerce Faculty's Ethics in Research Committee (see Appendix A). In addition to the aforementioned body, the author of the original study also obtained approval from the Executive Director of Student Affairs at UCT. Upon receiving approval, an email was sent to all registered UCT students requesting their voluntary participation in the study, which they could access via the link contained in the email.

The link leads participants to the Qualtrics survey tool which contained various sections to be completed with no time limit. The first section consisted of the introductory page which indicated the two instruction pages. The first section contained the interview transcripts of three candidates and indicated that participants would be asked the same two questions about the same two competencies (organisational skills and perseverance). Each competency was first defined before its related questions were displayed. Lastly, the section indicated that candidates' verbal and non-verbal cues were recorded.

The second instruction page related to DIM, where each DIM dimension was defined and a corresponding example of what each looked like was displayed. Participants were informed that they would be rating the frequency with which they observed the DIM dimensions and that there was no time limit so as to allow them to take their time and carefully consider their evaluations.

Participants then moved onto Section 1 of the questionnaire which contained the interview transcript of each candidate and an evaluation section with a reminder of competency and DIM definitions provided. Each candidate's transcript and corresponding evaluation appeared on separate pages. Upon completion of Section 1, participants moved to Sections 2 and 3 which consisted of the personality and dispositional reasoning measures. Finally, participants moved to Section 4 which contained demographic information.

Once participants had completed the questionnaire, they were given an option to enter themselves into the competition for the cash prize by entering their email address. It was repeated that their email addresses would only be used to contact participants regarding their prize. The survey was open for 10 days.

3.7 Demographic Characteristics

At the end of the Qualtrics online survey, participants were asked to indicate their race, gender, age, academic level, faculty, frequency of subjective evaluations of people, work experience and device utilised to complete the survey.

3.8 Statistical Analysis

The statistical analysis of the present study was conducted through the use of various software programmes. Raw data from the original study (Pieterse, 2016) was cleaned and coded using the Statistical Programme for Social Sciences (SPSS; IBM, 2020) and Microsoft Excel respectively. The reliability and validity of the study measurement instruments were assessed using SPSS and RStudio (RStudio Team, 2020). Cronbach's alpha (α) was used to assess reliability and confirmatory factory analysis (CFA) was used to assess validity (Rosseel, 2012). In addition to the above, SPSS was used to calculate descriptive statistics as well as to conduct tests for multivariate assumptions of the present study variables. SPSS was also used to test the study's hypotheses; tests specifically run in SPSS include the Pearson product-moment correlation and hierarchical multiple regression. G*Power 3.1 was used to assess the power of the results of supported hypotheses of the present study. The results of the abovementioned analyses are discussed in Chapter 4.

Borman's DA was used to calculate normative profile accuracy. Borman's DA index was computed by correlating a judge's ratings for each interview dimension with corresponding true scores across targets which yielded a DA score for each dimension (Sulsky & Balzer, 1988). An overall differential score per participant was obtained by averaging each correlation using a Fisher's *r*-to-*z* transformation (Pieterse, 2016). Cronbach's DA component indices were used to calculate distinctive profile accuracy by using edited pre-existing syntax that was compiled manually and transferred into SPSS. The formulae for each index were derived from Sulsky &

Balzer (1988) and Powell (2008). A score for each participant was calculated using the transferred syntax and applying the relevant variables in SPSS.

3.9 Conclusion

The above chapter illustrated that the present study made use of secondary data from an original study that used a within-person research design. It described the study participants, acquired through non-probability convenience sampling and indicated the appropriateness of this approach for the original study based on contextual factors. It also looked at appropriateness of secondary analysis research design to investigate the research question of the present study. The chapter also discussed and reviewed the stimulus material development (vignettes) that would enable the investigation of DIM accuracy in the interviews. The review indicated the considerations and appropriateness of the study. Finally, the chapter outlined the statistical analyses which included computing Pearson's product-moment correlation coefficients and hierarchical multiple regressions used to investigate the proposed hypotheses.

Chapter 4: Results

The present research aims to examine the relationship between dispositional reasoning in the judgement accuracy of DIM use in interviews when different accuracy measures are employed. This chapter begins by investigating the psychometric properties of the measurement scales used. Descriptive statistics of all variables are then presented. Finally, the detailed results of the testing of the three study hypotheses and additional analyses are discussed.

4.1 Measurement Scales

Internal consistency and construct validity of the study measures were assessed. Internal consistency refers to the degree to which scale items measure the same underlying construct, indicating whether or not a scale is reliable (Pallant, 2007). The commonly used Cronbach's α was the coefficient used to assess scale reliability. This measure indicates the average correlation between items of the same scale. When a scale shows acceptable internal consistency, it gives greater confidence that results generated from data yielded from the scale are reliable. The general guideline for an acceptable Cronbach's α value lies between .70 and .80 (Field, 2009; Pallant, 2007).

A couple of factors are however highlighted for consideration given their potential impact on the Cronbach's α value. The first factor includes the assessment of psychological constructs by a scale. For such scales, it can reasonably be expected that a Cronbach's α yield may be lower than .70 because of the diversity of psychological constructs (Field, 2009). The second factor relates to the number of items on a scale. When scales have a small number of items (e.g., < 10), it may yield a lower Cronbach's α. This makes the corrected item-total correlation statistic quite useful in the assessment of scale reliability. The present study applied a cut-off of .30 for consideration of acceptable corrected item-total correlations (Field 2009; Pallant, 2012).

Construct validity refers to the extent to which a scale measures what it claims to measure theoretically. As such, the construct validity of the study's psychometric measures is also discussed (Field, 2009). While Exploratory Factor Analysis (EFA) techniques are commonly used in the initial development of a scale, they are not recommended as a technique with which to confirm dimensionality for the present study given that the study measurement instruments are established scales. This makes Confirmatory Factor Analysis (CFA) the appropriate technique to employ in this regard (Field, 2009; Worthington & Whittaker, 2006). In the case of the amended (Revised Interpersonal Judgement Inventory) RIJI, results found in previous studies are considered a good indication of fit for the whole scale as well as the three subscales (De Kock et al., 2015; Pieterse, 2016).

Results show that data were not multivariate normal for both scales. In addition to this, in the case of the adapted RIJI scale, item-level correlations were dichotomous. In response to these two findings, certain indices formed part of the strategy to assess validity. The robust estimation method was considered the appropriate approach to employ to assess validity (Li, 2016). Indices used to assess and report on validity are in line with the minimum recommended number of fit indices (Jackson et al., 2009; Worthington & Whittaker, 2006). These include the Chi-square Test Statistic, including related degrees of freedom and significance level. This is an overall model fit index that tests the hypothesis that all variables are uncorrelated (Pallant, 2012).

Another index is the Root Mean-Square Error of Approximation (RMSEA), an absolute fit index that expresses the fit between a hypothesised model and observed data (Hu & Bentler, 1998, 1999). A cut-off of between .50 and .80 was applied for consideration of fair fit (Hu & Bentler, 1998). This cut-off is considered appropriate given the current debate around which fit indices are the best as well as the matter of how index performance may differ depending on study attributes (e.g., sample size) (Worthington & Whittaker, 2006).

The robust Comparative Fit Index (CFI), an incremental fit measure, was also employed. This index type looks at the incremental improvement of a model's fit by comparing a target model to a baseline model (null model), the variables of which are independent and uncorrelated (CFI cut-off ≥ .95; Hu & Bentler, 1998; Worthington & Whittaker, 2006).

The above combination of indices is considered appropriate given findings that suggest that they are not sensitive to sample size and they perform well where model misspecification detection is concerned (Jackson et al., 2009). In addition to the above indices, the Satorra-Bentler Scaled Chi-square was used. This statistic is important when multivariate non-normality is found as it manages the risk of inflated goodness-of-fit statistics resulting from data that is not multivariate normal (Bryant & Satorra, 2012).

Scale reliability and validity were evaluated. Assessing scale reliability alone for a scale is insufficient to obtain an indication of the appropriateness of a study scale (Field, 2009). It is important to also evaluate the scale's construct validity given that acceptable reliability does not necessarily guarantee a scale's validity or that it shows acceptable dimensionality (Field, 2009). The current study's results are presented below.

4.1.1 Dispositional Reasoning

Dispositional reasoning was measured with the 10-item *Trait Induction*, 4-item *Trait Extrapolation* and 4-item *Trait Contextualisation* subscales that make up the amended 18-item RIJI (De Kock, 2015). Data were transformed to create dichotomous scores for each participant (where 1 = correct and 0 = incorrect). For example, for *Trait Induction*, the accurate association with 'thorough' would be *conscientiousness*. Thus, participants who chose this category would receive a '1', while those who chose any of the remaining four categories would receive a '0'.

Evaluation of internal consistency and validity is considered inappropriate for scales comprised of items with varying difficulty (Scheepers, 2004). Such items tend to show low internal consistency and can yield a Cronbach's α that is unacceptable. Despite the above, internal consistency was assessed for this study and is discussed below; a discussion on scale

validity then follows. The validity of the RIJI subscales was not assessed for the present study; however, previous findings illustrate support for the validity of the subscale content (De Kock et al., 2015).

4.1.1.1 Reliability. Despite the challenge of items with varying difficulty and their impact on reliability, the scale displayed an acceptable internal consistency (Cronbach's α = 0.70) with internal item-total correlations ranging from -.24 to .45 (Pallant, 2007; Scheepers, 2004). The lower item-total correlations can be attributed to the nature of the RIJI scale (comprised of items with varying difficulty; Scheepers, 2004). While some item-total correlations were below the cut-off of .30, it was decided to retain all items given that the exclusion of items would not significantly improve the scale's internal consistency (Pallant, 2007). Lack of potential improvement could be attributed to the length of the scale (Worthington & Whittaker, 2006). This would further contract an already shortened version of the scale. Table 5 provides a summary of the results.

Table 5

RIJI: Cronbach α and Corrected Item-Total Correlations*

Item	Induction	Extrapolation	Contextualisation
RIJI_4	.35		
RIJI_8	.43		
RIJI_9	.45		
RIJI_12	.33		
RIJI_14	.29		
RIJI_15	.37		
RIJI_16	.24		
RIJI_17	-.30		

(continued)

Note. N = 516; α = standardised Cronbach alpha (1951). *RIJI = Revised Interpersonal Judgement Inventory

Table 5 (continued)

RIJI: Cronbach α and Corrected Item-Total Correlations*

Item	Induction	Extrapolation	Contextualisation
RIJI_18	.22		
RIJI_20	.38		
RIJI_21		.35	
RIJI_27		.29	
RIJI_43		.20	
RIJI_44		.19	
RIJI_52			.28
RIJI_57			.29
RIJI_61			.20
RIJI_64			.20

Note. N = 516; α = standardised Cronbach alpha (1951). *RIJI = Revised Interpersonal Judgement Inventory.

4.1.1.2 Validity. Confirmatory Factor Analysis (CFA) was conducted using RStudio (RStudio Team, 2020) to assess the RIJI scale. The model structure consists of Induction, Extrapolation and Contextualisation factors. The items of the subscales serve as indicators of each of the factors.

The assumptions of multivariate normality and linearity were assessed using IBM SPSS 27 (IBM, 2020). The data were not multivariate normal (see 4.3). Given that a non-normal distribution violates an assumption of CFA, robust estimation was used to conduct CFA for the RIJI scale (Li, 2016; Field, 2009). CFA was performed using data from 516 participants. There were no missing data.

Robust maximum likelihood estimation was employed to estimate all models. The Chi-square of independence model that tests the hypothesis that all variables are uncorrelated was rejected: χ^2 (135, N = 516) = 291.24, p < .001. The hypothesised single-factor structure model was tested thereafter, for which acceptable support was found: Satorra-Bentler χ^2 (135, N = 516) = 285.79, p < .001, robust CFI = .77, RMSEA = .07, 90% CI: (0.07; 0.08) (Hooper, Coughlan,

& Mullen, 2008). While the CFI was below .95, it is considered an appropriate indication of fit given that it is close to 1.0 (Hu & Bentler, 1998, 1999; Worthington & Whittaker, 2006). Given that the initial model fit was satisfactory, post-hoc model modifications were not performed.

All factor loadings, except three, were significant ($p < .05$) indicating that all items are reasonable indicators of their respective latent variables: *Induction, Extrapolation and Contextualisation* (Tabachnick & Fidell, 2001). The three items that were not significant represented the *Induction* factor of RIJI. Item coefficients ranged from -.41 to .70. Given that the scale was a shortened version of an existing one, it was considered appropriate to maintain the length of the scale for analysis purposes. Furthermore, based on previous research findings that demonstrate an acceptable fit for the RIJI scale, it was considered appropriate to retain the items.

4.1.2 Personality Factors

Personality was measured using 20 items from the BFI scale which comprises five personality factors or subscales. Participant personality traits were scored by first reverse coding negatively worded items then averaging participant scores per personality trait. A higher score indicates a higher level of a personality trait (John et al., 1991; Rammstedt & John, 2007). Table 6 provides a summary of items used in the study.

4.1.2.1. Reliability. The subscale (personality factors) reliabilities generally illustrated good internal consistency (Pallant, 2007). Results for Extraversion illustrated a good internal consistency (Cronbach's $\alpha = .75$) with all inter-item correlations displaying a value above .30; therefore, all were acceptable. The *Agreeableness* subscale displayed an initial internal consistency of $\alpha = .61$. It was decided to omit the item (BFI_2 "is generally trusting") given that it had an item-total correlation of less than .30. Removing this item only increased the internal consistency of the subscale to a Cronbach's α of .62. The *Conscientiousness* subscale displayed acceptable internal consistency (Cronbach's $\alpha = .70$).

Table 6

Big Five Inventory (BFI): Cronbach α and Corrected Item-Total Correlations

	Dimension/Subscale						
Item	E	A	C	O	N	α	*M*
BFI_1	.51						
BFI_6	.63					.75	3.23
BFI_15	.48						
BFI_20	.57						
BFI_7		.40					
BFI_12		.41				.62	3.44
BFI_19		.51					
BFI_3			.46				
BFI_8			.53			.70	3.65
BFI_13			.51				
BFI_16			.50				
BFI_10				.37			
BFI_14				.58		.69	3.78
BFI_17				.58			
BFI_4					.51		
BFI_9					.53	.73	3.20
BFI_11					.45		
BFI_18					62		

Note. *N* = 516; α = standardised Cronbach's alpha (1951). E = Extraversion; A = Agreeableness; C = Conscientiousness; O = Openness to experience; N = Neuroticism.

The *Openness* subscale initially illustrated a low internal consistency of .49 with two item-total correlations below .30 (.04 and .28). The initial removal of the lowest item (BFI_5, "has few artistic interests") yielded a more acceptable internal consistency and improved item-total correlations for the remaining items (Cronbach's α = .69). The removal of item BFI_10 would have led to an improvement of the alpha value (α = .76); however, the item was retained to meet the recommendation of three observed variables per latent variable (Field, 2009; Worthington & Whittaker, 2006). The *Neuroticism* subscale showed an acceptable internal

consistency (Cronbach's α = .73), with acceptable item-total correlations ranging from .45 to .62. Table 6 provides a summary of the results.

4.1.2.2. Validity. CFA was conducted using RStudio (RStudio Team, 2020) to assess the five-factor structure of the BFI. The model structure consists of the factors Extraversion, Agreeableness, Conscientiousness, Openness and Neuroticism. The items of the subscales serve as indicators of each of the factors. The assumptions of multivariate normality and linearity were assessed using IBM SPSS 27 (IBM, 2020). As mentioned above, the data were not multivariate normal. Given that a non-normal distribution violates an assumption of CFA, robust estimation was used to conduct CFA for the BFI scale (Field, 2009; Li, 2016). CFA was performed using data from 516 participants. There were no missing data.

Robust maximum likelihood estimation was employed to estimate all models. The Chi-square of independence model was rejected: χ^2 (125, N = 516) = 403.46, p < .001. The hypothesised structure model was tested thereafter, for which acceptable support was found: Satorra-Bentler χ^2 (125, N = 516) = 455.52, p < .001, robust CFI = .77, RMSEA = .07, 90% CI: (0.07; 0.08) (Hooper, Coughlan, & Mullen, 2008). Given that the CFI is close to 1.0, it was considered appropriate despite it being below .95 (Hu & Bentler, 1998, 1999; Worthington & Whittaker, 2006). Post-hoc model modifications were not performed given the satisfactory initial model fit.

All factor loadings were significant (see Table 7) indicating that all items are reasonable indicators of their respective latent variables: *Extraversion*, *Agreeableness*, *Conscientiousness*, *Openness* and *Neuroticism* (Tabachnick & Fidell, 2001).

Table 7

Completely Standardised Factor Loadings of the Adapted Big Factor Index (BFI)

Item	Extraversion	Agreeableness	Conscientiousness	Openness	Neuroticism
1	.65				
6	.83				
15	.66				
20	.67				
7		.59			
12		.41			
19		.85			
3			.61		
8			.54		
13			.63		
16			.58		
10				.38	
14				.76	
17				.66	
4					.61
9					.78
11					.53
18					.85

Note. $N = 516$.

4.2 Descriptive Statistics

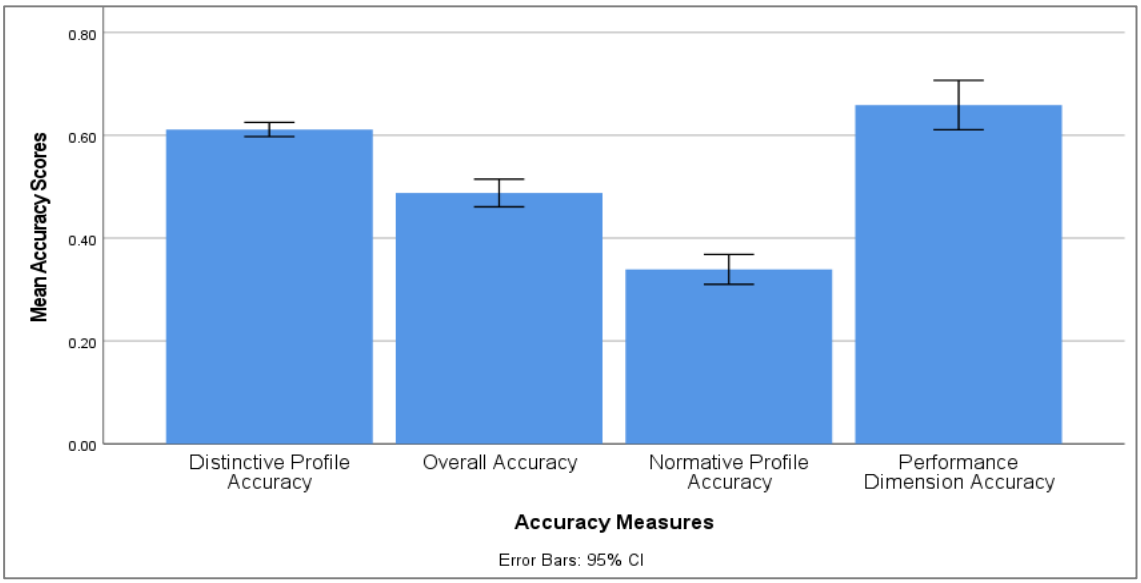
Table 6 illustrates the means, standard deviations and Pearson correlation matrix of all variables in the study. With regards to individual constructs, the table illustrates that on average, participants show moderate dispositional reasoning ability (M = 71.9%, SD = 16.0%). The means for personality show a similar spread of scores for all five traits. Participants scored higher on Openness and lower on Neuroticism. In terms of accuracy (Fisher-transformed),

results indicate that the means for normative and distinctive profile accuracy are low (Profile Accuracy[a], *M* = .34; Distinctive Accuracy[b], *M* = .61).

Results suggest that participants had better distinctive profile accuracy than normative profile accuracy. This indicates that participants showed the ability to accurately judge DIM when required to account for profile normativeness (see Figure 4). This indicates that participants show the ability to judge candidate interview dimensions accurately from vignettes when assessing accuracy using traditional profile accuracy measures. On average, overall accuracy scores were above the midpoint of zero and varied between participants, suggesting individual differences in accuracy (*M* = .49, *SD* = .31).

Figure 4

Graph depicting mean accuracy scores of distinctive profile accuracy, overall accuracy, normative profile accuracy and performance dimension accuracy.



[a] Profile accuracy is operationalised as Borman's differential accuracy. High scores imply high accuracy.

[b] Distinctive accuracy is operationalised as Cronbach's differential accuracy. High scores imply low accuracy.

Table 8

Descriptive Statistics and Inter-Correlations[a] of Study Variables

	Variable	*M*	*SD*	1	2	3	4	5	6	7
1	Gender[b]	1.61	.50	-						
2	Extraversion	3.23	.82	-.03	-					
3	Agreeableness	3.46	.70	.06	.05	-				
4	Openness	3.77	.71	-.11**	.29**	-.01	-			
5	Conscientiousness	3.65	.72	.09*	.14**	.10*	.12**	-		
6	Neuroticism	3.19	.80	.25**	-.27**	-.17**	-.05	-.08	-	
7	Dispositional Reasoning[c]	71.92	16.01	.09	-.10*	-.05	-.02	.05	.02	-
8	Trait Induction	61.34	20.74	.06	-.06	-.04	.01	.06	-.01	.91**
9	Trait Extrapolation	86.63	19.03	.05	-.06	-.07	-.02	.00	.07	.63**
10	Trait Contextualisation	83.67	22.17	.11*	-.13**	.00	-.08	.03	.04	.59**
11	Distinctive Accuracy[d e]	.61	.16	-.04	.03	-.08	-.07	-.04	-.05	-.10*
12	Normative Accuracy[d f]	.34	.34	.06	.05	.01	.07	-.04	.02	.09*
13	Performance Accuracy[d]	.66	.55	.00	.00	-.02	-.05	.12**	-.10*	.11**
14	Overall Accuracy[d]	.49	.31	.04	.02	.02	-.02	.10*	-.09*	.15**

(continued)

Note. N = 516. Trait induction = judges' ability to infer traits from behaviour cues; Trait extrapolation = understanding of how traits naturally co-vary; Trait contextualisation = to know how situations affect.

[a]These correlations have not been corrected for unreliability and restriction of range. [b]Gender was coded such that men were 1 and women were 2. [c]Dispositional reasoning and individual factor scores are presented as percentages.

Note. [d]Accuracy scores are Fisher-transformed (r-to-z) profile correlations between participants' ratings at item-level and true score estimates. [e]Distinctive accuracy = distinctive profile accuracy, operationalised as Cronbach's differential accuracy. High scores indicate higher accuracy. [f]Normative accuracy= normative profile accuracy, operationalised as Borman's differential accuracy. High scores indicate lower accuracy.

* $p < .05$. ** $p < .01$

Table 8 (continued)

Descriptive Statistics and Inter-Correlations[a] of Study Variables

	Variable	*M*	*SD*	8	9	10	11	12	13	14
1	Gender[b]	1.61	.50							
2	Extraversion	3.23	.82							
3	Agreeableness	3.46	.70							
4	Openness	3.77	.71							
5	Conscientiousness	3.65	.72							
6	Neuroticism	3.19	.80							
7	Dispositional Reasoning	71.92	16.01							
8	Trait Induction	61.34	20.74	-						
9	Trait Extrapolation	86.63	19.03	.38**	-					
10	Trait Contextualisation	83.67	22.17	.28**	.29**	-				
11	Distinctive Accuracy[d e]	.61	.16	-.06	-.15**	-.07	-			
12	Normative Accuracy[d f]	.34	.34	.08	.08	.05	-.57**	-		
13	Performance Accuracy[d]	.66	.55	.14**	.01	.05	-.04	-.07	-	
14	Overall Accuracy[d]	.49	.31	.16**	.05	.08	-.31**	.36**	.85**	-

Note. N = 516. Trait induction = judges' ability to infer traits from behaviour cues; Trait extrapolation = understanding of how traits naturally co-vary; Trait contextualisation = to know how situations affect.

[a]These correlations have not been corrected for unreliability and restriction of range. [b]Gender was coded such that men were 1 and women were 2. [c]Dispositional reasoning and individual factor scores are presented as percentages. [d]Accuracy scores are Fisher-transformed (r-to-z) profile correlations between participants' ratings at item-level and true score estimates. [e]Distinctive accuracy is operationalised as Cronbach's differential accuracy. High scores indicate higher accuracy. [f]Profile accuracy is operationalised as Borman's differential accuracy. High scores indicate lower accuracy.

* $p < .05$. ** $p < .01$

4.3 Tests for Assumptions

Assumptions of multivariate analysis were conducted for relevant study variables using IBM SPSS 27 (2020). Univariate and multivariate tests were conducted for normality, linearity, outliers, homoscedasticity and multicollinearity for study variables using IBM SPSS 27 (2020).

4.3.1 Normality, Skewness, Kurtosis

Assessment of normality, skewness and kurtosis indicated that data were not normally distributed. Significant findings were obtained for Kolmogorov-Smirnov tests ($p < 0.05$); thus the hypotheses for normality were rejected (Field, 2009). While a normality violation was found, skewness of data may be attributable to the homogeneity of the sample (similarity of participants who are all university students). Additionally, statistical tests that would be used for analysis are relatively robust to violations of normality (correlation and regression analyses; Field, 2009).

4.3.2 Outliers

Best practice guidelines by Aguinis et al. (2013) were used to manage outliers. A high cut-off point was used to prevent unnecessary exclusion of accurate scores; outliers may indicate top performers and be accurate. The exclusion cut-off point applied to transformed *z*-scores is |3.29|, $p < .001$. No data points were removed from the data set. Bivariate scatter plots were used to assess linearity and homoscedasticity and revealed no clear deviations.

4.3.3. Multicollinearity

Multicollinearity was assessed for the regression analysis discussed below. It is evaluated given that it can pose a risk to the accuracy of regression coefficient estimates (Field, 2009). Multicollinearity exists when there is a high correlation between independent variables ($r > .90$) (Pallant, 2011). The low correlations ($r < .10$) between the study predictors in Table 8 provide a

preliminary indication that multicollinearity does not exist (see Dispositional Reasoning, Extraversion, Agreeableness, Conscientiousness, Openness, and Neuroticism in section 4.2).

This is further supported by results of Tolerance Values (TV) and Variance Inflation Factors (VIF). A value below .10 for TV and a value exceeding 10 for VIF is indicative of a violation of multicollinearity (Field, 2009). Results for this study's independent variables are as follows: Extraversion (TV = .84; VIF = 1.18); Agreeableness (TV = .97; VIF = 1.04); Openness (TV = .91; VIF = 1.10); Conscientiousness (TV = .96; VIF = 1.04); Neuroticism (TV = .90; VIF = 1.11); and Dispositional Reasoning (TV = .98; VIF = 1.02). Results thus show that multicollinearity was not a concern.

4.4 Hypotheses Testing

4.4.1 Hypothesis 1:[3] *Normative Accuracy as Criterion*

The first hypothesis postulated that interviewers' dispositional reasoning is positively related to normative profile accuracy[4] of DIM in interviews. Results showed a positive statistically significant relationship between scoring higher on dispositional reasoning and the ability to obtain normative profile accuracy of DIM in interviews ($r = .09$, $p < .05$). While significant, the strength of this relationship is considered small (Pallant, 2007). Findings do however indicate that judges who scored higher in dispositional reasoning were better able to accurately discriminate the level of DIM relative to the other interview dimensions for a candidate.

[3] Pearson product–moment correlation was used to test all hypotheses.
[4] Profile accuracy is operationalised as Borman's differential accuracy (see Methods section).

4.4.2 Hypothesis 2:[5] *Distinctive Accuracy as Criterion*

The second hypothesis postulated that interviewers' dispositional reasoning are positively related to distinctive profile accuracy[6] of DIM in interviews. Results showed a small, statistically significant relationship between scoring higher on dispositional reasoning and the ability to obtain the distinctive profile accuracy of DIM in interviews (r = -.11, p < .05). Given that lower scores on the distinctive accuracy measure denote greater accuracy, this result showed support for the hypothesis and indicated that judges who scored higher in dispositional reasoning were better able to accurately discriminate the level of DIM relative to the other interview dimensions and distinguish this from the average candidate profile.

4.4.3 Hypothesis 3[7].

A forward hierarchical regression was performed to test whether dispositional reasoning explains additional variance in normative and distinctive profile accuracy (used in H3a and H3b, respectively) of DIM, over interviewers' personality. Tables 9 and 10 illustrate regression results for models where the five personality factors were entered as a set in Step 1, followed by dispositional reasoning, in Step 2.

For normative profile accuracy, R was not significantly different from zero $F(6, 509)$ = 1.74, p = .11, with R^2 = 0.02. The adjusted R^2 value (.01) indicates that dispositional reasoning explains .1% variability in increment to explain accuracy. In the case of distinctive profile accuracy, R was significantly different from zero $F(6, 509)$ = 2.60, p = .02, with R^2 = 0.03. The

[5] Pearson product–moment correlation was used to test all hypotheses.
[6] Distinctive accuracy is operationalised as Cronbach's differential accuracy (see Methods section).
[7] Hierarchical multiple regression was used to test Hypothesis 3.

adjusted R^2 value of .02 indicates that dispositional reasoning explains .2% variability in increment to explain accuracy. Support was thus found for H3b.

Table 9

Results of Hierarchical Regression Analyses of Normative Profile Accuracy Scores[a] on Dispositional Reasoning and Personality Factors

Predictor	Step 1	Step 2
	β	β
Step 1		
Extraversion	.04	.05
Agreeableness	.02	.03
Openness	.07	.07
Conscientiousness	-.05	-.06
Neuroticism	.04	.04
Step 2		
Dispositional reasoning		.10
Total R^2	.01	.02
ΔR^2	.01	.01

Note. N = 516. [a]Accuracy scores are Fisher-transformed (r-to-z) profile correlations between participants' ratings at dimension level and subject matter expert true score estimates.

In summary, for normative profile accuracy, the addition of dispositional reasoning to the equation with personality factors resulted in a non-significant increment in R^2. For distinctive profile accuracy, on the other hand, the addition of dispositional reasoning to the equation with personality factors resulted in a significant increment in R^2.

Table 10

Results of Hierarchical Regression Analyses of Distinctive Profile Accuracy Scores[a] on Dispositional Reasoning and Personality Factors

Predictor	Step 1	Step 2
	β	β
Step 1		
Extraversion	.05	.04
Agreeableness	-.09	-.10
Openness	-.08	-.08
Conscientiousness	-.03	-.03
Neuroticism	-.06	-.03
Step 2		
Dispositional reasoning		-.11
Total R^2	.02	.01
ΔR^2	.03	.02*

Note. N = 516. [a]Accuracy scores are Fisher-transformed (r-to-z) profile correlations between participants' ratings at dimension level and subject matter expert true score estimates. *p < .05 (two-tailed).

4.5 Further Analyses

4.5.1 Correlations

4.5.1.1 Personality. The relationships between the remaining personality factors (Extraversion, Neuroticism, and Conscientiousness) and accuracy were analysed using Pearson's product-moment correlation. Table 8 illustrates a significant small positive relationship between Conscientiousness and interview accuracy (r = .12, p < .01) and overall accuracy (r = .10, p < .05). These findings are consistent with the findings of the original study. Findings also show a significant small negative relationship between Neuroticism and interview accuracy (r = -.1, p < .05) and overall accuracy (r = -.1, p < .05). No significant relationship was found between the personality factors and both the normative and distinctive profile accuracy of DIM in interviews.

4.5.1.2 Dispositional reasoning components. Findings show a significant small positive relationship between induction and interview accuracy ($r = .14$, $p < .01$) and overall accuracy ($r = .16$, $p < .01$). This illustrates that judges who have higher behaviour-trait knowledge are better able to understand how traits relate to manifest behaviours, which enables them to infer the level of performance on interview dimensions (e.g., perseverance) and overall interview performance more accurately (De Kock et al., 2015, 2017).

A significant small negative relationship was found between extrapolation and the distinctive profile accuracy of DIM ($r = -.15$, $p < .01$). This suggests that judges who have a higher understanding of how traits and behavioural manifestations co-occur are more accurate at detecting DIM in interviews (De Kock et al., 2015, 2017). No significant relationship was found between contextualisation and accuracy.

4.5.1.3 Accuracy measures. Consistent with the meta-analytic findings of Sulsky & Balzer (1988), a significant strong relationship was found between normative and distinctive profile accuracy; study findings however show that this relationship was negative ($r = -.57$; $p < .01$; Cohen, 1988). This different may relate to the choice of formula employed for the computation of the Cronbach accuracy component scores in the previous (Sulsky & Balzer, 1988) and current study. The meta-analytic study used the square roots of the component scores as the final estimates, whereas square roots were not applied to component scores in the present study. A negative relationship is expected between the two accuracy measures given that a lower score indicates higher accuracy in the case of Cronbach's component accuracy scores (De Kock et al., 2015; Powell, 2007). Thus, a higher Borman's DA score is related to a lower Cronbach's DA score which illustrates a higher level of judgement accuracy.

4.5.2 Statistical Power

To assess whether there is sufficient power to support the study results, power analyses were conducted for the hypotheses using G*Power 3 (Cunningham & McCrum-Garner, 2007; Faul, Erdfelder, Lang, & Buchner, 2007). A significance level of $p < .05$ was considered appropriate for study purposes. Table 11 provides a summary of power calculations for the supported hypotheses. Analyses indicate that the supported hypotheses show a good, acceptable level of power.

Table 11

Summary of Supported Hypotheses Results

Hypothesis	Correlation	*p*	Power
Hypothesis 1c	.09	.03	1.00
Hypothesis 2c	-.11	.02	.80

4.6 Conclusion

This chapter demonstrated findings of the measurement properties of the study scales, namely, the shortened RIJI and the BFI scales. Descriptive statistics were then presented. Pearson correlation coefficient calculations were completed for the first two hypotheses and hierarchical regressions were computed for the third hypothesis. Additional analyses of study variables (correlations) were also discussed. The main findings from the statistical analyses are discussed further in the following chapter.

Chapter 5: Discussion

In this section of the research dissertation, a summary and interpretation of the main findings will be provided, followed by study limitations and recommendations for further research, before concluding remarks.

5.1 Summary of Main Findings

This study sought to investigate the relationship between dispositional reasoning and the judgement accuracy of DIM use in interviews when different accuracy measures are used. In particular, the research sought to determine whether or not the operationalisation of accuracy mattered, that is, would criterion validity results generalise between alternative accuracy measures that distinguish between normative (Borman's Differential Accuracy; Borman, 1977) and distinctive profile accuracy (Cronbach's Differential Accuracy; Cronbach, 1955)? It was important to compare study results when using different accuracy measures to shed light on the potential role of decisions about accuracy operationalisation, given the suggestion that different accuracy measures may measure different aspects of accuracy (Sulsky & Balzer, 1988).

Additionally, it was important to assess the generalisability of the accuracy results of the original study across different accuracy measures within the context of judgments about DIM in interviews (Pieterse, 2016; Sulsky & Balzer, 1988). To answer the call for further research in the judgement accuracy of DIM use in interviews (Roulin et al., 2015), the study examined the relationship between the individual interviewer construct of dispositional reasoning in the judgement accuracy (normative and distinctive profile accuracy) of DIM.

Study results revealed that not only are individuals higher in dispositional reasoning skill able to accurately judge DIM from interview transcripts (when employing traditional profile accuracy measures), they are also able to account for normativeness. Apart from being able to

discern the levels of traits within the average target profile in interviews, such judges are also able to differentiate applicants' profile trait levels from the average interview candidate (Hall et al., 2018). Findings thus suggest that dispositional reasoning is a consistent predictor of judgement accuracy of DIM in interviews.

5.2 Dispositional Reasoning and Deceptive Impression Management Judgement Accuracy

Study findings showed that dispositional reasoning (considered a type of intelligence; De Kock, 2015) was positively related to both normative and distinctive profile accuracy. This is consistent with research findings that suggest that higher cognitive ability shows better accuracy of DIM detection (Melchers et al., 2020).

Profile normativeness was highlighted by Cronbach (1955) as a potential shortcoming in judgement accuracy research given that it risks inflating overall profile accuracy, thus confounding accuracy results (Allik et al., 2015; Furr, 2008; Hall et al., 2018). Given that profile normativeness relates to the average or typical profile within a particular group, it shows the similarity of two average profiles and not the unique similarity of the profiles of two individuals; in other words, it shows the extent to which their trait ratings or dimension levels are similar to each other (Allik et al., 2015; Biesanz & Human, 2010; Furr, 2008; Hall et al., 2018; Krzyzaniak et al., 2019).

The calculation of both normative and distinctive profile accuracy scores allowed the study to ascertain the effect that accounting for profile normativeness has on the judgement accuracy of DIM and interview dimensions. Results show that the strength of the relationship between dispositional reasoning and DIM judgement accuracy did not change when different measures with different conceptual bases were used. Additionally, findings indicate a strong relationship between the accuracy measures used in the study (Borman's DA and Cronbach' DA). A significant relationship is consistent with the findings by Sulsky & Balzer (1988). These findings suggests

that dispositional reasoning explains different aspects of accuracy and that there is a relationship between rating ability and ability to diagnose the trait levels within an candidate's profile (Powell, 2007; Sulsky & Balzer, 1988).

The study findings indicating a positive relationship between dispositional reasoning and normative profile accuracy were consistent with previous findings and expected given the ability afforded by dispositional reasoning to a judge (Christiansen et al., 2005; De Kock et al., 2020; Pieterse, 2016; Powell & Bourdage, 2016). Dispositional reasoning enables the understanding and knowledge of traits, behaviours and how situations influence the expression of traits (Christiansen et al., 2005; De Kock et al., 2020; Powell & Goffin, 2009). Study findings are thus in line with the organising framework by De Kock et al., (2020).

Judges who scored higher on the individual interviewer construct of dispositional reasoning, were better able to complete the final key step in the judgement accuracy process, namely, utilisation. Interviewers who were able to detect relevant deceptive behavioural cues made available in interview transcripts were better enabled through the specific interviewer characteristic of dispositional reasoning to complete the final key step in the judgement process.

Dispositional reasoning is considered a form of intelligence related to facilitating effective behaviour information processing that supports a key step in the judgement accuracy process (trait cue utilisation); this, in turn, influences judgement quality (De Kock et al., 2020). Based on the above information processing, it can be described as a mental ability. Through their dispositional reasoning ability, interviewers can detect and utilise the trait-relevant behavioural cue information depicted in interview transcripts to make judgements about the level of DIM and other interview dimensions within a target's profile.

Not only does dispositional reasoning allow a judge to achieve normative profile accuracy, but it also allows a judge to discern the degree to which a candidate's profile trait levels differ from the average interview candidate. In other words, the extent to which a candidate's profile of DIM and other interview dimension levels are above or below the average candidate's profile of DIM and other interview dimension levels (Allik et al., 2015; Biesanz & Human, 2010; Furr, 2008; Hall et al., 2018; Krzyzaniak et al., 2019). Thus, even when accounting for normativeness, judges high in dispositional reasoning show the ability to accurately judge DIM in interviews. The present study findings thus further illustrate that despite the accuracy measure used, dispositional reasoning is a consistent predictor of judgement accuracy and may predict different aspects of judgement accuracy (De Kock et al., 2020; Sulsky & Balzer, 1988), thereby adding to research in the field of dispositional reasoning and DIM judgement accuracy.

As per the recommendation of the original study (Pieterse, 2016), the subcomponents of dispositional reasoning were also evaluated and they proved worthwhile to note. For example, significant relationships were found for both induction (knowledge of how traits manifest in behaviour) and extrapolation (understanding of how traits and their manifest behaviours naturally co-vary; Christiansen et al., 2005; De Kock et al., 2015, 2017, 2020).

Findings do however differ slightly from the original study (Pieterse, 2016). A positive relationship was only found for extrapolation and the distinctive profile accuracy of DIM, while in the case of induction, a positive relationship was found for performance dimension accuracy and overall accuracy. Given that trait extrapolation enables a judge to understand how traits co-vary, a judge can accurately identify trait-based behaviours that will most likely occur and vary together using their mental models of how traits and behaviours relate to each other. This understanding enables a judge to use limited information to form accurate impressions of a

candidate such that he or she can see that a candidate who exhibits deceptive behavioural cues of extensive image creation is also likely to exhibit deceptive ingratiation cues (De Kock et al., 2015).

This finding illustrates that extrapolation enables judgement accuracy when profile normativeness is accounted for. Through the understanding of how traits co-occur, a judge can accurately discern the extent to which a candidate's levels of extensive image creation and deceptive ingratiation differ from the levels of these cues in the average candidate profile (Allik et al., 2015; Biesanz & Human, 2010; Furr, 2008; Hall et al., 2018; Krzyzaniak et al., 2019).

In the case of trait induction, a judge is enabled to accurately deduce which trait drives an individual's behaviour (De Kock et al., 2015). An accurate judge can therefore infer which behavioural cues relate to DIM and other interview dimensions, respectively. The process of induction happens when a judge identifies behaviours and evaluates those behaviours in terms of the relevant underlying latent characteristics of a candidate. These behaviours are then encoded in terms of trait concepts and it is the correct behaviour-trait inferences that enable the accurate impression of a candidate (De Kock et al., 2015).

Similar to the findings of the original study (Pieterse, 2016), trait contextualisation (ability to identify situations relevant to different traits) did not yield a significant relationship with any accuracy score. This may be related to the participants' level of understanding of the situations that elicit the behavioural cues of DIM. Research does however suggest that structured and situational interviews tend to elicit IM tactics (honest and deceptive) from candidates (Lievens & Peeters, 2008; Peeters & Lievens, 2006; Roulin et al., 2015).

Given that the dispositional reasoning component of contextualisation relates to the knowledge of how situations influence trait manifestations, this may suggest that it is an ability that can be further developed (e.g., through training). Increased development may result in

increased contextualisation and dispositional ability as a whole, which could in turn improve judgement accuracy (De Kock et al., 2015). While preliminary study findings have shown little support for this notion (Powell, 2007), it is believed that assessing this relationship in terms of the distinct three components of dispositional reasoning may be of benefit for this line of research.

The abovementioned organising framework specifically differentiates judgement and rating accuracy (De Kock et al., 2020). Once the implicit process of judgement accuracy has been completed (a judge has processed behavioural information and formed a judgement), a rendering phase takes place which determines the explicit expression of the rating a judge assigns to a target (De Kock et al., 2020). The framework suggests that this rating is related to motivating factors that have a moderating effect on rating accuracy.

Motivation is considered a moderating factor between judgements and ratings given that a judge's attitude, personal agenda or political influences may encourage him or her to apply a higher or lower rating. In the case of this study, it can be argued that study participants were motivated to take their participation seriously and do well in the rating exercise due to the incentive attached to participating in the study and obtaining accurate ratings (De Kock et al., 2020). Given that motivation was not explicitly measured in the present study, further research along these lines can be completed to understand the moderating effects of motivation factors on DIM judgement accuracy, particularly on the relationship between dispositional reasoning and DIM detection in interviews (De Kock et al., 2020).

5.3 Personality and Deceptive Impression Management Judgement Accuracy

Consistent with the original study (Pieterse, 2016), findings illustrate a significant positive relationship between conscientiousness and interview performance accuracy, as well as overall accuracy. In addition to these findings, the present study found a significant small negative

relationship between neuroticism and performance dimension accuracy as well as overall accuracy. None of the personality factors showed a positive relationship with DIM accuracy.

Conscientiousness manifests in greater detail orientation. Individuals possessing this trait are typically more thorough, organised, careful and systematic (De Kock, 2015; De Kock et al., 2020; Goldberg, 1992). Judges high in conscientiousness are more likely to be more attentive which may enable them to better detect behavioural cues in interviews. These judges are also more likely to be consistent in the utilisation of behavioural cues. These factors in turn enable these judges to make accurate judgements in interviews.

Conscientiousness is also related to motivation across a range of tasks and the desire to have successful social relationships (Christiansen et al., 2005). The characteristics of conscientiousness suggest a relation to rater motivation referred to in the organising framework mentioned previously (De Kock et al., 2020). It can therefore be argued that conscientiousness may help to explain rater motivation. Further studies that evaluate this trait's moderating effect on judgement accuracy and rating quality in interviews could be explored.

Neuroticism, also known as emotional stability, relates to the degree to which an individual is self-conscious and easily distressed (Christiansen et al., 2005; Goldberg, 1992). These characteristics may discourage social interaction. Judges who are high in neuroticism may therefore be less likely to have the opportunity to observe behavioural cues and therefore be less able to make accurate judgements about behaviour. In contrast, judges who are low in neuroticism are more likely to engage in social interactions and be more focused due to reduced self-consciousness; therefore, they have more opportunity to observe or detect and utilise behavioural cues to make accurate judgements within social contexts such as interviews (Christiansen et al.,

2005; De Kock et al., 2020). Similar to the above, further studies could explore the moderating effect of personality traits on judgement accuracy and rating quality (De Kock et al., 2020).

The above findings illustrate that personality traits that form the Big Five Index may support different forms of accuracy in the selection process, for example, interview performance accuracy. It does not however show to be a consistent predictor of DIM in interviews. Further studies may benefit from investigating narrow personality traits and their relationship with DIM judgement accuracy (De Kock, 2015; De Kock et al., 2020), for example, honesty-humility or integrity as predictors of judgement accuracy, given the potential illustrated by results in this line of research which show that narrow traits may better predictors of certain academic criteria, namely grade point average (GPA) and counterproductive academic behaviour (CAB; De Vries et al., 2011).

5.4 Limitations

One of the limitations to note for this study relates to the relative homogeneity of the sample. Study participants consisted only of university students from the same tertiary institution. Caution should therefore be taken regarding generalising the study findings to other populations within the selection field (Field, 2009). Future research could assess the generalisability of the study results to interviewers in different organisational settings and industries (Field, 2009; Pieterse, 2016).

While the characteristics of the sample represent homogeneity, it can be argued that the choice of sample is considered appropriate for both the original and the present study for several reasons. Firstly, it allowed the researcher to complete the study within the available limited time (Pieterse, 2016). Secondly, it was reasonable for the researcher to make use of university students given previous research findings that indicate that interview experience does not increase DIM judgement accuracy (Roulin et al., 2014). Thirdly, it has been previously found that university

students can accurately judge personality (Christiansen et al., 2005; De Kock et al., 2015; Letzring, 2008; Powell & Goffin, 2009); therefore, it was reasonable to expect that they would be able to accurately judge DIM.

A second limitation of the study relates to a few identified low non-significant factor loadings on the RIJI. The scale consisted of three factors (induction, extrapolation and contextualisation) that had 10, 4 and 4 items, respectively. Low factor loadings may suggest that other factors may explain the variance in an observed variable (Field, 2009). A possible reason that could explain poor factor loadings is sample size (Li, 2016; MacCallum & Widaman, 1999). The study sample size, however, does not appear to be an issue given that it was larger than 200 (N = 516; Li, 2016; MacCallum & Widaman, 1999). It could be argued that the dispositional reasoning scale used was a shortened version of an original scale and that could relate to the low factor loadings; however, the shortened scale had well-determined factors (De Kock et al., 2017; MacCallum & Widaman, 1999). As mentioned above, low factor loadings may also result from scales with varying difficulty (Scheepers, 2004).

Another potential limitation to consider relates to the issue of realism (Pieterse, 2016). The study made use of 'paper people' as candidates (vignettes). The risk of unrealistic stimuli in a study may mean results that do not represent the true relationships between study variables (Aguinis & Bradley, 2014). This was mitigated by the researcher of the original study by making use of best practice guidelines for constructing vignettes and asking subject matter experts to evaluate transcripts for realism; the experts, in turn, indicated that the vignettes were fairly realistic (Aguinis & Bradley, 2014).

While realism can potentially be increased through the use of richer stimuli such as videotaped vignettes, this may lead to increased cognitive demand. Participants were guarded

against this because written vignettes were used. The use of vignettes allows a researcher to control extraneous variables that would otherwise be more difficult to control in a field study (Aguinis & Bradley, 2014; Erfanian et al., 2020). Vignettes also allow for a fully crossed research design, such that all judges rate all the same candidates in a controlled environment. The control of information and behavioural cues is better ensured with the use of vignettes, allowing for inferences to be made about the participants' ability to make accurate judgements given that judgements are based on an isolated stimulus (Aguinis & Bradley, 2014; Kinicki et al., 1995).

The researcher of the original study (Pieterse, 2016) and the present researcher were able to effectively investigate the effect of individual differences on judgement accuracy through the verbal and behavioural cues presented to the participants in the written text. Additionally, the data portrayed an actual interview process which requires a recruiter to form impressions of a candidate based on available information during the interaction and subsequently evaluate and rate the candidate at the end of the interview based on impressions and collected data (Barrick et al., 2009). Vignettes (text stimuli) are considered an appropriate stimulus to use to investigate individual differences in judgement accuracy as previous research has shown the ability of participants to make accurate personality judgements from the text (Hall, et al., 2016).

While it can be argued that the isolation of behavioural cues may have made it easier to judge DIM, results show the ability of participants to accurately judge DIM in interviews and, more specifically, to account for normative profile accuracy in the judgement of DIM (Pieterse, 2016). Similar to the original primary study (Pieterse, 2016), there were significant effect sizes found but these were low, with DIM accuracy effect sizes being smaller than performance dimension accuracy. A possible explanation for the lower DIM effect sizes compared to the other interview dimension effect sizes may relate to participants' efforts to reduce cognitive dissonance,

leading them to focus more on performance dimension evaluations to reduce the cognitive tension brought about by needing to evaluate all variables (Simon, et al., 1995). Another reason may relate to the general impressions of candidates that may underlie the rating of interview dimensions and DIM (Roulin et al., 2015).

The transcripts used for the study were very short (Pieterse, 2016). While research suggests that interviewers can make accurate judgements from short video segments (Schmid Mast et al., 2011), there is a suggestion that DIM judgement may become more effective after a longer period of exposure (Roulin et al., 2015). However, the potential risk that comes with prolonged exposure is increased cognitive load that may make it more difficult for an interviewer to make accurate judgements (Buller & Burgoon, 1996). Cognitive load was minimised given that participants did not have to conduct the interviewers in addition to making accurate judgements; they simply had to observe the interviews.

5.5 Implications for Theory and Future Research

The present study findings suggest that the individual difference construct of dispositional reasoning may predict the judgement accuracy of DIM in interviews through enabling the two key processes in judgement explained by RAM (Funder, 1995). In other words, dispositional reasoning may better enable the processes of *cue detection* and *cue utilisation.* Through the knowledge of traits, behaviours and how traits manifest in certain situations, interviewers are better able to identify the behavioural cues that relate to DIM and to use the information gained from the interaction to form impressions about a candidate (De Kock et al., 2015). This study thus further supports the theory of RAM in the context of DIM in interviews.

The current study further contributes to research in the field of personnel selection by evaluating dispositional reasoning and the detection of DIM in interviews. Previous studies have

investigated the interviewer's dispositional reasoning ability and how it relates to the judgement accuracy of personality (Christiansen et al., 2005; Powell & Goffin, 2009) and interview accuracy (De Kock et al., 2015). The findings of the current study found no support for the dispositional reasoning component of *contextualisation*. Further research could explore the effects of increasing interviewer knowledge and understanding of situations that relate to DIM in interviews (De Kock et al., 2015).

The findings of the current study illustrate that dispositional reasoning enables the consistent ability to predict DIM across different accuracy measures. Furthermore, dispositional reasoning predicts judgement accuracy when accounting for normativeness. The current study findings thus also expand on past research on normative and distinctive profile accuracy as operationalised by using different accuracy measures, namely Borman's DA and Cronbach's DA (Hall et al., 2018; Sulsky & Balzer, 1988).

To further explore and contribute towards research regarding conceptualisation and operationalisation of accuracy measures, researchers may also investigate the relationship between dispositional reasoning and DIM in interviews when using the other Cronbach Accuracy component scores (Elevation, Differential Elevation, and Stereotype Accuracy). Given the suggestion that different accuracy measures may measure different aspects of accuracy, it is recommended that further research investigate whether or not dispositional reasoning represents an interviewer construct that makes a judge accurate across various aspects of accuracy.

The findings of the present study showed no support for the individual difference construct of personality and its predictive ability on the judgement accuracy of DIM in interviews. Some findings did however illustrate a relationship between conscientiousness and neuroticism and between performance dimension accuracy and overall accuracy. These findings are consistent with

previous research (De Kock, et al., 2020; Pieterse, 2016) and thus further imply that personality is not a predictor of DIM judgement accuracy in interviews.

The organising framework established by De Kock et al. (2020) suggests that rating quality may be moderated by rater motivation. Once a rater has formed a judgement, the rater's goals or objectives (motivation) drive how the rater assigns a rating (e.g., high or low rating). Given the associated characteristics of conscientiousness, for example, of motivation to complete tasks (Christiansen et al., 2005; De Kock, 2015; De Kock et al., 2015), future research may benefit from assessing conscientiousness as it relates to the motivation of assigning accurate ratings. Similarly, neuroticism can be evaluated in the same way given that the characteristics of neuroticism suggest a degree of drive (motivation) to participate in social interactions. The above suggestions are in line with the recommendations by De Kock et al. (2020) to assess personality as a moderator of judgement accuracy.

It is recommended that future research is carried out in the field to assess the abovementioned relationships within organisational settings, using working professionals whose responsibilities include conducting interviews. This will not only add to the body of knowledge in the selection field relating to DIM in interviews but also to the research on dispositional reasoning in real-life settings.

5.6 Implications for Practice

Given that the study findings were established from a sample comprised solely of university students, results may have limited implications for practice (Field, 2009). Further field studies are required to be able to more confidently make such recommendations for practice. Future studies may, for example, consider recruiters within organisations, or recruitment agencies.

Organisations can however be aware of the potential threat to the validity of selection decisions that are posed by the use of DIM tactics during interviews given that faking is quite common in interviews (Melchers et al., 2020). Such behaviour may affect recruiters' perceptions of candidates, which could, in turn, lead to the appointment of individuals who are not suitably matched to a job which can have monetary and non-monetary implications on an organisation (Kristof-Brown et al., 2002; Roulin et al., 2015; Roulin et al., 2016).

The potential cost of hiring an employee who is not a good fit for a job includes not only the time lost from an inefficient recruitment process but also the future associated costs that will be experienced by the organisation. These include poor performance, poor culture fit and the potential additional monetary and time costs of replacing such an employee (Marsden, 2016; Roulin et al., 2015). Given that an organisation's competitiveness lies in its talent or employees, it is imperative that the correct hiring decisions are made. It is therefore important for recruiters to be able to identify candidates who employ DIM tactics and exclude them from the applicant pool.

Previous research suggests that making accurate judgements of DIM in interviews is not an easy task for recruiters (Melchers et al., 2020; Roulin et al., 2015; Roulin & Powell, 2018). The findings of the current and the original study (Pieterse, 2016) do however imply that some judges are better at detecting DIM than others. This is explained by a judge's dispositional reasoning ability. Those higher in dispositional ability are better able to accurately judge DIM in interviews. This interviewer construct can thus form a criterion for selection of interviewers by organisations.

Given that dispositional reasoning relates to the knowledge of traits, behaviours, and how context influences expression of behaviours, it implies that it can be developed or enhanced in recruiters. Organisations can therefore support recruiters by expanding their knowledge of different types of DIM behavioural cues (slight image creation, extensive image creation,

deceptive ingratiation and image protection) and the types of situations that elicit specific DIM tactics to increase the judgement accuracy of DIM in interviews (De Kock et al., 2015). The interview transcripts (vignettes) formulated by Pieterse (2016) could be relevant to future researchers wishing to use text-based stimuli for training purposes.

It is suggested that DIM judgement accuracy may benefit from longer exposure to behavioural cues (Roulin et al., 2014), however, the potential for cognitive load may impede judgement accuracy (Buller & Burgoon, 1996). Organisations can attempt to strike a balance between these two factors by having interview recruiters only watch the interview and rate candidates instead of both conducting the interviews themselves and rating candidates. Watching could include in-person observations or videotaped interviews of candidates. This would potentially be useful especially for organisations that do not currently practice this method (Langer et al., 2017).

5.7 Conclusion

This study sought to investigate the relationship between dispositional reasoning and the judgement accuracy of deceptive impression management use in interviews when different accuracy measures are used. One of the accuracy measures typically employed in accuracy studies is Borman's Differential Accuracy (1977), which may also be considered a 'normative' profile accuracy measure. A second accuracy measure often uses is Cronbach's Differential Accuracy (1955) measure, which is considered to be a measure of 'distinctive' profile accuracy. The importance of comparing results of studies when relying on different accuracy measures stems from the suggestion that different accuracy measures may measure different aspects of accuracy. Additionally, it is considered important to assess the generalisability of accuracy study research across different accuracy measures (Sulsky & Balzer, 1988).

The current secondary research study findings illustrate that dispositional reasoning consistently may predict DIM judgement accuracy across different accuracy measures, irrespective of whether or not these are normative or distinctive accuracy measures. As such, findings suggest that results of the original study (Pieterse, 2016) may be generalisable across different accuracy measures. By implication, these findings reinforce the notion that dispositional reasoning is a consistent predictor of judgment accuracy when deceptive impression management is being evaluated. Results illustrate that recruiters with greater dispositional reasoning ability may be better able to make accurate judgements of deceptive impression management in interviews, even when the accuracy measure being used is based on normative or distinctive accuracy operationalisations. Stated otherwise, findings confirm that the results of the primary study were not, in part, based on measurement artefacts as a potential confound.

Despite particular limitations (see 5.3), this study contributes insights regarding theory and potential implications for HRM and accuracy research practice. It also makes recommendations of opportunities for future research (see 5.4). Findings suggest that dispositional reasoning represents a relatively stable characteristic of a good judge, particularly in the case of deceptive impression management detection in interviews. In sum, this study thus concludes that when selecting interviewers on the basis of their dispositional reasoning ability, the accuracy measure does not appear to matter.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351-371. https://doi.org/10.1177/1094428114547952

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods*, *16*(2), 270-301. https://doi.org/10.1177/1094428112470848

Allik, J., Borkenau, P., Hřebíčková, M., Kuppens, P., & Realo, A. (2015). How are personality trait and profile agreement related. *Frontiers in Psychology, (6),* 785. https://doi.org/10.3389/fpsyg.2015.00785

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology, 94*(6), 1394-1411. https://doi.org/10.1037/a0016532

Battaglia, M. P. (2011). Nonprobability sampling. In P. J. Lavrakas, (Ed.), *Encyclopedia of survey research methods* (pp. 524-526). Sage Publications. https://dx.doi.org/10.4135/9781412963947

Benet-Martínez, V., & John, O. P. (1998). Los cinco grandes across cultures and ethnic groups: Multitrait multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology, 75*(3), 729–750. https://doi-org.ezproxy.uct.ac.za/10.1037/0022-3514.75.3.729

Biesanz, J. C., & Human, L. (2010). The cost of forming more accurate impressions: Accuracy-motivated perceivers see the personality of others more distinctively but less normatively than perceivers without an explicit goal. *Psychological Science, 21*(4), 589-594. https://doi.org/10.1177/0956797610364121

Borman, W. C. (1977). Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance*, *20*(2), 238-252. https://doi.org/10.1016/0030-5073(77)90004-6

Borman, W. C., Hough, L. M., & Dunnette, M. D. (1978). *Performance ratings: An investigation of reliability, accuracy, and relationships between individual differences and rater error* (Contract DAHC 19-75-C-0006). Alexandria, VA: U.S. Army Research Institute. https://apps.dtic.mil/sti/citations/ADA061149

Borman, W. C. (1979). Individual differences correlates of accuracy in evaluating others' performance effectiveness. *Applied Psychological Measurement*, *3*(1), 103-115. https://doi.org/10.1177/014662167900300111

Bradbury-Jones, C., Taylor, J., & Herber, O. R. (2014). Vignette development and administration: a framework for protecting research participants. *International Journal of Social Research Methodology*, *17*(4), 427-440. https://doi.org/10.1080/13645579.2012.750833

Bryant, F. B., & Satorra, A. (2012). Principles and practice of scaled difference chi-square testing. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(3), 372–398. https://doi.org/10.1080/10705511.2012.687671

Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory, 6*(3), 203-242. https://doi.org/10.1111/j.1468-2885.1996.tb00127.x

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance*, *18*(2), 123-149. https://doi.org/10.1207/s15327043hup1802_2

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Erlbaum Publications.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297-334. https://doi.org/10.1007/BF02310555

Cronbach, L. J. (1955). Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, *52*(3), 177-193. http://dx.doi.org.ezproxy.uct.ac.za/10.1037/h0044919

Cunningham J. B., & McCrum-Gardner, E. (2007). Power, effect and sample size using GPower: Practical issues for researchers and members of research ethics committees. *Evidence Based Midwifery, 5*(4), *132-136*. https://link.gale.com/apps/doc/A172050741/HRCA?u=lasiera&sid=googleScholar&xid=6acd735c

De Kock, F. S. (2015, December, 17). *Individual differences in judgment accuracy in personnel selection: What makes the 'good judge'?* (Doctoral dissertation, Erasmus University Rotterdam) http://hdl.handle.net/1765/79493

De Kock, F. S., Lievens, F., & Born, M. P. (2015). An in-depth look at dispositional reasoning and interviewer accuracy. *Human Performance, 28*(3), 199-221. https://doi.org/10.1080/08959285.2015.1021046

De Kock, F. S., Lievens, F., & Born, M. P. (2017). A closer look at the measurement of dispositional reasoning: Dimensionality and invariance across assessor groups. *International Journal of Selection and Assessment, 25*(3), 240-252. https://doi.org/10.1111/ijsa.12176

De Kock, F. S., Lievens, F., & Born, M. P. (2020). The profile of the 'good judge' in HRM: A systematic review and agenda for future research. *Human Resource Management Review, 30*(2), 1-21. https://doi.org/10.1016/j.hrmr.2018.09.003

De Vries, A., De Vries, R. E., & Born, M. P. (2011). Broad versus narrow traits: Conscientiousness and honesty-humility as predictors of academic criteria. *European Journal of Personality, 25*(5), 336-348. https://doi-org/10.1002/per.795

Ellis, A. P. J., West, B. J., Ryan, A. M., & DeShon, R. P. (2002). The use of impression management tactics in structured interviews: A function of question type? *Journal of Applied Psychology*, *87*(6), 1200-1208. https://doi.org/10.1037/0021-9010.87.6.1200

Erfanian, F., Latifnejad Roudsari, R., Heydari, A., & Noghani Dokht Bahmani, M. (2020). A narrative on using vignettes: Its advantages and drawbacks. *Journal of Midwifery and Reproductive Health*, *8*(2), 2134-2145. https://doi.org/10.22038/jmrh.2020.41650.1472

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics,* 5*(1),* 1-4. https://doi.org/10.11648/j.ajtas.20160501.11

Farrell, A. M., & Rudd, J. M. (2009). *Factor analysis and discriminant validity: A brief review of some practical issues.* Anzmac.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical

power analysis program for the social behavioral and biomedical sciences. *Behavior Research Methods, 39,* 175-191. https://doi.org/10.3758/BF03193146

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage Publications.

Funder, D. C. (1995). Funder, D. C. (1995). On the accuracy of personality judgement: A realistic approach. *Psychological Review, 102*(4), 652-670. http://dx.doi.org.ezproxy.uct.ac.za/10.1037/0033-295X.102.4.652

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science, 21*(3), 177-182. https://doi.org/10.1177/0963721412445309

Funder, D. C., & Colvin, R. (1997). Congruence of others' and self-judgements of personality. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 617-647). Orlando, FL: Academic Press. https://doi.org/10.1016/B978-012134645-4/50025-1

Furr, R. M. (2008). A Framework for Profile Similarity: Integrating Similarity, Normativeness and Distinctiveness. *Journal of Personality, 76*(5), (1267-1316). https://doi.org/10.1111/j.1467-6494.2008.00521.x

Furr, R.M., & Wood, D. (2013). On the similarity between exchangeable profiles: A psychometric model, analytic strategy, and empirical illustration. *Journal of Research in Personality, 47*(3), (233-247). https://doi.org/10.1016/j.jrp.2013.01.009

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment, 4*(1), 26-42. https://doi.org/10.1037/1040-3590.4.1.26

Hall, J. A., Andrzejewski, S. A., & Yopchick, J. E (2009). Psychosocial correlates of interpersonal sensitivity: A meta-analysis. *Journal of Nonverbal Behavior, 33*(3), 149-180. https://doi.org/10.1007/s10919-009-0070-5

Hall, J. A., Back, M. D., Nestler, S., Frauendorfer, D., Schmid Mast, M., & Ruben, M. A. (2018). How do different ways of measuring individual differences in zero-acquaintance personality judgment accuracy correlate with each other? *Journal of Personality*. https://doi.org/10.1111/jopy.12307

Hall, J. A., Goh, J. X., Mast, M. S., & Hagedorn, C. (2016). Hall, J. A., Goh, J. X., Schmid Mast, M., & Hagedorn, C. (2016). Individual differences in accurately judging personality from text. *Journal of Personality, 84*(4), 433-455. https:doi.org/10.1111/jopy.12170

Hooper, D., Coughlan, J., & Mullen, M. (2008, September). Evaluating model fit: a synthesis of the structural equation modelling literature. *In 7th European Conference on research methodology for business and management studies* (pp. 195-200).

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary vs. secondary. *Encyclopedia of Social Measurement, 1,* 593-599.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. https://doi.org/10.1037/1082-989X.3.4.424

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Huffcutt, A. I., Weekley, J. A., Wiesner, W. H., Groot, T. G., & Jones, C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology, 54*(3), 619-644. https://doi.org/10.1111/j.1744-6570.2001.tb00225.x

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6-23. https://doi.org/10.1037/a0014694

Jackson, D. N. (1972). A model for inferential accuracy. *Canadian Psychologist/Psychologie Canadienne, 13*(3), 185-195. https://doi.org/10.1037/h0082183

Keller, S., & Meaney, M. (2017, November, 24). *Attracting and retaining the right talent*. McKinsey. https://www.mckinsey.com/business-functions/organization/our-insights/attracting-and-retaining-the-right-talent

Kenny, D. A., & Albright, L. (1987). Accuracy in interpersonal perception: A social relations analysis. *Psychological Bulletin*, *102*(3), 390-402. http://dx.doi.org.ezproxy.uct.ac.za/10.1037/0033-2909.102.3.390

Kinicki, A. J., Hom, P. W., Trost, M. R., & Wade, K. J. (1995). Effects of category prototypes on performance-rating accuracy. *Journal of Applied Psychology, 80*(3), 354-370. https://doi.org/10.1037/0021-9010.80.3.354

Kristoff-Brown, A., Barrick, M. R., & Franke, M. (2002). Applicant impression management: Dispositional influences and consequences for recruiter perceptions of fit and similarity. *Journal of Management*, *28*(1), 27-46. https://doi.org/10.1177/014920630202800103

Krzyzaniak, S. L., Colman, D. E., Letzring, T. D., McDonald, J. S., & Biesanz, J. C. (2019). The effect of information quantity on distinctive accuracy and normativity of personality trait judgments. *European Journal of Personality, 33*(2), 197-213. https://doi.org/10.1002/per.2196

Langer, M., König, C. J., & Krause, K. (2017). Examining digital interviews for personnel

selection: Applicant reactions and interviewer ratings. *International Journal of Selection & Assessment, 25*(4), 371–382. https://doi.org/10.1111/ijsa.12191

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of research in personality*, *42*(4), 914-932. https://doi.org/10.1016/j.jrp.2007.12.003

Li, C-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods, 48*(3), 936-949. https://doi.org/10.3758/s13428-015-0619-7

Lievens, F., & Peeters, H. (2008). Interviewers' sensitivity to impression management tactics in structured interviews. *European Journal of Psychological Assessment*, *24*(3), 174-180. http://dx.doi.org/10.1027/1015-5759.24.3.174

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor anlaysis. *Psychological Methods, 4*(1), 84-99. https://doi.org/10.1037/1082-989X.4.1.84

Marsden, T. (2016). What is the cost of attrition? *Strategic HR Review, 15*(4), 189-190. https://doi.org/10.1108/SHR-05-2016-0039

Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology, 64*(1), 53-87. https://doi.org/10.1111/j.1744-6570.2010.01202.x

Melchers, K. G., Roulin, N., & Buehl, A.-K. (2020). A review of applicant faking in selection interviews. *International Journal of Selection and Assessment, 28*(2), 123-142. https:doi.org/10.1111/ijsa.12280

Miller, R. L., & Brewer, J. D. (Eds.). (2003). T*he AZ of social research: a dictionary of key social science research concepts*. Sage.

Murphy, K. R., & Balzer, W. K. (1989). Rater errors and rating accuracy. *Journal of Applied Psychology, 74*(4), 619-624 https://doi.org/10.1037/0021-9010.74.4.619

Murphy, K.R., & Cleveland, J.N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Sage Publications, Inc.

Neuman, W.L. (2000). *Social Research Methods: Qualitative and Quantitative Approaches 4th ed*. Allyn & Bacon: A Pearson Education Company https://doi.org/10.2307/3211488

Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS version 15*. McGraw Hill

Peeters, H., & Lievens, F. (2006). Verbal and nonverbal impression management tactics in behavior description and situational interviews. *International Journal of Selection and Assessment, 14*(3), 206-222. https://doi.org/10.1111/j.1468-2389.2006.00348.x

Pieterse S. (2016). *Detecting Deceptive Impression Management in Interviews: The Role of Interviewers' Individual Differences*. (Master's Thesis, University of Cape Town).

Powell, D. M. (2008). *Assessing personality in the employment interview: The impact of rater training and individual differences on rating accuracy* (Doctoral dissertation, Faculty of Graduate Studies, University of Western Ontario). ProQuest Dissertations. https://central.bac-lac.gc.ca/.item?id=NR39315&op=pdf&app=Library&oclc_number=665191410

Powell, D. M., & Bourdage, J. S. (2016). The detection of personality traits in employment interviews: Can "good judges" be trained? *Personality and Individual Differences, 94*, 194-199. https://doi.org/10.1016/j.paid.2016.01.009

Powell, D. M., & Goffin, R. D. (2009). Assessing personality in the employment interview: The impact of training on rater accuracy. *Human Performance, 22*(5), 450-465. https://doi.org/10.1080/08959280903248450

Rammstedt B., & John, O. P. (2007). Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of Research in Personality, 41*(1), 203-212. https://doi.org/10.1016/j.jrp.2006.02.001

Roch, S. G., Woehr, D. J., Mishra, V., & Kieszczynska, U. (2012). Rater training revisited: An updated meta-analytic review of frame-of-reference training. *Journal of Occupational and Organizational Psychology*, *85*(2), 370-395. https://doi.org/10.1111/j.2044-8325.2011.02045.x

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Roulin, N. (2016). Individual Differences Predicting Impression Management Detection in job interviews. *Personnel Assessment and Decisions*, *2*(1), 1-11. https://doi.org/10.25035/pad.2016.001

Roulin, N., Bangerter, A., & Levashina, J. (2014). Interviewers' perceptions of impression management in employment interviews. *Journal of Managerial Psychology, 29*(2), 141-163. https://doi.org/10.1108/JMP-10-2012-0295

Roulin, N., Bangerter, A., & Levashina, J. (2015). Honest and deceptive impression management in the employment interview: Can it be detected and how does it impact evaluations? *Personnel Psychology, 68*(2), 395-444. https://doi.org/10.1111/peps.12079

Roulin, N., & Powell, D. M. (2018). Identifying applicant faking in interviews. *Journal of Personnel Psychology, 17*(3), 143-154. https://doi.org/10.1027/1866-5888/a000207

RStudio Team (2020). RStudio: Integrated Development for R. http://www.rstudio.com/

Ryu, E., Couper, M. P., Marans, R. W. (2005). Survey incentives: cash vs. in-kind; face-to-face vs. mail; response rate vs. nonresponse error. *International Journal of Public Opinion Research, 18*(1), 89-106. https://doi.org/10.1093/ijpor/edh089

Scheepers, J.M. (2004). Overcoming the effects of differential skewness of test items in scale construction. *SA Journal of Industrial Psychology, 30*(4), 27-43. https://doi.org/10.10520/EJC89026

Schmid Mast, M., Bangerter, A., Bulliard, C., & Aerni, G. (2011). How accurate are recruiters' first impressions of applicants in employment interviews? *International Journal of Selection & Assessment, 19*(2), 198-208. https://doi.org/10.1111/j.1468-2389.2011.00547.x

Schneider, L., Powell, D. M., & Roulin, N. (2015). Cues to deception in the employment interview. *International Journal of Selection and Assessment, 23*(2), 182-190. https://doi.org/10.1111/ijsa.12106

Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology, 68*(2), 247-260. https://doi.org/10.1037/0022-3514.68.2.247

Smither, J. W., Barry, S. R., & Reilly, R. R. (1989). An investigation of the validity of expert true score estimates in appraisal research. *Journal of Applied Psychology, 74*(1), 143-151. https://doi.org/10.1037/0021-9010.74.1.143

Sulsky, L. M., & Balzer, W. K. (1988). Meaning and measurement of performance rating accuracy: Some methodological and theoretical concerns. *Journal of Applied Psychology, 73*(3), 497-506. https://doi.org/10.1037/0021- 9010.73.3.497

Sulsky, L. M., & Day, D. V. (1994). Effects of frame-of-reference training on rater accuracy under alternative time delays. *Journal of Applied Psychology, 79*(4), 535-543. https://doi.org/10.1037/0021-9010.79.4.535

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics.* Allyn and Bacon

Tziner, A., Murphy, K., Cleveland, J. N., Yavo, A., & Hayoon, E. (2008). A new old question: Do contextual factors relate to rating behaviour: An investigation with peer evaluations. *International Journal of Selection and Assessment, 16*(1), 59-67. https://doi.org/10.1111/j.1468-2389.2008.00409.x

Wade, M. R., & Parent, M. (2001). Relationships between job skills and performance: A study of webmasters. *Journal of Management Information Systems, 18*(3), 71-96. https://doi.org/10.1080/07421222.2002.11045694

Worthington, R.L., & Whittaker, T.A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counselling Psychologist, 34(6)*, 806-838. https://doi.org/10.1177/0011000006288127

Appendix A

Figure A1

Ethical clearance for original study (Pieterse, 2016)

UNIVERSITY OF CAPE TOWN

Faculty of Commerce
Ethics in Research Committee
Courier: Room 2.21 Leslie Commerce Building Upper Campus University of Cape Town
Post: University of Cape Town • Private Bag • Rondebosch 7701
Email: Irwin.brown@uct.ac.za
Telephone: +27 21 650-2311
Fax No.: +27 21 689-7570

23 June, 2015

Susan Pieterse

School of Management Studies

Project title:
Detecting Deceptive Impression Management in Interviews: The Role of Interviewers' Individual Differences

Proposal ref: 23-06-2015 Pieterse Susan

Dear Researcher,

This letter serves to confirm that this project as described in your submitted protocol has been approved.

Please note that if you make any substantial change in your research procedure that could affect the experiences of the participants, you must submit a revised protocol to the Committee for approval.

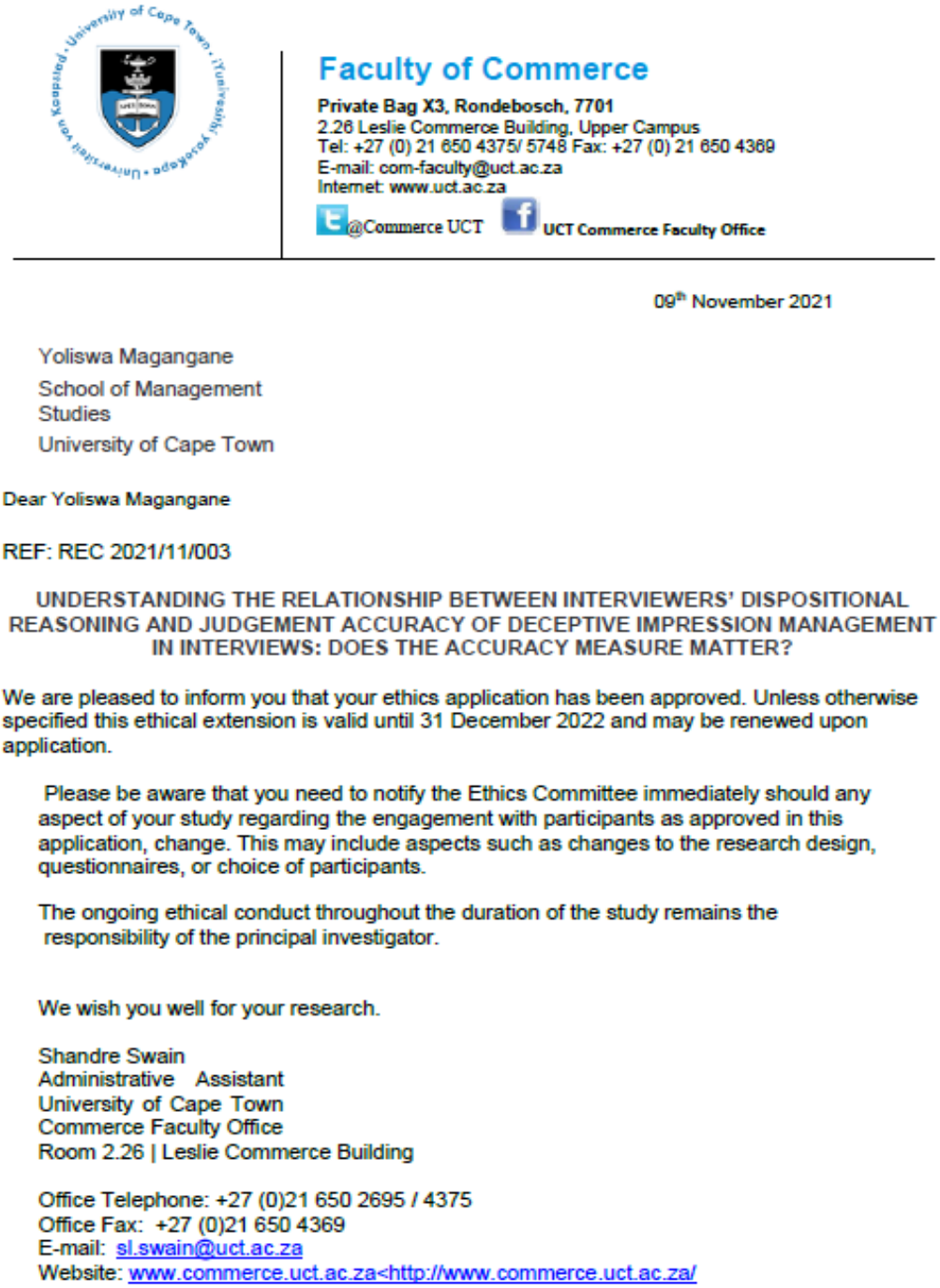
Regards,

Professor Michael Kyobe

Commerce Faculty Ethics in Research Committee

Figure A2

Ethical clearance for present study



Faculty of Commerce

Private Bag X3, Rondebosch, 7701
2.26 Leslie Commerce Building, Upper Campus
Tel: +27 (0) 21 650 4375/ 5748 Fax: +27 (0) 21 650 4369
E-mail: com-faculty@uct.ac.za
Internet: www.uct.ac.za
@Commerce UCT **UCT Commerce Faculty Office**

09th November 2021

Yoliswa Magangane
School of Management Studies
University of Cape Town

Dear Yoliswa Magangane

REF: REC 2021/11/003

UNDERSTANDING THE RELATIONSHIP BETWEEN INTERVIEWERS' DISPOSITIONAL REASONING AND JUDGEMENT ACCURACY OF DECEPTIVE IMPRESSION MANAGEMENT IN INTERVIEWS: DOES THE ACCURACY MEASURE MATTER?

We are pleased to inform you that your ethics application has been approved. Unless otherwise specified this ethical extension is valid until 31 December 2022 and may be renewed upon application.

Please be aware that you need to notify the Ethics Committee immediately should any aspect of your study regarding the engagement with participants as approved in this application, change. This may include aspects such as changes to the research design, questionnaires, or choice of participants.

The ongoing ethical conduct throughout the duration of the study remains the responsibility of the principal investigator.

We wish you well for your research.

Shandre Swain
Administrative Assistant
University of Cape Town
Commerce Faculty Office
Room 2.26 | Leslie Commerce Building

Office Telephone: +27 (0)21 650 2695 / 4375
Office Fax: +27 (0)21 650 4369
E-mail: sl.swain@uct.ac.za
Website: www.commerce.uct.ac.za<http://www.commerce.uct.ac.za/

"Our Mission is to be an outstanding teaching and research university, educating for life and addressing the challenges facing our society."

Appendix B

Deceptive Impression Management Task

Bending the truth and telling white lies to increase the likelihood of getting a job is very common during employment interviews. This is known as deceptive impression management or interview faking.

My research investigates whether **you can spot the fakers** in this 30-minute experiment.

Incentive

Use your skills to win some easy cash: **Cash prizes of R1250, R1000, and R750** will be awarded to the top three ratings. There will also be **two additional lucky draw** cash prizes of **R500** each. To qualify for these, please enter your email address once you have completed the survey.

This study has been approved by the Commerce Faculty's Ethics in Research Committee, and your participation is voluntary. You can choose to withdraw at any stage during the research. Your responses will be used for research purposes only.

If you have any questions about the research, please contact me, Suki Pieterse, on ptrsus002@myuct.ac.za.

Instructions

Scenario: Three candidates applied for a management trainee position. These interviews were transcribed, and the verbal and non-verbal cues were recorded.

All three candidates were asked **the same questions** relating to **two performance dimensions** important for future success on the job. To elicit more information on the performance dimension, the interviewer asked a prompting question as well. These were:

Performance Dimension	**Definition**	**Question**
Perseverance	The tendency to persist in the face of obstacles and work activities.	"Imagine you're finding the first months at your new job very difficult. The tasks you're assigned are very demanding and you think your boss isn't entirely satisfied with your work. Please describe briefly how you would behave in this situation." + prompt
Organisational Behaviour	Scheduling work and actives, coordinating the work and activities of others, and the ordering of information.	"Imagine you return to work after a holiday. You discover a stack of unopened letters on your desk and over 100 unread emails in your email inbox. There are already some important meetings planned for today. These meetings will take about an hour each. Your boss also wants to speak to you urgently about an issue. He has sent you details about it via email. Please describe briefly how you would behave in this situation." + prompt

Your Task

Put yourself in the position of a recruiter, and please try your best, as you will conduct interviews in your career to come. You will be asked to rate each candidate on the dimensions above. Don't worry, I will remind you of the definitions again.

The transcripts have also been structured so you can easily read the candidates' responses without rereading the questions.

Please note candidates' actions (non-verbal behaviours) are recorded in *italics.*

Now the real fun starts: Detecting Deception

Scenario: The candidates displayed deceptive impression management (DIM) cues to varying degrees in their interviews. There are 4 types of DIM tactics.

DIM Tactic	Definition	Example
Slight image creation	To embellish and tailor skills, capabilities, and characteristics to enhance fit to the job.	"I am good at working with people, and I am a team player"
Extensive image creation	To construct, invent or borrow experiences and accomplishments in response to a question.	"I perform at my best when working under pressure. I have only received firsts this year whilst holding down two part time jobs"
Deceptive Ingratiation	To express insincere beliefs and values to conform to those held by the interviewer/organisation.	"As you said earlier, teamwork is important to the company, and I think you get more value when you collaborate"
Image protection	To mask or omit negative experiences.	"I know I don't have much work experience, but I wanted to focus on my studies"

Your task: You will be asked to rate how frequently you observed the types of DIM tactics. Don't worry, I will remind you of the definitions again.

There is no time limit, so take as long as you may need.

Get ready to evaluate some candidates!

Candidate 1

Q1 Interviewer: "Imagine you're finding the first months at your new job very difficult. The tasks you're assigned are very demanding and you think your boss isn't entirely satisfied with your work. Please describe briefly how you would behave in this situation.

> **Candidate 1**: I would obviously speak to my supervisor and colleagues about the tough things I am experiencing, and really try to get insight into whether it is normal to feel that way *(nodding)*. By voicing my concerns, showing that I care, and engaging with my peers I will see others deal with the demands and learn even more effective ways of dealing with the pressure *(sustained eye contact throughout speaking)*.

Q1 prompt: It is a good idea to engage with others about your difficulties *(candidate nodding)*, but what if you feel your performance is below standard?

> **Candidate 1**: Uhm *(readjusting in seat)*, I would definitely try speak to my immediate supervisor and ask for feedback on my performance, specifically ask how he or she recommend I improve. It is a graduate job, and coming from university into the working world I have no experience, so asking about all of the opportunities on offer to learn more and improve is very important for me to make sure I perform very *(emphasis and gesturing hands, speaking quickly)* well in the future.

Q2 Interviewer: I can see that you value feedback to improve your performance, onto the next question. You return to work after a holiday. You discover a stack of unopened letters on your desk and over 100 unread emails in your email inbox. There are already some important meetings planned for today. These meetings will take about an hour each. Your boss also wants to speak to you urgently about an issue. He has sent you details about it via email. Please describe briefly how you would behave in this situation.

> **Candidate 1**: Oh, that is quite a lot going on *(pause)*. But I think I would go about it one step at a time *(gesturing with hands)*. I would carefully read the important message from my supervisor first to see what it is about. I would then scan over all of the other emails and letter to see which are important and require my immediate attention. I am generally very *(emphasis)* organised.

Q2 prompt: Okay, I can see that you try and structure your work, but how would you deal with all of the meetings scheduled, especially with your supervisor?

> **Candidate 1:** I would try prepare for the meetings as best I can, with my focus being on the meeting with my boss. This is my first day back and I know that I would have to catch up on everything that has happened, but again, I will talk with my colleagues and try to see about the most important things I have missed, because teamwork is a big part of the job *(no silences)*. I am also not scared of asking for help *(eye contact throughout answer)*.

Key for Vignettes:

Slight Image Creation

Extensive Image Creation

Ingratiation

Image Protection

Appendix C

Revised Interpersonal Judgement Inventory

Trait induction

Circle the letter that corresponds most to the trait you think is represented by the word:

	Trait				
Behaviour	Emotional stability	Extraversion	Openness	Agreeableness	Conscientious-ness
Sloppy					X
Irritable	X				

Trait extrapolation

For example, one item depicted 'John' as "John's co-workers all describe him as efficient, thorough, and persistent. MOST likely John also:". Next, respondents had to choose the best answer from the following options:

A. feels the need to be around lot of people,

B. has a great deal of sympathy for those less fortunate,

C. **doesn't often give in to his impulses**,

D. enjoys fantasizing and daydreaming.

Clearly, only option (C), 'doesn't often give in to his impulses' relates to the focal trait (conscientiousness) in the original person description.

Trait contextualization

For example, one item stated "Which of the following situations is most relevant to the trait of sociability?". Then, respondents had to select the most appropriate answer from three options (correct answer in bold):

A. A team member upon whom you rely allows her unanswered emails to accumulate and frustrate your co-workers in the process.

B. You notice that the time has just turned 1 p.m. (which is your lunchtime) and you see a few of your colleagues walking to the tea room.

C. You see that you colleague has been working non-stop since the morning.