When More is not Better: Understanding the Potential Nonlinear Relationship Between Intelligence and Rating Accuracy

A dissertation submitted in partial fulfilment of the requirements for the award of the Degree of Master of Commerce in Organisational Psychology

Supervisor: Prof Francois de Kock

Marizanne Schade
SCHMAR134
Faculty of Commerce
University of Cape Town
Master's in Industrial and Organisational Psychology

COMPULSORY DECLARATION:
This work has not been previously submitted in whole, or in part, for the award of any degree. 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, cited, and referenced.

Signature: Signed by candidate Date: 14 February 2022

Acknowledgements

I am eternally grateful to my supervisor, Prof Francois de Kock, for all his guidance and knowledge on the topic. I also thank my colleagues, peers and friends for their unwavering support and motivation. Lastly, I dedicate this body of work to my parents for their unconditional patience and encouragement that made this possible.

Abstract

Employers rely on judges or raters to accurately rate the potential or performance of candidates through interviews or assessment centre evaluations. As the judgment process places heavy demands on information processing, cognitive ability (of raters) is important to detect and interpret behavioural cues presented by those being rated. A consistent empirical finding is that intelligence is the strongest predictor of rating accuracy, but prior research has largely been based on linear models. However, researchers have yet to investigate whether these variables could be nonlinearly related. By studying nonlinear models in judgment and accuracy, we can not only deepen our understanding of the 'good judge' in HRM, but we may further enhance methods to select and train raters in applied practice. This secondary research study re-analysed data from a prior published study to evaluate the relationship between rater intelligence and accuracy of interview ratings provided by 146 South African managers. The predictiveness of an ordinary least squares (OLS) linear regression model was compared to two nonlinear models (quadratic and cubic) to determine which statistical approach explained the most variance in rating accuracy scores. Findings provided further support of a linear relationship between intelligence and rating accuracy suggesting no quadratic or cubic interactions. Judges, therefore, produced more accurate ratings at higher levels of intelligence. Possible explanations of the findings include the sample size and task complexity. Study limitations and recommendations for future research are discussed in detail.

Keywords: intelligence, accuracy, nonlinear, interview, judgement, rating

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

In the field of Human Resource Management (HRM), ratings play an integral role in the effective functioning of an organisation. It is standard practice for interviewers, recruiters, and assessment centre assessors to produce evaluations of interview candidates' potential or for line managers to conduct performance appraisals for direct reports (Christiansen et al., 2005; De Kock et al., 2020). All of these circumstances require judges or raters[1] to observe individuals (also known as targets) and synthesize information obtained through questions or tasks (Kuncel & Highhouse, 2011). Raters then use this information to make inferences about an individual's underlying traits based on the behaviours or characteristics they have identified to produce ratings (Funder, 2012). These ratings in turn determine whether an employee is hired, promoted or positively appraised (Guion & Gibson, 1988; Sulsky & Keown, 1998).

Unsurprisingly, the accuracy of ratings is of concern to employers when they are used to inform selection and performance management decisions (Christiansen et al., 2005). For example, more employers have turned to the recruitment and selection process to secure their competitive advantage by hiring and retaining top talent (Campbell et al., 2012). This has also meant committing more financial and labour resources over time. A serious threat to the success of these investments exists when assessors or judges with poor rating abilities administer and oversee the selection and promotion process. Poor judges are less likely to hire and promote deserving employees which could result in increased turnover rates and profit losses for employers (Randall & Sharples, 2012). For this reason, there is a growing body of research focused on how to achieve and measure accurate ratings. (Davis & Kraus, 1997; Funder, 2012; Lippa & Dietz, 2000)

A Brief History of Rating Accuracy in Personnel Selection

Research into how individuals rate others dates as far back as the 1930s (Funder, 1995). Early challenges were related to inconsistent empirical findings and developing reliable accuracy measurements (Funder, 1995). As a result, many researchers shifted focus away from studying accuracy towards inferential biases and errors in judgment in the 1950s (Christiansen et al., 2005; Funder, 2012). After a period of dormancy, the topic received renewed interest in the 1980s when more sophisticated accuracy measures were introduced to overcome some of the initial criterion issues (Funder, 1995; Lippa & Dietz, 2000). Recent researchers have developed an interest in understanding how the characteristics of judges

[1] In this study, the terms 'judge', 'rater' and 'assessor' are used interchangeably.

influence the rating process and determine the accuracy of ratings (De Kock et al., 2020; Letzring, 2008; Lippa & Dietz, 2000). Consequently, most of the current studies aim to identify the determinants of a good judge to assist employers with accurate personnel selection and retention decisions (Christiansen et al., 2005; Colvin & Bundick, 2001; De Kock et al., 2020; Letzring, 2008).

Determinants of the Good Judge

In an attempt to identify specific characteristics and abilities associated with good judges, researchers have considered an extensive list of factors (De Kock et al., 2020). A systematic review of over 80 years' worth of empirical findings by De Kock et al. (2020) found that most studies (around one-third) in the HRM field examined whether Big 5 personality traits were associated with higher rating accuracy (Borman, 1979; Christiansen et al., 2005; Davis & Kraus, 1997; Powell & Goffin, 2009). Researchers have also assessed the role of cognitive abilities in assisting raters with forming judgements (Borman, 1979; Lippa & Dietz, 2000) with dispositional reasoning attracting increased attention over the last decade (Christiansen et al., 2005; De Kock et al., 2015; Janovics, 2003). Additionally, the motivation of raters to accurately assess targets has been considered (Salvemini et al., 1993; Wood & Marshall, 2008) as well as the influence of demographics like gender (Ambady & Rosenthal, 1992), age (Borman, 1979) and rating experience (Borman & Hallam, 1991; Wood & Marshall, 2008).

The Role of General Intelligence in Predicting Rating Accuracy

The systematic review by De Kock et al. (2020) further reported that general intelligence remains the strongest predictor of judgement accuracy compared to any other individual differences. This is not surprising since the process of accurately assessing others is considered extraordinarily complex and cognitively demanding (Christiansen et al., 2005). The reasoning for this is best explained through Funder's (1995) Realistic Accuracy Model which posits that raters go through four distinct and consecutive stages to reach an accurate judgment about a target: relevance, availability, detection and utilization. Strong cognitive abilities become instrumental in the two final stages which require raters to process, integrate and recall information about targets to reach an accurate assessment (Letzring, 2008). Intelligent raters are therefore considered better equipped with the skills and abilities required for good task performance (Hauenstein & Alexander, 1991).

When assessing the ideal level of intelligence required for accurate raters, there is a lack of consensus among empirical findings. There are a number of studies that have found positive correlations or linear relationships between IQ[2] and rating accuracy (Borman, 1979; Borman & Hallam, 1991; Christiansen et al., 2005; Lippa & Dietz, 2000) suggesting that more intelligent raters will always be more accurate. But there are also studies that have reported negative correlations (Hauenstein & Alexander, 1991; Smither & Reilly, 1987) or no relation whatsoever (Letzring et al., 2006; Powell, 2007). Most notably, in a study conducted by Smither and Reilly (1987), they found that participants with the highest intelligence scores were less accurate than those with moderate intelligence scores. More intelligent judges, therefore, do not always produce the most accurate ratings as first assumed. This points to a gap in the literature examining how rating accuracy is influenced at extreme levels of intelligence. More specifically, assessing whether this relationship could be more complex than a linear interaction where IQ and accuracy move in the same direction in the same increments.

Theoretical Support for Nonlinear Relationships

Nonlinear relationships are best characterised by bends in the curvature of a line where the slope changes (Cohen et al., 2003). Changes in the independent variable could correspond with increases or decreases in the dependent variable at varying rates (Bobko, 2001). Based on a review of theoretical and empirical findings in the HRM field, there is enough evidence to suggest the presence of two specific nonlinear interactions: quadratic and cubic effects (Hauenstein & Alexander, 1991; Smither & Reilly, 1987).

Exploring a Quadratic Relationship

Reasons for a possible nonlinear interaction between IQ and rating accuracy can best be described by the 'too-much-of-a-good-thing' effect popularized by Pierce and Aguinis (2013). This meta-theoretical principle posits that positive antecedents can lead to positive outcomes but may become harmful or less impactful after a certain threshold (Brown et al., 2020). It has been useful in explaining other nonlinear findings in the person-perception field for example between impression management tactics in interview performance (Robie et al., 2020) and self-presentation in job interviews (Baron, 1986).

Evidence suggests that more intelligent individuals struggle to communicate with others given their advanced comprehension and reasoning abilities (Antonakis et al., 2017). For this reason, they might grapple more than moderately intelligent interviewers with eliciting information or behavioural cues from targets during interviews (Christiansen et al.,

[2] The terms 'IQ' and 'intelligence' are used interchangeably in the present study to represent general mental abilities and cognitive function.

2005). This interaction is best represented by an inverted U-shape or quadratic function depicted in Figure 1.

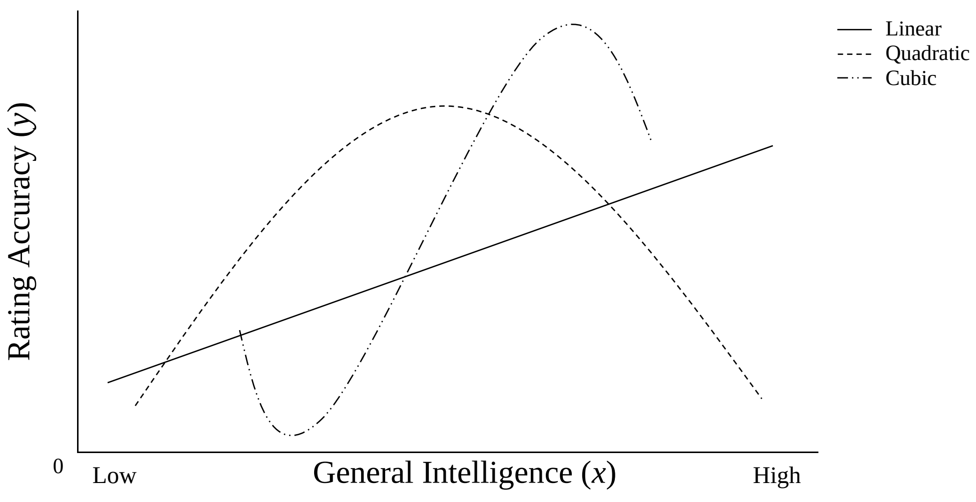
Exploring a Cubic Relationship

Problem-solving skills are also required when accurately assessing others and Antonakis et al. (2017) found that an increase in IQ points after a certain threshold has a smaller impact on these abilities. This is also true for IQ levels below a certain threshold. This means that at the extreme ends of IQ, increased intelligence has a smaller impact on rating accuracy, whereas increased IQ is more impactful on accuracy for moderately intelligent individuals. This relationship is also represented as an S-shaped function in Figure 1 with two bends in the curve.

These factors combined with contradictory empirical findings make a strong case for revisiting how cognitive abilities interact with rating accuracy and proposing a possible nonlinear relationship.

Figure 1

Proposed Linear, Quadratic and Cubic Relationships Between IQ and Accuracy



The Present Study

To address the lack of conclusive evidence in the current literature, the present study aimed to examine how increases in IQ relate to judgement accuracy among raters. More specifically, it compares the variance in rating accuracy produced by linear and nonlinear models to answer the following research question: *Can a nonlinear model better predict the relationship between intelligence and rating accuracy than a linear model?*

Research Objectives

Data was sourced from a study conducted by De Kock et al. (2015) with 146 South African managers and re-analysed in accordance with the two study objectives.

- Comparing the predictiveness of a linear model to nonlinear models (quadratic and cubic) in order to determine which statistical approach explains the most variance in rating accuracy scores
- Adding to the dearth of research currently available on this topic (De Kock et al., 2020)

Proposed Value of Study

From a theoretical perspective, findings are expected to enhance our current understanding of the variables and provide further insights that can guide future researchers. Given that the majority of previous studies employed student participants as judges, the previously reported effect sizes could underestimate the relationship between intelligence and accuracy in practice (Lippa & Dietz, 2000). Practically, this would greatly benefit employers who screen raters and could do so with the ideal level of intelligence required to confidently produce accurate ratings (De Kock et al., 2020).

Overview of Dissertation

Chapter two of this dissertation outlines the existing body of research that covers both general intelligence and rating accuracy constructs and summarizes previous studies supporting nonlinear relationships between the variables. Chapter three details the methodology followed to conduct the study and chapter four presents the results after conducting the analyses. Lastly, chapter five discusses the study findings and further implications for theory and practice.

Chapter 2: Literature Review

This chapter examines the available literature on general intelligence (IQ) and rating accuracy within the scope of the present study. It includes conceptualizations and operationalizations of the study variables, as well as an examination of historical claims of a linear relationship. It provides evidence to support nonlinear interactions between IQ and rating accuracy to better explain the complex phenomenon. It also presents two hypotheses developed to answer the study's primary research question.

Defining General Intelligence

Intelligence research spans over a century and has been studied through a range of disciplines (Hernández-Orallo et al., 2014). As more research is conducted to understand this multifaceted construct, the conceptualization thereof has understandably changed over time. Generally, however, it is understood to represent an individual's intellectual and cognitive abilities. For the purpose of this study, intelligence was conceptualized as defined by Humphreys (1979, p. 115): "…the resultant of the processes of acquiring, storing in memory, retrieving, combining, comparing, and using in new contexts information and conceptual skills; it is an abstraction". The conceptualization also includes a more recent definition provided by Salgado (2017, p. 116) as "…the capacity of an individual to learn quickly and accurately a task, a subject matter or a skill, under optimal instructional conditions". Individuals with higher levels of intelligence can "…[solve] problems correctly, [make] rapid but sound decisions, [judge] situations accurately…use abstract reasoning, to acquire knowledge and to be able to use it in new contexts" (Salgado, 2017, p. 116). General intelligence is also often denoted by a range of interchangeable terms: general mental abilities, general cognitive function, IQ, and *g* factor (Borman, 1979; Salgado, 2017; Schmidt & Hunter, 1993).

Spearman (1904) was the first to suggest that individuals' mental abilities can be represented by a single factor he labelled the *g* factor. Spearman argued all cognitive abilities share a general factor that can be used as a predictive measure (Jensen, 1993). This general factor is present in all tests that measure intelligence, cognitive ability or related constructs even if it might not be suspected. Jensen (1993, p. 308) stated that "…the *g* factor is remarkably stable across different collections of mental tests, even collections of tests that bear hardly any superficial resemblance to one another". In a study of intelligence by Johnson et al. (2004), they found correlations above .90 between three distinct test batteries assessing various components of cognitive function supporting the idea that individuals possess a general mental ability that can be predictive of performance across a range of mental tasks. In response to Spearman's emphasis on a general factor, various other researchers proposed the existence of multiple factors pointing to a hierarchical structure (Embretson & McCollam, 2000): Cattell (1963), Horn and Cattell (1967), Thurstone (1938) and Guilford (1967). Most notably, Carroll (1993) proposed a three stratum hierarchy after conducting systematic exploratory factor analysis of over 460 cognitive ability datasets (McGrew, 2009). The *g* factor occupies the apex of the hierarchy, also known as stratum three, with each ability representing the variance in the abilities below. Stratum two includes broad abilities like visuospatial ability, fluid intelligence, crystallized intelligence and working memory (Woodley, 2010). Stratum one, the final stratum, accounts for narrow or more specialized abilities (Bickley et al., 1995).

The Link Between IQ and Accuracy

The link between intelligence and the ability to judge others is not new. Almost 85 years ago Allport (1937) noted that "Understanding people is largely a matter of perceiving relations between past and present activities, between expressive behavior and inner traits, between cause and effect, and intelligence is the ability to perceive just such relations as these" (p. 514). Decades later, researchers have found strong evidence to suggest that individuals assessing others require more advanced cognitive abilities to produce accurate assessments of targets (Davis & Kraus, 1997). The study of how accurately individuals can judge others gained prominence in the 1980s and is known in the person-perception field as rating or rater accuracy (Funder, 1999).

Defining Rating Accuracy

Rating accuracy is a measurement of the strength and type of relation between scores developed by raters and a corresponding set of 'true' scores (true reflections of a ratee's underlying traits) considered to be an accepted standard for comparison (Engelhard Jr, 1996; Murphy et al., 1982; Sulsky & Balzer, 1988). Rating accuracy received increased attention from the 1980s after a distinct shift was made from finding the 'perfect' rating system to finding what makes the 'perfect' rater (Sulsky & Keown, 1998). It became clear that the impact of a rater in the rating process was underestimated and more was needed to understand how a rater's performance can be improved or predicted.

Conceptualization of Rating Accuracy

Rating accuracy can be classified as either diagnostic or predictive based on the nature of the judgments performed by raters. A rater's ability to determine whether a target displays certain characteristics is termed diagnostic accuracy which can range from dispositional inferences (surface level assessments of whether a character trait is present or not) to more complex behavioural inferences (an understanding of what caused the observed behaviours) (Christiansen et al., 2005; De Kock et al., 2015). Predictive accuracy is the rater's ability to correctly predict a target's behaviour in the future based on diagnostic judgments or results from standardized tests. It is important to note that the focus of the current study was on inferences about behaviours rather than assessing the accuracy of predictions of future behaviour.

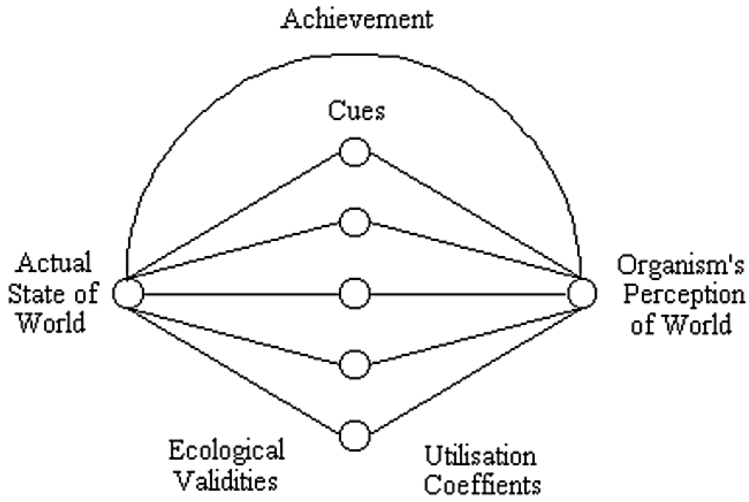
Rating accuracy can also be referred to as inferential accuracy, signal detection or rater validity, but should not be confused with the absence of rating errors (De Kock et al., 2015). Rating errors occur when a rater's assessment of the target individual is influenced by personal biases or heuristics like central tendency, leniency or restriction of range (Engelhard Jr, 1996; Kassim, 2011). Rating error measures have been shown to correlate with rating accuracy measures, but there is little to no empirical support that the former predicts the latter (Murphy & Balzer, 1989). The absence of rating errors, therefore, does not guarantee higher levels of rating accuracy (Funder, 1995).

To understand the cognitive processes involved in how raters develop accurate ratings, researchers have devised a variety of theoretical models (Funder, 2012). In the present study, two of the most influential frameworks were used as reference: the Lens Model and the Realistic Accuracy Model.

The Brunswik Lens Model. Brunswik's (1956) Lens Model theorized that individuals rely on cues to make inferences when a phenomenon of interest cannot be observed directly within their environment. Cues are defined as pieces of information that can be readily and easily accessed about the phenomenon (Miller & Kirlik, 2006). In Figure 2, the real environment is depicted on the left side and an organism or individual's perception of that environment is on the right. Brunswik argued that individuals cannot obtain perfect, objective descriptions of their environment and must instead rely on proximal stimuli or probabilistic cues for guidance (Vicente, 2003). For example, when an interviewer needs to assess whether a candidate is extroverted, he or she will rely on behavioural cues from the interviewee like willingness to engage in conversation and making direct eye contact (Powell, 2007). The observed cues also present the individual with varying degrees of utility (Deffenbacher & Hamm, 1972). Ecological validities measure the optimal relevance of a behavioural cue to the individual whereas utilization coefficients describe the actual weight placed on the cues by the rater (Vicente, 2003). According to the Lens Model, an accurate judgment is achieved when "…there is a correlation between the actual properties of a stimulus and the judgment of those properties" (Powell, 2007, p. 14).

Figure 2

Brunswik's (1956) Lens Model[3]



Funder's Realistic Accuracy Model. The Realistic Accuracy Model (RAM) was developed by Funder (Funder, 1995, 1999, 2012) and slightly departs from the Lens Model by placing equal importance on a target's ability to exhibit behavioural cues and the rater's perceptiveness of those cues. It posits that raters need to follow four interdependent stages (relevance, availability, detection and utilization) in order to formulate an accurate inference of a target's underlying traits (Powell & Goffin, 2009). The first two stages of the model take place in the social environment shared by the target and rater whereas the third and fourth stages entirely depend on the rater's abilities (Powell & Bourdage, 2016).

Relevance. For a rater to assess whether a target individual displays a predetermined trait, the target must provide the rater with some information relevant to that trait (Powell & Bourdage, 2016). For example, a rater can only assess whether an individual is friendly if they display characteristics or behaviours associated with friendliness like trying to actively engage in conversation with the rater and making strong eye contact (Funder, 2012).

Availability. The rater must have the opportunity to observe the target displaying the behavioural cue. If the rater does not share the context in which the behaviour is present, the rater cannot make an accurate assessment thereof (Powell & Goffin, 2009).

Detection. The rater must be observant and attentive throughout the interaction with the target individual to ensure that he or she can detect behavioural cues (Letzring, 2008).

[3] Reprinted from "Beyond the lens model and direct perception: Toward a broader ecological psychology", by K. Vincente, 2003, *Ecological Psychology*, 15(3), 241-267, p. 247. Copyright 2003 by Kim Vincente.

Any distractions or impairments on the rater's part will prevent accurate detection (Funder, 2012).

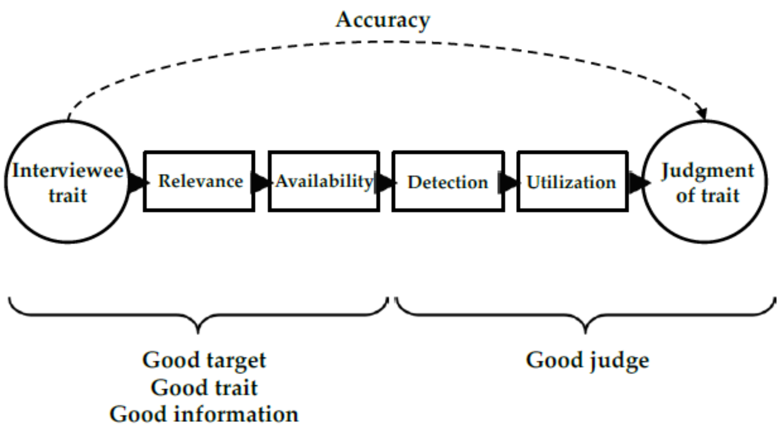
Utilization. The rater must utilize the behavioural cue to make an accurate inference about the target (Funder, 2012). For example, if the target smiles and attempts to be friendly, that should not mistakenly be interpreted as insincerity or sarcasm.

Raters must go through each of the four stages in order to make an accurate judgment (Letzring et al., 2006; Powell & Goffin, 2009). For example, if the target individual displays the relevant behavioural cue in the presence of the rater, but the rater is not observant, he or she will not be able to successfully detect it. This means that the rater cannot utilize the information and make an accurate judgment.

Good Targets, Traits, Information and Judges. Funder (1995) also argued that there are four factors that moderate the RAM judgement process. *Good targets* who exhibit clear behaviours that are representative of their underlying traits make it easier to judge (Chen et al., 2018). *Good traits*, like expressiveness, are easier to identify than more complex traits like deception (De Kock et al., 2015). *Good information* about the target provides the rater with richer insight. This information can be obtained in various ways including meaningful time spent with the target and better engagement with a target who feels comfortable (De Kock et al., 2015). *Good judges* possess characteristics and abilities that allow them to elicit better responses from targets and accurately interpret those cues (Powell & Goffin, 2009). Figure 3 illustrates the four stages that determine judgement accuracy as well as the impact of the moderators.

Figure 3

Funder's (1995) Realistic Accuracy Model[4]



[4] Reprinted from *Individual differences in judgment accuracy in personnel selection: What makes the 'good judge'?* (p. 20), by F. De Kock, 2015, Erasmus University Rotterdam. Copyright 2015 by Francois de Kock.

Operationalization of Rating Accuracy

The original study conducted by De Kock et al. (2015) employed two widely-used operationalizations of rating accuracy: Cronbach's (1955) four accuracy components and Borman's (1977) Differential Accuracy. The use of both measures allowed researchers to validate whether rating accuracy effects were stable across different operationalizations (Neumayer & Pluemper, 2020).

Cronbach's Four Accuracy Components. Cronbach's (1955) assertion that rating accuracy is best represented by measuring specific components rather than deriving a general value, was extremely influential in the person perception field. He developed four distinct indices to measure the difference between 'true scores' assigned to individuals' performance and scores determined by raters where lower score values indicate higher accuracy (Brooks & Brooks, 1990; Coleman et al., 2001). A description of each component is provided below and formulae can be found in Table 3.

The first measure, *Elevation* , represents the difference between an evaluator's average scores for related dimensions compared to the average true scores developed by expert raters (Coleman et al., 2001). In practice, this would indicate, for example, how accurately different leaders would rate the same employee's overall performance (Murphy et al., 1982). The second measure, *Differential Elevation*, is an average of the mean ratings assigned to each ratee across all the performance dimensions (Borman, 1977; Brooks & Brooks, 1990). Raters whose mean ratings deviate less from the true mean ratings are considered more accurate e.g., a hiring manager who can accurately rank potential candidates' overall performance to assess job fit. *Stereotype Accuracy* is a measure of a rater's ability to accurately predict the average rating for a ratee across each performance dimension (Brooks & Brooks, 1990). It might impact how different leaders rate the performance of the same team to select the best performing one. And lastly, *Differential Accuracy (DA)*, is a measure of a rater's ability to identify individual differences between ratees' performance when everything else remains equal (Brooks & Brooks, 1990; Murphy et al., 1982). In practice, DA could influence the same hiring manager's ability to accurately rate each candidate on specific performance items to identify the most suited individual (Murphy et al., 1982). Wiggins (1973) observed that when rating accuracy is referred to in laymen's terms, it is most often implied to mean differential accuracy.

Borman's Differential Accuracy. Since then Borman (1977) developed a procedure to measure overall differential accuracy by conducting correlational analyses rather than comparing differences between components of accuracy (Becker & Cardy, 1986). Borman's

Differential Accuracy correlates raters' ratings for each dimension with corresponding true scores across ratees which produces a DA score for each dimension (Sulsky & Balzer, 1988). Thereafter the overall DA score is calculated by using Fisher's r-to-z transformation after averaging the correlations across dimensions. Higher scores imply more accurate ratings (Sulsky & Balzer, 1988). Borman's DA is also considered an accurate measure of rating validity (Sulsky & Day, 1994).

Prior Findings of a Linear Relationship

There is overwhelming theoretical support for the role of intelligence in determining how accurately raters assess others (Borman, 1979; Christiansen et al., 2005; De Kock et al., 2020). More recently, a systematic review of over 80 years' worth of determinants of a good judge by De Kock et al. (2020), confirmed that cognitive factors continue to show stronger and more consistent relationships with rating accuracy than any other factors. More specifically, cognitive abilities play an instrumental role in a rater's ability to detect and utilize behavioural cues i.e. the last two stages of the RAM process shaped by the 'good judge' (De Kock et al., 2020).

Cue Detection

Accurate cue detection requires the rater to be attentive and elicit useful information from a ratee (Funder, 1995). It requires raters to focus on their target and broader environment so that they can absorb information and appear interested (Letzring, 2008). Higher levels of social intelligence, a facet of general intelligence, have helped raters better detect cues in interview settings (Speer et al., 2019).

Social Intelligence. Social intelligence is generally understood as a special intelligence linked to the *g* factor representing broader intelligence popularized by Spearman (1904). For this reason, raters with context-specific abilities and social-specific knowledge, can better appraise social demands and detect present cues (Speer et al., 2019). For example, they might be better at establishing a rapport that allows the ratee to feel comfortable which increases disclosure during an interview setting (Christiansen et al., 2005). A study of the good judge's characteristics by Letzring (2008) reported significant correlations between judgmental accuracy and social behaviours exhibited by judges. Judges who were more accurate, "…engage[d] in constant eye contact, expresse[d] warmth, seem[ed] to enjoy the interaction, seem[ed] to like partners, seem[ed] interested in what the partners ha[d] to say, expresse[d] sympathy…" (p. 925). Letzring (2008) concluded by stating that "these behaviours are consistent with the idea that good judges have good social skills and therefore have more relevant information available to them when making judgments" (p. 925). It, therefore, stands to reason that if individuals with higher levels of social intelligence are more accurate, then the same should hold true for individuals with higher levels of general intelligence.

Cue Utilisation

Arguably the most important step of the RAM process, cue utilisation requires the rater to form an accurate judgment about a target (Christiansen et al., 2005). Raters are required to combine the cues detected in the previous stage and accurately interpret whether behaviours are diagnostic of a specific trait (Funder, 1995). Determinants of accurate cue utilization can best be described by working memory and dispositional reasoning.

Working Memory. A strong link exists between working memory capacity and general intelligence that allows more intelligent raters to store, retrieve and manipulate information about ratees quicker (Conway et al., 2003; Murphy & Davidshofer, 1988; Speer et al., 2019). A series of studies conducted by Kyllonel and Christal demonstrated a strong correlation between Spearman's *g* factor and working memory capacity which have since been reproduced (Kyllonen, 1993; Kyllonen, 1996; Kyllonen & Christal, 1990). It can therefore be concluded that more intelligent raters can retrieve and manipulate memories about targets with greater speed and accuracy which produces better quality judgements.

Dispositional Reasoning. Accurate cue utilisation also requires the rater to connect the behaviour displayed with known concepts used to describe it (Christiansen et al., 2005; De Kock et al., 2020). This ability is known as dispositional reasoning and requires knowledge of the relationships among behaviour, traits, and situations (Christiansen et al., 2005). The original study by De Kock et al. (2015) found that dispositional reasoning met the criteria for an intelligence since it positively correlated with intelligence scores and predicted interview accuracy. Like social intelligence, raters with higher levels of dispositional reasoning would also have higher levels of intelligence resulting in higher accuracy scores.

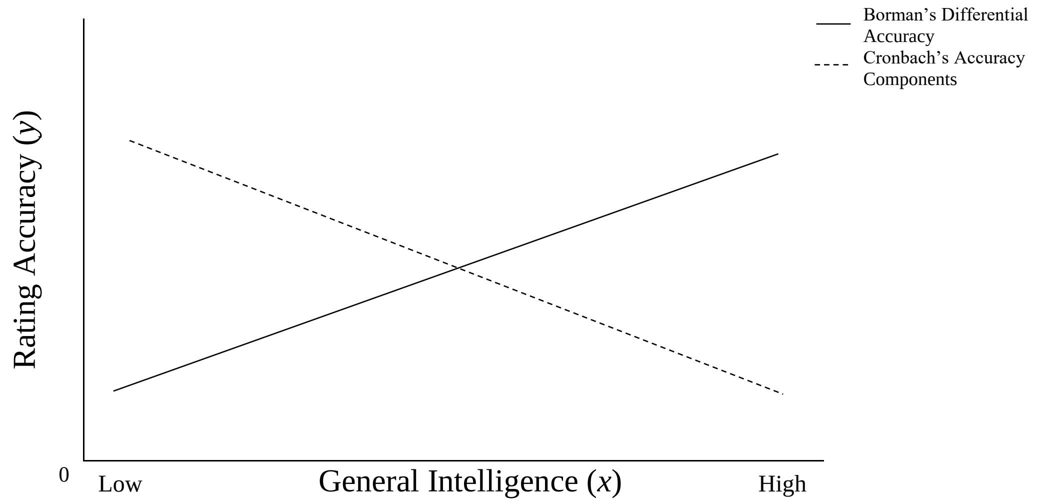
Problem-Solving Abilities. This process of identifying a behavioural cue and finding a relevant explanation also requires the rater to apply problem-solving abilities (Lievens et al., 2006). Researchers have found extensive evidence that individuals with higher cognitive function can also better integrate information and identify creative ways of finding solutions (Greiff & Neubert, 2014). For example, a meta-analysis of 47 studies performed by Stadler et al. (2015) found a statistically significant relationship between intelligence and problem-solving (r = .43).

Empirical Support for a Linear Relationship

A linear relationship between two variables is defined by a straight line where a one-unit increase or decrease in a variable corresponds with the same rate of increase or decrease in another variable (Elliott et al., 2008). It is the simplest and most common relationship used by researchers to predict the value of a criterion from a predictor variable (Cohen et al., 2003). Most of the researchers in the person-perception field have assumed that intelligence and rating accuracy relate in a linear fashion, that is to say, individuals with higher levels of intelligence will always produce more accurate ratings of others at all levels of intelligence. Figure 4 depicts this predicted linear association between the variables using the present study's accuracy measures. Since lower scores for Cronbach's accuracy measures denote higher levels of accuracy, plotted lines would follow the opposite direction compared to Borman's Differential Accuracy scores.

Figure 4

Representation of Linear Relationships Between Study Variables



At the time this study was conducted, only 14 studies from the HRM field reported correlational results between intelligence and rating accuracy documented in Table 1. Correlational coefficients are used to measure the "strength of a linear association between two continuous variables that are both normally distributed" (Elliott et al., 2008, p. 93). It relies on the assumption that both variables increase or decrease at the same rate in either the same (positive correlation) or opposite (negative correlation) directions. Values close to plus or minus one represent perfect correlation and values close to zero suggest no correlation (Field, 2009). Notably, a strong correlation coefficient does not guarantee a causal relationship between the variables (Elliott et al., 2008).

Table 1

Research Evidence on the Correlation Between Intelligence and Rater Accuracy

Author(s)	*N*	Sample description	Effect size [a]	Correlation score (r) [b]
Borman (1979)	146	Students	Small to medium	.26*
Smither and Reilly (1987)	90	Students	Small to medium	-.22*
Brecker (1989)	120	Students	Not available	Not available [c]
Borman and Hallam (1991)	79	Mechanics	Small to medium	.24
Hauenstein and Alexander (1991)	100	Students	Small to medium	-.25* and -.34** [d]
Davis (1999)	82	Assessment centre assessors	Not available	Not available
Lippa and Dietz (2000)	109	Students	Medium	.36**
Janovics (2003)	410	Students	Not available	Not available
Christiansen et al. (2005)	122	Students	Small to medium	.25**
George (2006)	301	Students	Small	Not available
Letzring (2008)	138	Students	Negligible	-.01
Powell (2007)	164	Students	Negligible	-.20 to .09 [e]
De Kock et al. (2015)	146	Managers	Small to medium	.20*
Speer et al. (2019)	128	Students	Medium	.32**

Notes. The studies above do not include work conducted outside of Industrial Organisational Psychology literature.

[a] Effect size (*r*) is expressed according to Cohen 's (1988) guidelines: negligent or no effect (r = 0), small (|.10| < r < |.30|), medium (|.30| < r < |.50|) and large (r > |.50|).

[b] Relationships are positive unless stated as negative.

[c] Not all information was available for certain studies since the original dissertation could not be sourced. The original authors may be contacted for additional information.

[d] Accuracy was related to total intelligence scores for two sample groups: favourable lecturers (*r* = - .34) and unfavourable lecturers (*r* = - .25).

[e] Two of the 44 correlations that researchers tested for were statistically significant which represented the number expected by chance. The overall effect size was therefore considered negligible.

** p < .05 **p < 0.1 (two-tailed)*

The absence of a strong correlation could point to a weak (or non-existent) linear relationship or a potential nonlinear association that better represents the interaction (Elliott et al., 2008).

Two studies found medium effect sizes ($r > .30$) suggesting a strong, positive correlation between the variables (Lippa & Dietz, 2000; Speer et al., 2019). Six studies reported small-to-medium effect sizes ($.10 < r < .30$), with Smither and Reilly (1987) and Hauenstein and Alexander (1991) finding negative correlations. And the last of the accessible studies reported negligible correlations.

It is therefore well-established that there is a connection between rating accuracy and general intelligence, but the nature of this remains misunderstood. Given the inconsistent and counterintuitive findings, the present study sets out to explore whether a potential nonlinear relationship could be causing the slope of the interaction to change after some point. A nonlinear interaction could better account for the conflicting positive and negative results and provide a more nuanced explanation of how the constructs relate.

Empirical and Theoretical Support for Nonlinear Relationships

In nonlinear relationships, an increase or decrease in the independent variable (IV) does not correspond with the same rate of increase or decrease in the dependent variable (DV) (Elliott et al., 2008). Nonlinear interactions are therefore characterised by a lack of consistency in the slope of the relationship (Cohen et al., 2003). Although nonlinearity can take different forms, Robie and Ryan (1999) found that they most often take on four different patterns in the field of personnel selection:

> (a) a nonlinear, positive monotonic relationship; (b) a relationship that is nonlinear and positive monotonic up to a point, after which there is no relationship; … (c) a nonmonotonic relationship in which increased scores are associated with increased performance only to a point, after which further score increases are associated with decreased performance… [(d)] … a nonlinear function which is characterized by an `S' shape. (p. 159)

Available theoretical and empirical evidence of the interaction between intelligence and accuracy points towards the existence of the last two nonlinear forms also known as quadratic and cubic effects (Smither & Reilly, 1987). The following sections will discuss empirical findings and provide theoretical support for each nonlinear interaction.

Support for a Quadratic Effect

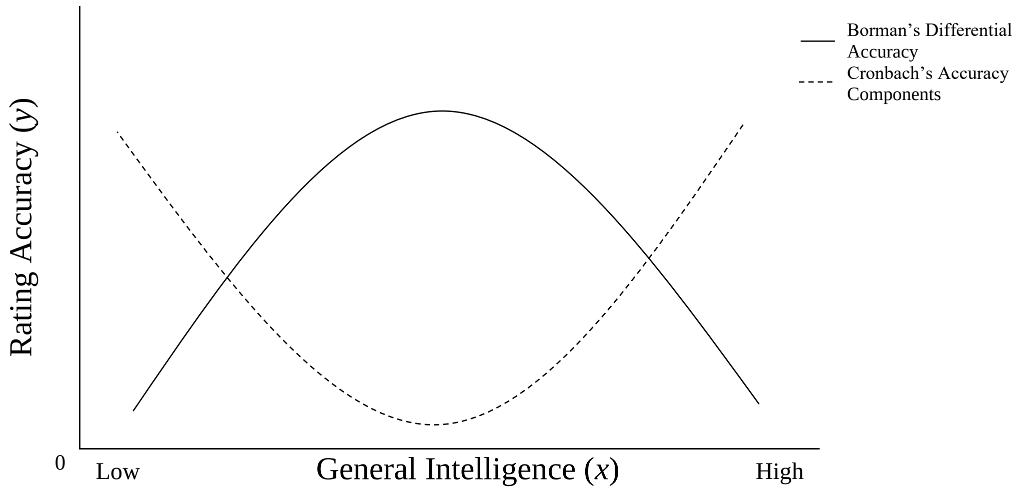
A quadratic relationship can be described as dependent and independent variables that correlate positively or negatively until a certain point after which the direction of the slope changes (Osborne, 2017). The resultant relationship is described as convex when it is negatively accelerated and concave when it is positively accelerated (Ganzach, 1997). Figure 5 depicts how increases in a rater's general intelligence would correspond with changes in rating accuracy in a quadratic model until the relationship reaches a specific IQ threshold after which higher intelligence levels correspond with lower accuracy scores.

Empirical Support for a Quadratic Relationship. Given that so few studies have reported effects between intelligence and rating accuracy, it is not surprising that there is a dearth of research examining data for nonlinear patterns. There are currently two studies that have found quadratic interactions: Smither and Reilly (1987) and Hauenstein and Alexander (1991).

Smither and Reilly (1987) investigated the intercorrelation among job components, time delay in rating, and rater intelligence as determinants of accuracy in performance ratings. They found that intelligence significantly correlated with rating accuracy (r = -.22, p < .05) and with two of Cronbach's accuracy components: stereotype accuracy (r = -.24) and differential elevation (r = -.28). A review of scatterplots prompted researchers to test for quadratic and cubic relationships based on the distribution of data points.

Figure 5

U-Shape Representing Quadratic Relationships Between Study Variables



They subsequently found a significant quadratic relationship between intelligence and differential accuracy, $F(1,87) = 6.04$, $p < .05$, and stereotype accuracy, $F(1,87) = 6.25$, $p < .05$. Researchers concluded that participants with moderate intelligence scores reported higher differential and stereotype accuracy scores producing scatterplots with inverted U-shapes. To account for the nonlinear relationship, Smither and Reilly (1987) argued that the most intelligent raters perceived the tasks as boring and paid less attention to ratee performances compared to participants of moderate intelligence. Although this explained the higher differential accuracy scores, it did not necessarily justify the stereotype accuracy scores.

Hauenstein and Alexander (1991) subsequently found a significant linear relationship between intelligence and rating accuracy when studying how rater implicit theory and intelligence impacted rater ability. Dimensional accuracy was related to general mental ability within both the favourable lecturer ($r = -.34$, $p < .01$) and the unfavourable lecturer ($r = -.25$, $p < .05$) conditions. Using hierarchical regression, researchers tested for a nonlinear relationship between intelligence and rating accuracy (by entering a squared term for intelligence scores to the regression model). General mental ability had a significant nonlinear relationship with elevation ($\Delta R^2 = .07$, $F(1,96) = 7.67$, $p < .01$) but not with dimensional accuracy. They posited that the nonlinear findings are a direct result of the operationalization of elevation:

> Elevation is operationalized by taking the square root of the squared deviation between a rater's overall mean rating and the overall target performance level. By squaring this deviation, there is no distinction between a rater exhibiting an inappropriate elevation due to being too lenient versus being too severe. Intelligent raters were also more stringent than less intelligent raters…Together, these findings suggest that intelligent raters exhibited inappropriate elevations because they were too severe, whereas, the inappropriate elevations exhibited by less intelligent rates were due to leniency effects. (p. 317).

Notably, this finding contradicted the study conducted by Davis (1999) who found that more intelligent raters gave more lenient ratings. Researchers also argued that the type of rater might provide further insight into Smither and Reilly's (1987) nonlinear findings.

Theoretical Support for a Quadratic Effect. Theoretical support for a nonlinear relationship between intelligence and rating accuracy can best be summarized by the notion that too much of a good thing, can be detrimental after reaching a certain threshold (Pierce &

Aguinis, 2013). This was first introduced by Pierce and Aguinis (2013) to explain why researchers in the management field came across paradoxical findings. The too-much-of-a-good-thing principle posits that antecedent variables with associated monotonic positive relations reach context-specific inflection points after which the relations turn asymptotic and lead to negative outcomes (Pierce & Aguinis, 2013). It challenges the conventional understanding that all positive antecedents produce positive outcomes by considering the harm caused to individuals at the highest levels of the predictor variable (Brown et al., 2020). In practice, this would mean that raters with higher levels of intelligence are more accurate than raters of lower levels of intelligence up to a certain point. Thereafter, additional IQ points lower a rater's accuracy scores and the relationship becomes negative (quadratic model).

With reference to Funder's Realistic Accuracy Model, the too-much-of-a-good thing effect can be found in both the cue detection and utilization stages.

Cue Detection. The influence an individual can exert is determined by the manner in which they communicate with others. Simonton (1985) proposed four nonlinear models that explain how more intelligent individuals become less influential after peaking at various intelligence points. Simonton (1985, p. 563) stated that "…if an individual is too intelligent, his or her persuasive communications may be less effective due to low comprehensibility". According to Simonton's second model (comprehension), at higher levels of intelligence, an individual's influence is determined by his or her communication skills – the same can be said for accuracy. Letzring (2008) found that good judges can elicit meaningful cues from targets through active engagement. This provides them with more information about the target to make an accurate judgment. It is therefore argued that intelligent raters are more accurate judges since they can communicate and engage better with targets. However, intelligence levels peak after which more intelligent raters find it difficult to communicate with raters and gather all the information they need to make an accurate assessment.

Cue Utilization. Individuals with higher levels of intelligence also possess different comprehension and reasoning abilities than others (Zizai, 2016). When engaging with targets who score lower on mental abilities, they will tend to overestimate the target's cognitive capacity and unintentionally leave out information or use more technical terms. This is also known as the Dunning-Kruger Effect (Kruger & Dunning, 1999) where individuals with high intelligence overestimate others' intelligence and individuals with lower intelligence overestimate their own intelligence levels. This poses another communication barrier during the cue detection and cue utilization stages. The rater's questioning technique also contributes to the quality of cues that can be observed. More intelligent raters often ask better questions and know when a follow-up is warranted (Christiansen et al., 2005). Based on the above, the following hypotheses were proposed:

Hypothesis 1: There is a statistically significant quadratic relationship between general intelligence and rating accuracy.

Hypothesis 1A: There is a statistically significant quadratic relationship between general intelligence and Borman's Differential Accuracy.

Hypothesis 1B: There is a statistically significant quadratic relationship between general intelligence and Elevation Accuracy.

Hypothesis 1C: There is a statistically significant quadratic relationship between general intelligence and Differential Elevation.

Hypothesis 1D: There is a statistically significant quadratic relationship between general intelligence and Stereotype Accuracy.

Hypothesis 1E: There is a statistically significant quadratic relationship between general intelligence and Cronbach's Differential Accuracy.

Support for a Cubic Model

Cubic relationships contain two slope changes where the dependent variable changes direction (Field, 2009). They can start with a curve that is concave upward followed by concave downward as the independent variable increases, or vice versa (Cohen et al., 2003). Although cubic relationships are rare in the field of HRM, they are not unheard of since most power polynomials are not estimated above the cubic level (Cohen et al., 2003). This is because every bend in a polynomial represents rapid growth, and it is highly unlikely to find data from a population where this pattern exists naturally (Field, 2009).

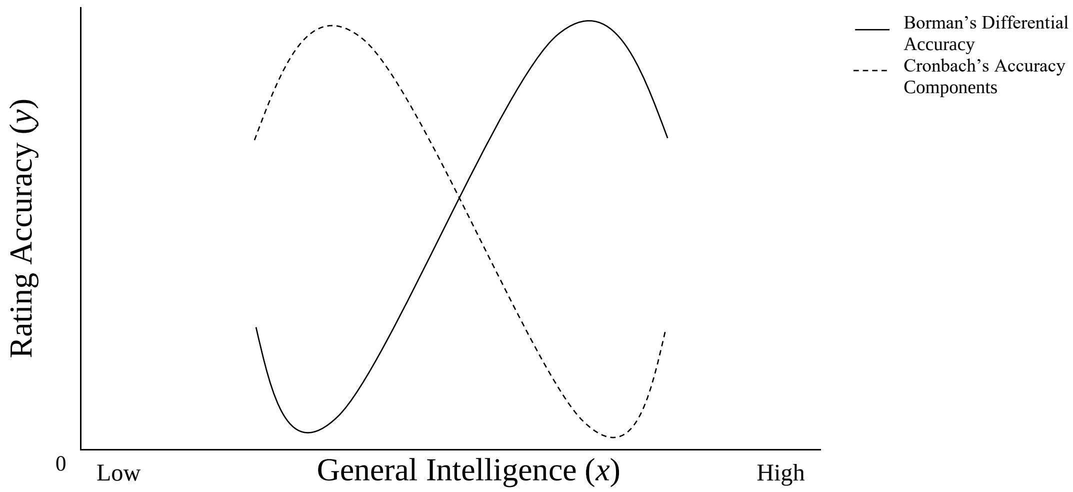
Figure 6 depicts how increases in a rater's general intelligence would correspond with increases and decreases in rating accuracy in a cubic model. It is predicted that increased IQ points at the lower or upper extreme ends would have inconsequential effects on increasing rating accuracy. But, as raters approach the mid-range, increased IQ points have a greater impact on accuracy.

Empirical Support for a Cubic Relationship. Admittedly, empirical support for a cubic relationship between intelligence and rating accuracy is limited since both studies that found quadratic effects, did not find significant cubic effects (Hauenstein & Alexander, 1991; Smither & Reilly, 1987). By replicating the approach followed by Hauenstein and Alexander

(1991) and Smither and Reilly (1987), the intention of this study was to either reproduce previous nonsignificant findings with a sample of managers to offer renewed support or discover a cubic interaction to be further investigated.

Figure 6

S-Shape Representing Cubic Relationships Between Study Variables



Theoretical Support for a Cubic Relationship. Increased problem-solving abilities allow raters to analyse behavioural cues and make an accurate assessment of which underlying traits they could represent (Lievens et al., 2006). Simonton's (1985) first model (intellectual superiority) states that individuals who have higher levels of cognitive ability have a greater degree of influence over others. It is argued that individuals look to find leaders that they can refer to as task specialists who can provide expert knowledge on subject matters – this requires extensive problem-solving abilities (Antonakis et al., 2017). However, Simonton also found that above an IQ score of 120 there is little to no variance between participants' creative problem-solving skills. This nonlinear relationship was also confirmed by Antonakis et al. (2017, p. 33) who stated that "…it is not that high IQ does not matter, but that it matters less at higher scores". Simonton (1985) further argued that after an IQ of 120, there is a negligible correlation between problem-solving ability and intelligence and therefore perceptions of influence increase at a lower rate.

The same can be said at the lower end of intelligence where low IQ scores correlate with low accuracy scores at a lower rate. Individuals with poor cognitive abilities are expected to struggle more with correctly connecting behavioural cues with the appropriate underlying traits (Christiansen et al., 2005). The proposed cubic model is therefore flat at the extreme ends so that increased intelligence matters most at moderate levels of intelligence. The following hypotheses were developed for the present study:

Hypothesis 2: There is a statistically significant cubic relationship between general intelligence and rating accuracy.

> **Hypothesis 2A:** There is a statistically significant cubic relationship between general intelligence and Borman's Differential Accuracy.
>
> **Hypothesis 2B:** There is a statistically significant cubic relationship between general intelligence and Elevation Accuracy.
>
> **Hypothesis 2C:** There is a statistically significant cubic relationship between general intelligence and Differential Elevation.
>
> **Hypothesis 2D:** There is a statistically significant cubic relationship between general intelligence and Stereotype Accuracy.
>
> **Hypothesis 2E:** There is a statistically significant cubic relationship between general intelligence and Cronbach's Differential Accuracy.

Chapter Summary

Funder's (1995) Realistic Accuracy Model argues that judges of targets' traits and abilities go through four distinct stages to reach an accurate assessment. The last two stages (cue detection and utilisation) place a significant cognitive load on raters to reach accurate judgments (Christiansen et al., 2005). More specifically, raters with increased levels of social intelligence, working memory and dispositional reasoning can store, retrieve, and manipulate information about targets faster and more precisely. As a result, general intelligence (IQ) has been shown to be the strongest and most consistent predictor of accurate judgments compared to other individual differences between judges.

Up to this point, researchers have assumed that more intelligent raters always produce more accurate judgments (linear association between variables). But inconsistent correlation coefficients from previous findings point to the need for further investigation. The present study hypothesises that a quadratic or cubic model could better account for the variance in rating accuracy by general intelligence. Reasons for this include judges' communication skills, the Dunning-Kruger effect and problem-solving abilities. The next chapter discusses the methodology applied to test the hypotheses proposed.

Chapter 3: Method

This chapter provides a detailed breakdown of how the present study was conducted to answer the following research question: Can a nonlinear model better predict the relationship between intelligence and rating accuracy than a linear model? It includes sections on the research design, sample and procedure, materials, measures, ethical considerations, and statistical analyses conducted to test the study hypotheses.

Research Design

A secondary quantitative research design was adopted to re-examine data from a prior published study. Quantitative research is useful to expand on existing theories within the literature and is argued to be the most appropriate since this study required the analysis of numerical data. This also allowed the production of generalizable results that can be replicated by future researchers (Williams, 2007). This approach is aligned with a positivist paradigm that aims to observe and measure variables so that they may be quantified and used to predict phenomena (Antwi & Hamza, 2015). The study also employed a descriptive research design to observe the relationship between general intelligence and rating accuracy and understand how these variables interact (Williams, 2007).

Using Secondary Research

Secondary research is the process of accessing and utilizing data collected by other parties (Johnston, 2017). Data for the study was sourced from research conducted by De Kock et al. (2015) who adopted a cross-sectional design to utilize data collected at one point in time (Levin, 2006). This data was considered suitable since researchers measured levels of intelligence and rating accuracy (among other variables) in a South African context with experienced professionals. Furthermore, this dataset was published in a reputable journal that requires peer-review and subjects research data to intense scrutiny. Any data used from this study is therefore deemed trustworthy and accurate (Parsons et al., 2010). It also enabled the use of the same dataset to generate linear and nonlinear models that could be compared to published work. This provided further credence to the findings and added to the growing person-perception literature. Using secondary data has also become increasingly important as it accelerates the generation of research, and pools intellectual resources that can provide more immediate insight into topical phenomena (Gilmore et al., 2018). Additionally, it also provides access to a larger dataset that could not easily be accessed by the researcher alone within the timeframe of this study. A larger dataset allows for stronger statistical power and better detection of significant findings (Andersen et al., 2011). Analysing secondary data means that time devoted to data collection could be redirected to data analysis which is the primary focus of this study.

Sample

A homogenous purposive sampling strategy was employed in the original study to recruit participants. Eligibility was determined by two criteria: participants were employed and had at least five years' experience in managerial positions to increase the generalizability of the study results (Ferguson, 2004). Participants were recruited from a sample of 146 South African police managers who attended a seven-week managerial training course to qualify for a promotion. Participants ranged from lower-level supervisors to senior management with titles Captain, Warrant Officer or Lieutenant. They were deemed an appropriate sample for the study as they were the most likely employees in the organisation to conduct interviews with potential candidates (De Kock et al., 2015). The sample consisted of 75.35% of participants who identified as male and 71.20% as Black African. The average age of participants was 43.67 years with a standard deviation of 5.34 (n = 133). More detailed demographic information can be found in Table 2 with the percentage of unspecified or missing fields.

Table 2

Demographic Statistics of Study Participants[5]

		Frequency	**Percentage of Total**
Gender	Male	107	73.29%
	Female	35	23.97%
	Unspecified	4	2.74%
Level of Education	Grade 10	6	4.11%
	Grade 12	76	52.05%
	National Diploma	35	23.97%
	Bachelor's Degree	13	8.90%
	Honours Degree	1	.68%
	Master's degree	1	.68%
	Unspecified	14	9.59%

[5] The study sample was not considered WEIRD (participants from Western, educated, industrialized, rich and democratic societies) and posed no concern to the study's generalizability (Henrich, et al., 2010). If more demographical information is required, feel free to consult the original published study.

Procedure

Researchers explained the rating procedure and materials to participants at the start of the training session introducing it as a component of the training course. To ensure that participants understood what was expected, they were shown a video-recorded interview segment on a projector with audio equipment. Thereafter they could discuss the ratings they provided and ask any questions to clarify the process. This was followed by four interview segments shown in the same format which participants were asked to rate independently. They were provided with interview dimension rating sheets and individual difference measures to complete. Researchers concluded the study by debriefing participants and thanking them for participating (De Kock et al., 2015).

Materials

Video Interviews and Rating Materials

Semi-structured interviews were recorded, between an expert interviewer and five graduate students recruited to put themselves in the position of someone applying for a fictional junior management position and told that the interview simulation would assist in preparing them for a real job application. The questions were situational and designed to test two dimensions: their communication and people management skills. All graduates were presented with the same, eight open-ended questions like "How would you handle a situation where your work colleagues ignore your ideas and input?" Study participants were shown the same, shortened video recordings of five minutes each and asked to rate interviewees' responses on a 7-point Likert Scale rated from 1 (poor) to 7 (excellent). Each question contained a rating guide to assist participants with understanding the scale points. For example, a score of 1 would be associated with an answer of "I stop giving ideas and input" to the question above.

Video Interviewee True Scores

Seven subject matter experts were asked to rate the interviewees' responses based on their communication and people management skills. This was done to obtain a true score i.e. an estimate of a candidate's true performance (Chirico et al., 2004). The panel consisted of professors of Industrial and Organisational (IO) Psychology and individuals with at least a master's degree in IO. Expert ratings were averaged to obtain true scores for interviewees and an intraclass correlation coefficient (ICC) was formulated to assess the correlation between groups of data (Bartko, 1966). According to LeBreton and Senter (2008), a score between .71 to .90 indicates a strong agreement between raters. De Kock et al. (2015) reported strong mean interjudge agreement for each dimension and across all dimensions [ICCtot(2, k) = .86].

Measures

Rating Accuracy

In the study conducted by De Kock et al. (2015), participant rating accuracy was selected as the dependent variable. To determine the rating accuracy score for each participant, researchers calculated Cronbach's (1955) four accuracy measures to determine the deviation between the true scores and rating scores using procedures developed by Sulsky and Balzer (1988). Table 3 provides a breakdown of the formulae used to calculate each component as outlined by Brooks and Brooks (1990) where:

- i is the number of items rated
- k is a specific ratee
- m is a specific item
- n is the number of ratees
- $\bar{t}$ is the grand mean of the true scores across all ratees and items
- $\bar{x}$ is the rater's grand mean across all ratees and items

Table 3

Formulae for Cronbach's (1995) Accuracy Components

Accuracy Component	Formula
Elevation	$(\bar{x} - \bar{t})^2$
Differential Elevation	$\frac{1}{n}\sum_{k=1}^{n}[(\bar{x}_k - \bar{x}) - (\bar{t}_k - \bar{t})]^2$
Stereotype Accuracy	$\frac{1}{i}\sum_{m=1}^{i}[(\bar{x}_m - \bar{x}) - (\bar{t}_m - \bar{t})]^2$
Differential Accuracy	$\frac{1}{ni}\sum_{k=1}^{n}\sum_{m=1}^{i}[(x_{km} - \bar{x}_k - \bar{x}x_m + \bar{x}) - (t_{km} - \bar{t_k} - \bar{t}_m + \bar{t})]^2$

Researchers then computed Borman's Differential Accuracy using the formula below where d is the number of dimensions, m is a specific item, r is the subject rating, t is the true score, T_{rt} the correlation between ratings and true scores for a particular dimension transformed to a Z-score using Fisher's r-to-z transformation to obtain average correlations (Mendoza, 1993).

$$\text{Borman's Differential Accuracy} = \frac{1}{d}\sum_{m=1}^{d}(T_{rt})$$

General Mental Ability

Before viewing the recorded interviews, participants were asked to complete the Wonderlic Personnel Test-Revised (WPT-R) to assess their general mental ability. The test comprises 50 multiple-choice items that need to be completed within twelve minutes (Gill & Brajer, 2012). The test is designed to assess "…mathematical, verbal, logical, and analogical reasoning skills…" (Matthews & Lassiter, 2007, p. 707). The WPT-R has been widely used for personnel selection and has proven to strongly correlate with another popular intelligence measure, namely the Wechsler Adult Intelligence Scale (Dodrill, 1981). The test has also been proven to be consistent with reliability scores above .80 (Wonderlic Inc, 2002).

Ethical Considerations

In line with recommendations from the American Psychological Association (2017) and Gilmore et al. (2018), authority was obtained from the principal researchers to access and analyse the data collected in the original study. Permission was also obtained from the University of Cape Town Faculty of Commerce Research Ethics Committee before proceeding with this study (refer to Appendix A). Although the researchers in the primary study obscured the purpose of the experiment, participants were debriefed afterwards in line with the American Psychological Association ethical guidelines. Researchers justified the choice of limited deception as it was necessary to ensure the study's internal validity (Linden et al., 2010). The choice to disclose the true nature of the study did not cause participants any physical pain or emotional distress (Sharpe & Faye, 2009). Participation was voluntary; consent was obtained; researchers did not collect any identifiable data from participants and no payments or gifts were given to incentivise participation.

Statistical Analyses

The 28th version of the IBM Statistical Package for the Social Sciences (SPSS) was used to clean and analyse the secondary data set once access was granted by the original researchers. After the data were checked for irregularities, descriptive statistics were run to summarize the dataset and identify intercorrelations between variables. To test hypotheses one and two, ordinary least squares regression analysis was conducted, but not before assessing whether the data met the assumptions for linear regression.

Analytical Approach to Linear and Nonlinear Regression

The present study employed the same technique to identify nonlinear effects as Hauenstein and Alexander (1991), Robie et al. (2020) and Smither and Reilly (1987) by testing different nonlinear models using linear regression analysis. This technique was best suited to test both study hypotheses since it could predict the quantitative value of a criterion variable from one or many predictor variables using the parameter estimates obtained (Kelley & Maxwell, 2010; Young, 2017). Additionally, regression can estimate the power of the independent variable on the dependent variable and compare the statistical significance between a nonlinear model and linear model (Cohen et al., 2003). Since linear regression requires the parameters or coefficients to be linear and not the interaction between variables, it is deemed suitable for nonlinear models (Cohen et al., 2003).

A simple linear regression function can also be expressed as Equation 1 where *y* represents rating accuracy, *x* is intelligence, *a* is the intercept (the point at which either axis of a graph is intersected by a line plotted on the graph), *b* is the regression coefficient, and *e* is the error term (the difference between the data collected and the population of the study) (Cohen et al., 2003; Seber & Wild, 1989). A one unit increase in *x* would correspond with a constant magnitude of increase in *y* across the entire scale (Cohen et al., 2003).

$$y = a + b_1x + e \tag{1}$$

Ordinary Least Squares (OLS) is a popular linear regression technique used to estimate the unknown parameters of a model (Guion & Gibson, 1988; Long, 2008) OLS minimizes the sum of the squared errors within a specific dataset to find a model that best fits the data (Long, 2008).

Steps Required to Examine Nonlinear Relationships

Although previous studies broadly followed the same approach, the most comprehensive outline for performing nonlinear regression was provided by Cohen et al. (2003). It was recommended that researchers follow four steps to compute accurate nonlinear functions.

Considering the Correct Approach. According to Cohen et al. (2003), there are four approaches to using regression to examine nonlinear relationships: power polynomials, orthogonal polynomials, nonlinear transformations and nonparametric regression. This study employed the first approach, power polynomials, which is the most commonly used and simplest technique as evidenced by its use in previous studies (Hauenstein & Alexander, 1991; Robie et al., 2020; Smither & Reilly, 1987). By using power polynomials, researchers can add higher order terms like x^2 and x^3 to transform the original *x* variable (Bobko, 2001). The highest order term in a polynomial equation determines the overall shape of the regression function and by extension, the number of bends in the function (Cohen et al., 2003). Orthogonal polynomials were deemed inappropriate since the process requires the predictor variables to be categorical in nature whereas the present study examined continuous variables (Cohen et al., 2003). Nonlinear transformations were also deemed inappropriate since they do not produce R squared values or p-values for parameter estimates which would make it difficult to compare the statistical significance of different models. And finally, since nonparametric regression does not require the specification of a nonlinear model type beforehand it heavily relies on a large sample size to inform the relationship (Cohen et al., 2003). A sample size of 141 participants would therefore not meet the requirement for producing accurate results.

Centering the Predictor Variable Around the Mean. As recommended by Aiken et al. (1991) and Cohen et al. (2003), the predictor variable should first be centered around the mean to improve the accuracy of the regression equations. Centering involves the linear transformation of the predictor values to the mean of the predictor by subtracting the mean score of the predictor from each predictor value (Cohen et al., 2003). By centering the predictor, it reduces nonessential[6] multicollinearity with higher order terms, for example, *x* would highly correlate with x^2 in a regression equation without mean centering. There was no need to center the criterion variable in the present study since it would have no impact on the regression coefficients in the equation (Aiken et al., 1991; Iacobucci et al., 2016).

Creating New Variables. Before researchers can add higher order terms to a regression equation, they have to first create them by multiplying the centered predictor scores to produce x^2 and x^3 (Cohen et al., 2003). This was done in SPSS after centering the predictor around the mean.

Using Power Polynomials to Test Nonlinear Relationships. The next step is to conduct linear regression analyses using the polynomial functions appropriate for the study

[6] Essential multicollinearity exists between variables that are likely to be correlated given the nature of the constructs e.g., the age of a child and his or her development stage (Aiken et al., 1991; Iacobucci et al., 2016). Nonessential multicollinearity "…describes correlations that arise due to issues of measurement or in the moderated multiple regression context, the fact that X_1 and X_2 are likely correlated with their product term X_1X_2 because, of course, they are contained within it" (Iacobucci et al., 2016, p. 1309).

(Cohen et al., 2003). To test for a quadratic or U-shape function, the predictor should be entered as a linear (x) and nonlinear (x^2) variable to capture one bend in the function (Cohen et al., 2003). It would produce Equation 2 where the slope of the regression of y on x is different for each value of x (Cohen et al., 2003).

$$y = a + b_1x + b_2x^2 + e \qquad (2)$$

In addition to lower order terms, a third-order predictor (x^3) should be added to Equation 2 to test for a cubic function (Cohen et al., 2003). The cubic function would have two points where the slope is equal to zero producing an S-shape function in Equation 3.

$$y = a + b_1x + \ b_2x^2 + b_3x^3 + e \qquad (3)$$

Test of Significance of Highest Order Coefficient. The highest order coefficients will indicate whether additional nonlinear terms added to the overall predictiveness of the model above and beyond the lower order terms (Cohen et al., 2003). Researchers should assess the incremental improvement in predictive value after the highest order term was added to determine whether the nonlinear model is more significant.

Regression Functions Used to Test Study Hypotheses

In line with Cohen's recommendations, the following equations were developed to test both study hypotheses.

Hypothesis 1. Since hypothesis 1 tested for a quadratic function between intelligence (IV) and rating accuracy (DV), a squared term was added to Equation 1 to test for a U-shaped curve similar to the studies conducted by Hauenstein and Alexander (1991), Robie et al. (2020) and Smither and Reilly (1987). For **hypothesis 1A**, the sign of the highest order term was made negative so that the curvature of the function is inverted or concave downward (Cohen et al., 2003). It produced Equation 4 where general intelligence was entered as a linear (x) and nonlinear (x^2) predictor.

$$y = a + b_1x - b_2x^2 + e \qquad (4)$$

Since lower values of Cronbach's accuracy components denote higher accuracy levels, **hypotheses 1B, C, D and E** were expected to produce U-shaped functions with a positive sign before the highest order term in Equation 5.

$$y = a + b_1x + b_2x^2 + e \qquad (5)$$

Hypothesis 2. The second hypothesis tested for an S-shape function where increases in intelligence scores would be inconsequential at extreme ends in predicting accuracy (Robie & Ryan, 1999). To compute this, a third-order predictor (x^3) was added to Equation 2 for the second bend in the curve. When testing **hypothesis 2A**, a negative sign was added before bx^3 in Equation 6 to produce a function that is concave downward and then concave upward (Cohen et al., 2003).

$$y = a + b_1x + b_2x^2 - b_3x^3 + e \tag{6}$$

The inverse would be true for **hypotheses 2B, C, D and E** where lower scores for Cronbach's accuracy measures represent higher accuracy levels. This was expected to produce S-shaped functions concave upward and then concave downward in Equation 7.

$$y = a + b_1x + b_2x^2 + b_3x^3 + e \tag{7}$$

Chapter Summary

This chapter discussed the research methodology applied in the present study to test the hypotheses laid out in the previous chapter. Special attention was paid to the choice of research design, study sample, procedure, materials, measures and ethical considerations. It concluded by explaining the statistical analyses performed to produce results discussed in the next chapter.

Chapter 4: Results

The primary goal of this study was to determine whether a nonlinear model could better predict the interaction between general intelligence and rating accuracy. This chapter presents the findings of ordinary least squares regression analyses after comparing the variance explained by the predictor variable (IQ) in the criterion variable (rating accuracy) across linear and nonlinear models.

Preparing Data for Analysis

Before analysis commenced, data for general intelligence and rating accuracy were examined in SPSS for errors and missing values to increase the quality of results obtained, as recommended by Tabachnick and Fidell (2013). No errors or coding issues were found, but five cases were identified with missing values for Cronbach's accuracy components and general intelligence scores. They were subsequently deleted and a total of 141 cases remained (97% of the original study participants). Thereafter the predictor variable was centered around the mean (Appendix B includes the SPSS syntax used the perform the mean centering) and higher-order variables (x^2 and x^3) were created in SPSS to compute power polynomials.

Reliability Analysis

A reliability analysis was conducted during the original study to determine whether the Wonderlic Personnel Test-Revised (WPT-R) measure consistently reflected the construct (general intelligence) that it was measuring (Field, 2009). De Kock et al. (2015) reported an internal consistency reliability score of .75 which is above the .70 score recommended for ability measures with a high level of internal consistency (Field, 2009), but below the recommended level of .80 for cognitive tests (Kline, 1999).

Descriptive Statistics

The means, standard deviations, and bivariate correlations are presented in Table 4. Both general intelligence and rating accuracy were analysed as continuous variables. The unstandardized mean score for general mental abilities (or IQ) across all participants was 12.42 (SD = 4.4). Understandably, the age of the student participants was significantly lower than that of this sample comprising of mid-level management (M = 43.73, SD = 5.38).

Table 4

Descriptive Statistics and Intercorrelations [a]

	M	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1. Age	43.73	5.38	–									
2. Level of education	–	–	-.16	–								
3. General mental ability [b]	.00	4.43	-.36**	.25**	–							
4. General mental ability [b] squared (x^2)	19.46	32.15	-.23**	-.07	.40**	–						
5. General mental ability [b] cubed (x^3)	55.95	432.14	-.25**	.07	.74**	.72**	–					
6. Borman's Differential Accuracy	.99	.65	-.15	.13	.20*	-.01	.08	–				
7. Cronbach's Elevation Accuracy	.64	.45	.07	.12	-.16	-.12	-.07	-.28**	–			
8. Cronbach's Differential Elevation	.66	.28	.19*	-.07	-.12	.10	-.01	-.23**	-.09	–		
9. Cronbach's Stereotype Accuracy	.24	.21	.18	-.07	-.03	-.09	-.09	-.21*	.04	-.03	–	
10. Cronbach's Differential Accuracy	.46	.22	-.09	.01	-.12	-.12	-.12	-.11	-.04	-.02	.09	–

Note. *N* ranged from 128 to 141.
[a] The Pearson Product-Moment Correlation was used to determine correlation coefficients.
[b] General mental ability scores have been centered around the mean of the predictor
$* p < .05$ $**p < 0.1$ (two-tailed)

GMA positively correlated with Borman's Differential Accuracy (*r* = .20, p < .01) and unsurprisingly with its squared (*r* = .40) and cubed (*r* = .74) terms as well. GMA was also associated with age (*r* = -.36) and level of education (*r* = .25). Borman's Differential Accuracy significantly correlated with three of Cronbach's components: elevation accuracy (*r* = -.28), differential elevation (*r* = -.23) and stereotype accuracy (*r* = -.21). Interestingly, there was also a positive correlation between age and Cronbach's differential elevation (*r* = .19) suggesting that younger participants were better at judging overall performance across different dimensions than aged participants.

Testing Regression Assumptions

To ensure that the results derived from the regression analyses could be generalized to the population of interest, six underlying assumptions were evaluated with general intelligence as the predictor and Borman's Differential Accuracy and Cronbach's accuracy components as criterion variables (Field, 2009).

Independence of Errors

Residual error terms should be uncorrelated or independent to achieve accurate regression results (Flatt & Jacobs, 2019). This assumption was tested using the Durbin-Watson coefficient which provides a value between zero and four. Values equal to two indicate no correlation between error terms; values below one imply that error terms are positively related; and values above three imply that they are negatively related (Field, 2009; Flatt & Jacobs, 2019). Results in Table 5 confirmed that there were no significant correlations between residual errors.

Table 5

Durbin-Watson Coefficients

Criterion Variable	Coefficient
Borman's Differential Accuracy	1.66
Elevation	1.66
Differential Elevation	1.91
Stereotype Accuracy	2.19
Cronbach's Differential Accuracy	1.87

Homoscedasticity

The scatter plots in Appendix C show the residuals randomly and evenly scattered around zero (Osborne & Waters, 2002). This confirmed that the variance between the error residuals was constant at each level of the predictor variable which met the assumption of homoscedasticity (Field, 2009).

Absence of Multicollinearity

Multicollinearity exists when two or more predictors correlate within a regression model (Daoud, 2017). (Disatnik & Sivan, 2016; Field, 2009). According to Disatnik and Sivan (2016), this can cause standard errors of the regression coefficients to become inflated, and consequently the entire analysis of the coefficients may be confounded. Given that this study employed a single predictor, tests for multicollinearity were not deemed necessary and therefore not performed (Field, 2009).

Normally Distributed Errors

The normal distribution of error terms was investigated to ensure that accurate p-values are generated for t-tests (Flatt & Jacobs, 2019). In the Probability-Probability plots in Appendix D, there are insignificant deviations from the 45° line confirming that the residual errors were normally distributed (Ghasemi & Zahediasl, 2012) for all accuracy measures.

Significant Outliers

According to Cohen et al. (2003), "polynomial equations may be highly unstable and can be grossly affected by individual outliers" (p. 212). It was, therefore, essential to examine the data for values that bias the mean and inflate the standard deviation (Field, 2009). Cook's distance is a measure of an observation's impact on regression coefficients and values above one indicate that there are outliers within the dataset that can distort the results (Field, 2009; Walfish, 2006). Values ranging from .00 to .17 were reported indicating that no concerning outliers were identified. Table 6 details the minimum and maximum values for Cook's distance across each accuracy measure.

Sample Size

To ensure that the study sample is of adequate size to detect significant results, Tabachnick et al. (2019) suggested that researchers calculate the required number of participants determined by the number of predictor variables. More specifically, the sample size should be larger than $50 + 8(k)$ where k is the number of predictors. Given that the present study employed one predictor (IQ), the sample of 141 participants meets this requirement.

Table 6

Cook's Distance Values

Criterion Variable	Minimum	Maximum
Borman's Differential Accuracy	.00	.14
Elevation	.00	.13
Differential Elevation	.00	.12
Stereotype Accuracy	.00	.17
Cronbach's Differential Accuracy	.00	.06

Tests of Hypotheses

Linear Regression Baseline

In order to compare the predictiveness of nonlinear models to a linear model, a baseline was established in line with the methodology employed in the original study by De Kock et al. (2015). The baseline represents the predominant theoretical understanding of how general intelligence interacts with rating accuracy (Hauenstein & Alexander, 1991). More specifically, a one unit increase in intelligence corresponds with a one beta unit increase in rating accuracy producing a positive linear relationship. The statistical significance of the baseline was compared to that of nonlinear models to establish which model can better predict rating accuracy.

Borman's (1977) Differential Accuracy as Criterion Variable. Using OLS regression analysis, general intelligence was added to the linear model as the predictor variable and rating accuracy as the criterion variable. The model was determined to be significant ($F_{1,139}$ = 5.92, R^2 value = .04, adjusted R^2 value = .03) at a p-value below .05 (the significance level for this study). This means that the null hypothesis was rejected when the likelihood of the observed results caused by random sampling error, was 5% or less (Field, 2009; Sullivan & Feinn, 2012). Although the significance level is considered arbitrary, researchers are in agreement that a level of .05 is adequate to detect material effects (Cohen, 1994).

Results showed that 4% of the variance in rating accuracy was explained by general intelligence (see Appendix E for detailed findings). It also produced the regression equation below which predicts that a one unit increase in a rater's intelligence corresponds with a .30 unit increase in their rating accuracy.

$$\textit{Rating accuracy} = .99 + .03(\textit{general intelligence})$$

Cronbach's (1955) Components as Criterion Variables. The linear regression model was also tested with general intelligence as the predictor variable and each of Cronbach's accuracy components as criterion variables. None of the components produced statistically significant results with 1% of rating accuracy explained by changes in general intelligence. The interaction with elevation was the only one to produce a significant result (p = .05) with a 95% confidence interval between -.03 and .0001 ($F_{1,139}$ = 3.85, R^2 value = .03, adjusted R^2 value = .02).

According to Du Prel et al. (2009), when a confidence interval includes zero a finding cannot be considered statistically significant even at a p-level equal to or below .05. The inclusion of zero means that one cannot rule out that there is no significant relationship in the population (Du Prel et al., 2009). For this reason, the linear model between general intelligence and elevation cannot be deemed statistically significant (see Appendix F for full regression output).

A summary of regression results with each criterion variable can be found in Table 7. Figure 7 includes graphic depictions of the linear curve estimations of each model. Graphs include observed data points from the Wonderlic Personnel Test-Revised (WPT-R) on the x-axis and the relevant accuracy measure on the y-axis.

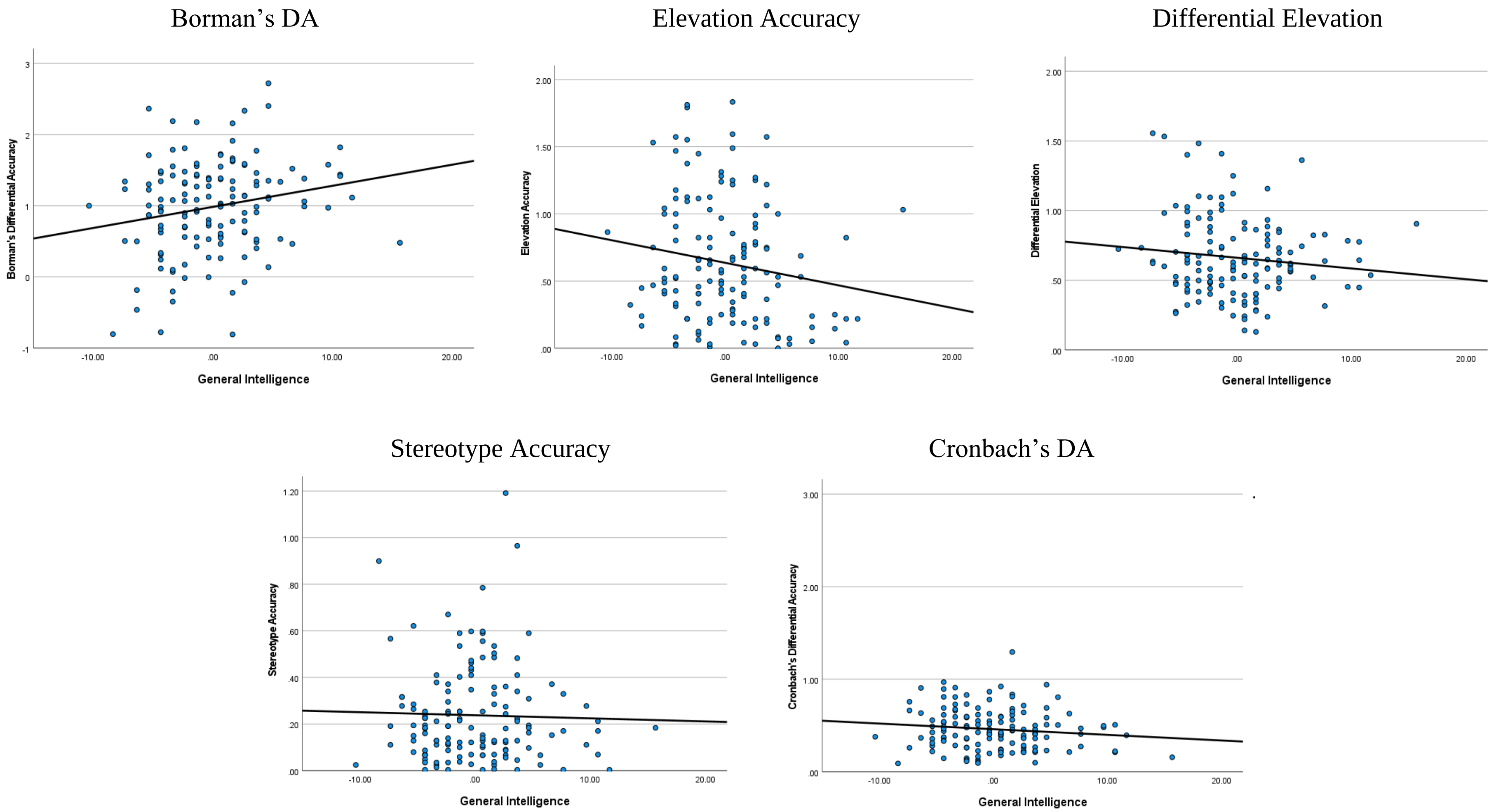
Table 7

Regression Results for Linear Models

Criterion Variable	Adjusted R^2	F-value	*df*	Significance level	Regression coefficient
Borman's Differential Accuracy	.03	5.92	1, 139	.02	.03
Elevation	.02	3.85	1, 139	.05	-.02
Differential Elevation	.01	2.02	1, 139	.16	-.01
Stereotype Accuracy	-.01	.11	1, 139	.74	-.01
Cronbach's Differential Accuracy	.01	2.06	1, 139	.15	-.01

Figure 7

Linear Curve Estimations



Hypothesis One: Testing for a Quadratic Effect

The first hypothesis tested whether there was a statistically significant quadratic relationship between general intelligence (predictor variable) and rating accuracy (criterion variable). Additionally, hypothesis 1A used Borman's DA as the DV and hypotheses 1B-E contained each of Cronbach's four accuracy components as DVs. This was done by adding a squared term to the predictor variable to produce a quadratic regression model similar to studies conducted by Hauenstein and Alexander (1991), Robie et al. (2020) and Smither and Reilly (1987). This produced an effect size estimate for the nonlinear term allowing for easy comparison with the linear term (Brown et al., 2020).

Results are displayed in Table 8 and no statistically significant effects were detected. The null hypothesis was therefore rejected. Significance levels ranged from .06 to .45 with all R^2 changes equal to or below .01. The only exception applied to the quadratic model that included differential elevation which accounted for a 3% increase in effect size from the linear model. Quadratic curve estimations of the above models can be found in Figure 8 and detailed results can be found in Appendix G and H.

Table 8

Polynomial Regression Results for Quadratic Models

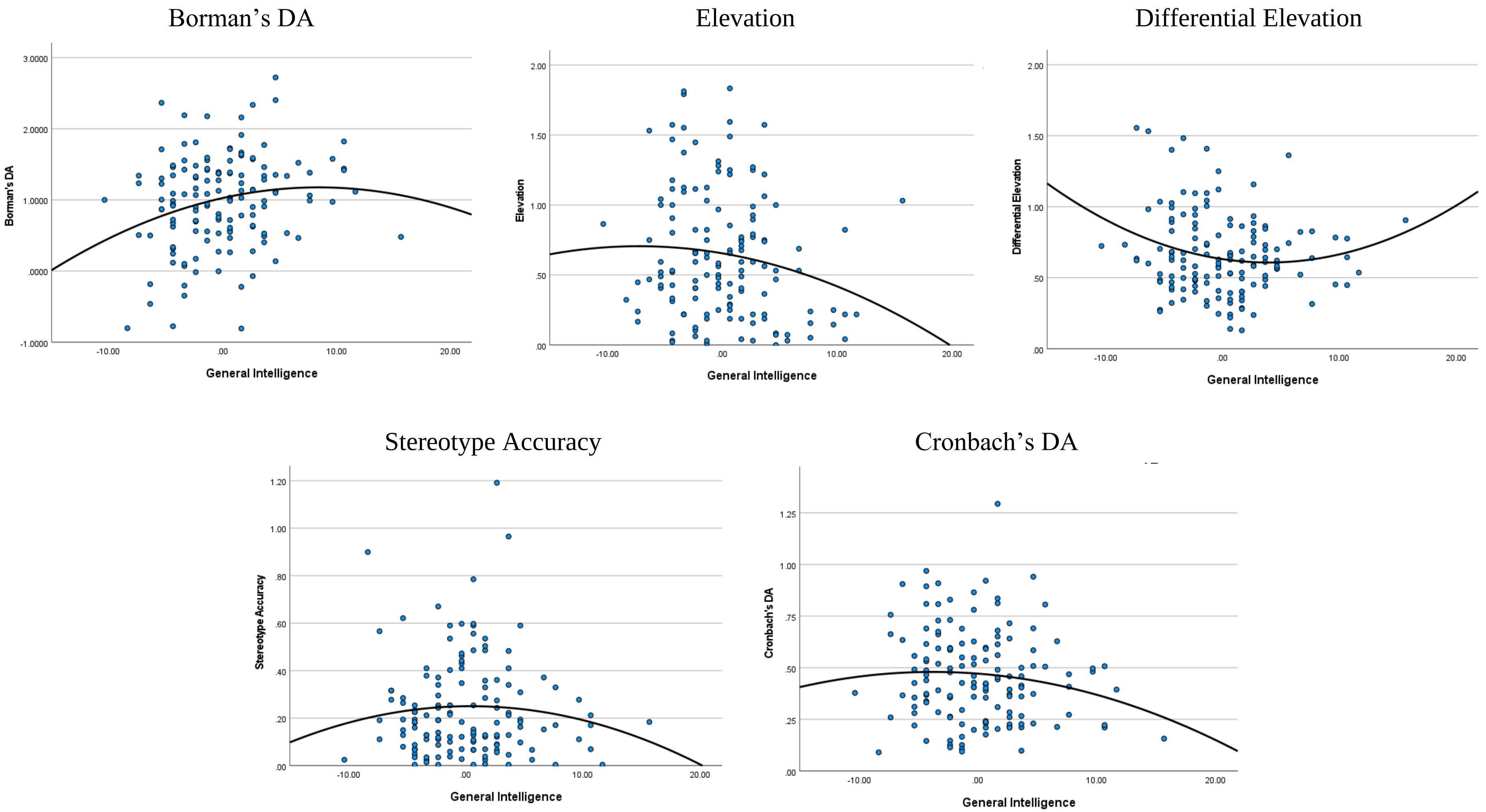
Criterion Variable	**Adjusted R^2**	**R^2 Change**	**F-value Change**	***df***	**Significance Level**	**Regression coefficient**
Borman's Differential Accuracy	.04	.01	1.35	1, 138	.25	-.002
Elevation	.02	.00	.56	1, 138	.45	-.001
Differential Elevation	.04	.03	3.74	1, 138	.06	.002
Stereotype Accuracy	-.01	.01	1.16	1, 138	.28	-.001
Cronbach's Differential Accuracy	.02	.01	.84	1, 138	.36	-.001

Hypothesis Two: Testing for a Cubic Effect

Hypothesis 2 tested whether there was a statistically significant cubic relationship between general intelligence (predictor variable) and Borman's DA (hypothesis 2A) and Cronbach's accuracy components (2B-E). As such, a third-order or cubic polynomial term was added to the regression model to produce an S-shaped function. Results in Table 9 reported no significant findings of a cubic interaction with Borman's DA, differential elevation, stereotype accuracy or Cronbach's DA. But a cubic effect was found between IQ and elevation ($F_{1,137}$ = 2.76, adjusted R^2 value = .04) with p-value = .05.

Figure 8

Quadratic Curve Estimations



On closer inspection, the confidence interval of 95% had a lower bound of -.000006 and an upper bound of .0007 which meant that the null hypothesis for hypothesis 2B was rejected. Appendix I and J contain the regression output and Figure 8 includes the cubic curve estimations of each model.

Table 9

Polynomial Regression Results for Cubic Models

Criterion Variable	Adjusted R^2	R^2 Change	F-value Change	*df*	Significance Level	Regression Coefficient
Borman's Differential Accuracy	.03	.00	.33	1, 137	.57	-.0001
Elevation	.04	.03	3.78	1, 137	.05	.0003
Differential Elevation	.02	.00	.00	1, 137	.92	.00001
Stereotype Accuracy	.00	.00	.51	1, 137	.45	.00006
Cronbach's Differential Accuracy	.00	.00	.04	1, 137	.84	.00002

Statistical Power Analysis

After testing all of the hypotheses, the statistical power of each model was measured to determine the likelihood of rejecting the null hypothesis when it is false (Arnold et al., 2011). Field (2009) stated that values closer to one indicate a higher likelihood that the model will detect an effect assuming that one exists in the study's population. More specifically, power values above .80 are considered sufficient to detect significant effects (Field, 2009). G*Power statistical software was used to conduct a post hoc test for linear multiple regression with an exact distribution (Faul et al., 2007). SPSS input and output can be found in Appendix K. Summarised findings are displayed in Table 10 and indicate that the models with Borman's Differential Accuracy and Elevation Accuracy as the criterion variables demonstrated sufficient power to detect statistically significant effects, but the remaining models most often did not.

Figure 9

Cubic Curve Estimations

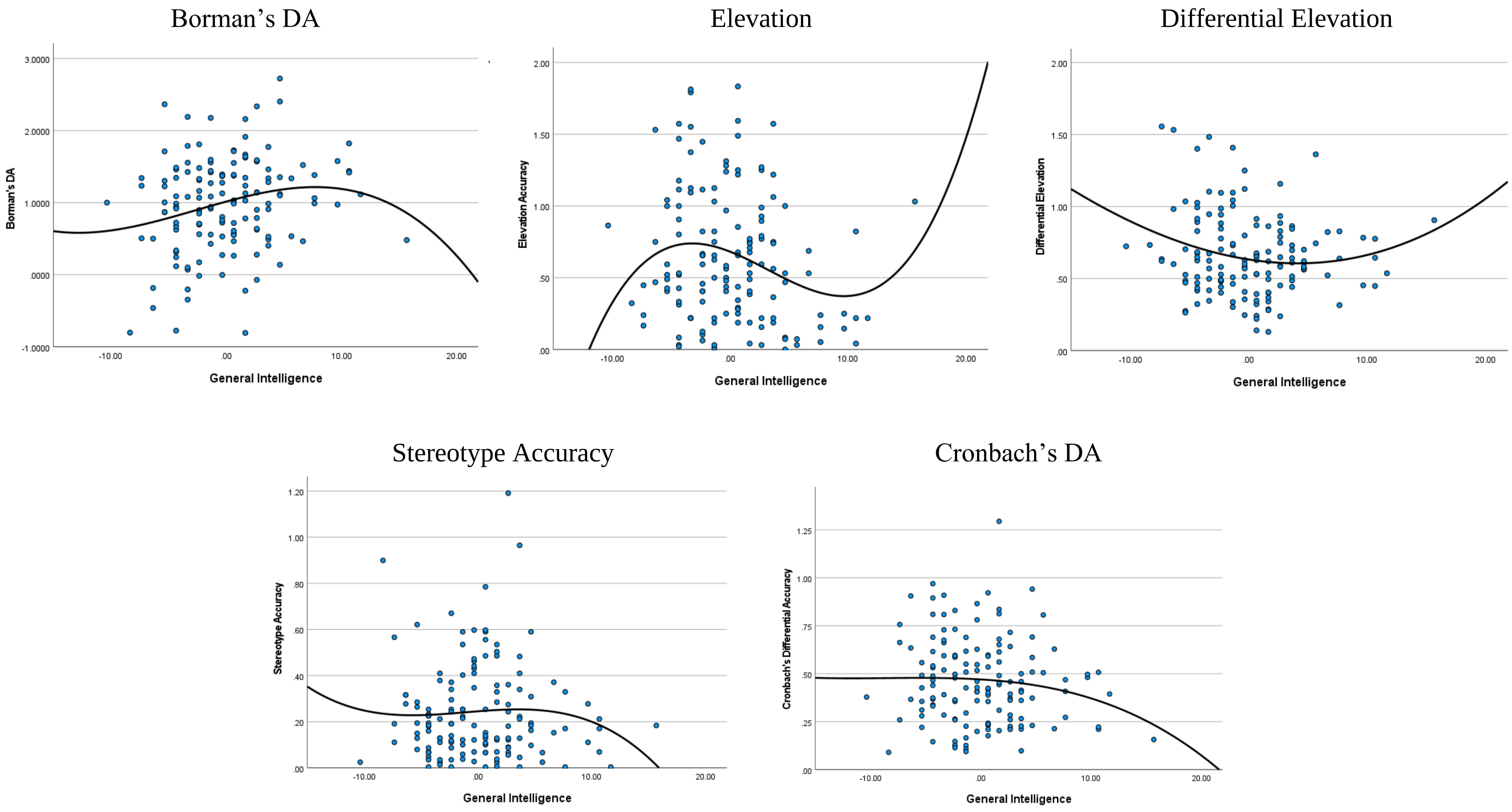


Table 10

Statistical Power of OLS Regression Models

Model	Statistical Power
Linear Models	
Borman's Differential Accuracy	.86
Elevation	.80
Differential Elevation	.60
Stereotype Accuracy	.36
Cronbach's Differential Accuracy	.60
Quadratic Models	
Borman's Differential Accuracy	.91
Elevation	.80
Differential Elevation	.86
Stereotype Accuracy	.60
Cronbach's Differential Accuracy	.72
Cubic Models	
Borman's Differential Accuracy	.91
Elevation	.94
Differential Elevation	.86
Stereotype Accuracy	.60
Cronbach's Differential Accuracy	.72

Chapter Summary

Correlations were found between Borman's Differential Accuracy (DA) score and most of Cronbach's accuracy indices. Although Borman's DA correlated significantly with IQ, none of Cronbach's measures did. Individuals who were younger and had higher levels of education scored higher on the GMA measure. In an attempt to test the predictiveness of a U-shaped and S-shaped model, a linear baseline was first established for comparison purposes. Using OLS regression analyses, linear associations were found with IQ between Borman's DA. No significant U-shaped (quadratic) or S-shaped (cubic) relationships were detected between the study variables. The following chapter will discuss the findings in more detail with reference to current person-perception literature.

Chapter 5: Discussion

Cognitive abilities appear to be the strongest individual predictor of how accurately judges' rate others in the HRM context (De Kock et al., 2020) given the cognitive demands associated with the cue detection and cue utilisation stages of Funder's (1995) RAM process (Christiansen et al., 2005). However, previous studies have been inconsistent in establishing whether judges with higher levels of intelligence produce more accurate scores at all levels of intelligence suggesting that differences in IQ become inconsequential or harmful after certain thresholds (Hauenstein & Alexander, 1991; Smither & Reilly, 1987). The interaction between IQ and accuracy could potentially be better represented by nonlinear models which remain unexplored in the person-perception literature (De Kock, 2015; De Kock et al., 2020).

The present study aimed to enhance our understanding of how judges' cognitive abilities informed their rating accuracy ability by considering the predictiveness of quadratic and cubic models compared to the traditional linear relationship. Using secondary data from a study conducted by De Kock et al. (2015), quadratic and cubic models were computed to determine which best predicts the relationship between intelligence and rating accuracy. Rating accuracy measures used included Borman's (1979) Differential Accuracy and Cronbach's (1955) four accuracy components. This chapter starts by restating the purpose of the study and highlights key findings from the previous chapter. It then discusses correlations between study variables and assesses support for linear, quadratic and cubic models. It also considers the implications for theory and practice; discusses the study limitations; and includes recommendations for future research.

Purpose of this Study

Past researchers have established a consistently strong link between general intelligence and rating accuracy (Borman & Hallam, 1991; Christiansen et al., 2005; Speer et al., 2019). For this reason, IQ is currently the best indicator available to researchers who wish to assess a judge's rating accuracy (De Kock et al., 2020). It is therefore concerning that the few previous studies that have reported interactions between IQ and rating accuracy in the HRM context, have almost always assumed an underlying linear relationship (Borman, 1979; Letzring et al., 2006; Lippa & Dietz, 2000) when contradictory findings have been reported by Hauenstein and Alexander (1991) and Smither and Reilly (1987). In our desire to expand our understanding of what makes a good judge in the HRM context it would be critical to assess this fundamental assumption of how IQ determines rating accuracy when empirical and theoretical support exists suggesting a more complex phenomenon than first assumed.

For this reason, the objective of the present study was to compare the predictiveness of a linear relationship between IQ and rating accuracy to that of nonlinear models to determine which IQ best explains variance in rating accuracy. The key findings of the present study are listed below:

- Judges' rating accuracy increases for all levels of intelligence supporting a linear relationship between IQ and rating accuracy
- There is no significant support for any quadratic (U-shaped) or cubic (S-shaped) effects suggesting that judges with higher levels of intelligence will always be more accurate

Correlations between Study Variables

Cognitive factors appeared to correlate with accurately assessing interviewees' abilities and traits. In accordance with studies from Borman (1979), Borman and Hallam (1991), Christiansen et al. (2005) and Speer et al. (2019), there was a positive, small-to-medium correlation between intelligence and rating accuracy. Findings, therefore, support the need for enhanced abilities to handle the cognitive load associated with storing, retrieving, and evaluating information used to judge others.

Surprisingly, Cronbach's Differential Accuracy (DA) was the only accuracy component not to significantly correlate with Borman's Differential Accuracy (DA). But, a review of two performance-rating studies by Sulsky and Balzer (1988) also revealed inconsistent correlations between Cronbach's accuracy components and Borman's DA. It can therefore be concluded that although Cronbach's DA and Borman's DA are most closely related conceptually, it is not unusual for these measures to have no correlation. Both speak to the ability to rank order targets based on set dimensions (Borman, 1979), but admittedly, Borman's DA is more broadly defined (Becker & Cardy, 1986). One should also consider how each accuracy measure calculates the distance between 'true' scores and ratings provided by judges. Borman's DA is an average correlation across all dimensions between ratings and true scores after Fisher's r-to-z transformation (Becker & Cardy, 1986). Whereas Cronbach's DA reflects the difference between ratings and true scores across each rater and each dimension. According to Becker and Cardy (1986), Cronbach's DA therefore reflects the "residual accuracy not reflected in the other three" accuracy components (p. 662).

Support for a Linear Relationship

Results indicated that a significant linear relationship exists between general intelligence and rating accuracy (specifically Borman's Differential Accuracy). Managers with higher levels of intelligence produced more accurate assessments of interviewees for all levels of intelligence. These findings are not surprising given that the prevailing assumption in the person-perception field is that this relationship is linear (Borman, 1979; Borman & Hallam, 1991; Lippa & Dietz, 2000; Speer et al., 2019). It can therefore be concluded that enhanced cognitive abilities assisted judges during the cue detection and utilization stages of the RAM process. Interestingly, no linear relationship was detected between Cronbach's accuracy components and IQ . The use of two accuracy operationalizations in this study worked to its benefit and allowed researchers to assess the stability of the construct outside of the measures used. This is highly recommended to future researchers hoping to replicate study findings.

Support for Nonlinear Relationships

Although person-perception literature supports the theory that more intelligent raters could become less accurate after reaching a specific intelligence threshold (Hauenstein & Alexander, 1991; Smither & Reilly, 1987), no evidence was found in the present study to support this claim. The addition of a squared IQ term did little to improve the predictiveness of rating accuracy. Findings did not support a U-shaped relationship where managers with higher levels of intelligence produced less accurate scores after a certain IQ level.

The Role of Task Complexity

In an attempt to explain the lack of significant findings, it might be useful to consider the role of task complexity. Smither and Reilly (1987) originally posited that raters at higher levels of intelligence found rating tasks mundane and therefore produced less accurate ratings compared to moderately intelligent raters. This was in response to finding a quadratic relationship between intelligence and rating accuracy. In a review of literature on what makes a good judge by De Kock et al. (2020), researchers noted that task complexity can impact the detection of behavioural cues. Put simply, "…intelligence may explain accuracy better in high-complexity tasks as compared to low-complexity tasks" (p. 9). It is argued that when the cue detection stage places a greater demand on raters' cognitive abilities, intelligence is expected to correlate more strongly with accuracy (Ambady & Rosenthal, 1992; Lippa & Dietz, 2000). For example, low-structure interviews produce more unique information and require enhanced mental abilities from a rater to elicit meaningful cues compared to high-structure interviews with less behavioural variation (De Kock et al., 2020; Funder, 2012). The same can be said for ratings in an assessment centre context where raters judge different targets simultaneously on various dimensions that require stronger information processing skills (De Kock et al., 2020).

It is therefore expected that variation in the rating context stimuli could highlight stronger differences between the accuracy scores of raters with different levels of intelligence. For example, the present study asked managers to view recordings of semi-structured interviews where candidates answered a list of questions. Participants could have experienced the task as more complex given the lack of opportunity to elicit cues from participants and not being in the same room as interviewees to observe and influence their behaviours. This would explain why judges with lower levels of intelligence produced lower accuracy scores compared to more intelligent judges. By varying task complexity, researchers could assess whether quadratic interactions are more likely to be present when highly intelligent individuals are disengaged and expend less time and effort in accurately assessing targets (Smither & Reilly, 1987).

Sampling Managers

Another consideration for a lack of a quadratic effect relates to the study sample which consisted primarily of experienced managers. It is possible that intelligent individuals who have been promoted to this position have displayed exemplary social and problem-solving skills protecting them against common pitfalls. Managers are often required to demonstrate strong communication skills which would assist in eliciting cues from targets. It is therefore recommended that future researchers continue to collect data from various rater types (e.g., managers, psychologists, and assessment centre assessors) to conclude whether it plays a role in moderating the effect of IQ on accuracy scores.

Intellectual Stratification

Adding a cubic term to detect an S-shaped curve, did not identify any nonlinear interactions either. Increases in IQ points did not have an increased impact on managers' accuracy compared to increases in IQ at low or high extremes. This finding corresponded with the only other study to test for this relationship by Smither and Reilly (1987) who also did not detect significant results. A possible explanation for the lack of cubic findings can include Simonton's (1985) argument of 'intellectual stratification'. This was put forward to explain how the optimal level of intelligence needed for an individual to exert influence over a group is determined by the intelligence of the group itself and not the population in general. It is pertinent to this study as it points to the possibility of intellectual stratification within subgroups in the population (like managers) where variance in intelligence is limited (Antonakis et al., 2017; Simonton, 1985).

One would expect to find more varied intelligence levels in the workplace since different occupations require different educational backgrounds (Simonton, 1985). However, since this sample is comprised of mid-level managers it is possible that this role attracts an intellectually homogeneous subgroup (Simonton, 1985). There were therefore not enough data points at the extreme ends in this sample to detect cubic effects. This would justify the need to replicate this study with employees who hold different positions with different employers to assess whether this finding exists outside of the current sample.

Theoretical Implications

The present study had three key theoretical implications for understanding the impact of general intelligence on rating accuracy discussed below.

Very few studies have primarily focused on the potential nonlinear relationship between general intelligence and rating accuracy in an attempt to expand the limited literature available. This study was the third of its kind to test for quadratic and cubic effects that could better account for changes in rating accuracy scores after reaching specific IQ thresholds (Hauenstein & Alexander, 1991; Smither & Reilly, 1987). It, therefore, contributes to the growing body of person-perception literature in South Africa and internationally (De Kock et al., 2020).

This is also one of the few studies to date using a sample of practitioners rather than students when assessing the effect of intelligence and rating accuracy. Outside of the original study conducted by De Kock et al. (2015), Davis (1999) was the only other researcher to report on findings from a sample of assessment centre assessors. Although researchers acknowledge there are instances where student sampling is valid and necessary, it can pose a problem with external validity (Henry, 2008). The present study therefore added to the limited research available describing the phenomenon in practice (De Kock et al., 2015). This is particularly useful since level of education is known to correlate strongly with IQ (Deary et al., 2007) such that students' IQ scores could be markedly different from practitioners in the field of HRM responsible for assessing the performance of potential of targets. Students enrolled at tertiary education institutions already represent IQ scores at the higher end of the population sample whereas practitioners with a diverse educational background could better represent scores across all IQ levels making it easier to detect nonlinear interactions (Antonakis et al., 2017).

Most importantly, it produced conclusive evidence to support the traditional linear relationship between rating accuracy in the context of the present study. Findings validate the important role that cognitive abilities play in the final stages of Funder's (1995) RAM process and confirm that higher levels of intelligence will always correspond with more accurate assessments of targets. This is in line with similar results reported by Christiansen et al. (2005), Lippa and Dietz (2000) and Speer et al. (2019).

Implications for Practice

Since accurate judgments are required to make decisions that support an organisation’s functioning (Christiansen et al., 2005), the present study has implications for practice. More specifically, how employers appropriately select accurate raters. Findings support the need for cognitive tests to screen raters based on their mental abilities considering the strong link with rating accuracy (De Kock et al., 2020). General intelligence remains the best predictor of rating accuracy in person-perception literature, and the present study confirmed that more intelligent raters will produce more accurate ratings at all levels of intelligence (De Kock et al., 2020).

Study Limitations

Although the present study added to the existing person-perception body of knowledge, it is not without its constraints. Limitations associated with using secondary data and the design of the primary study are discussed below.

Using a Secondary Data Source

When using data collected in a previous study, there are limitations inherent to the process. By using a secondary dataset, the researcher is not directly involved in collecting and capturing the data or determining the methodological approach employed (Andersen et al., 2011). This limits control over the quality of data and can influence the data analysis process if researchers in the original study followed poor data collection and capturing practices (Hox & Boeije, 2005). Since the primary study conducted by De Kock et al. (2015) had clearly defined variables, limited missing values and no coding errors, data management and analyses were easy to conduct. Data were also stored with restricted access and peer-reviewed reducing potential quality concerns (Hox & Boeije, 2005).

Data Collection Timeframe

The primary data were collected in 2011 by De Kock et al. (2015) – over a decade ago. In this study, it is not expected to negatively impact the results as the variables are not directly related to the timeframe during which they were measured. In other words, there is no reason to expect that the population’s level of intelligence or rating accuracy has significantly changed over time to justify the need for newly collected data.

Materials

The use of video-recorded interviews standardized participants’ exposure to stimuli which allowed researchers to minimize error variance and produce true expert scores (Chung et al., 2019). In the original study, De Kock et al. (2015) argued that although realistic conditions were created for interviewees who used the opportunity to prepare for ‘real’ interviews, recordings could not substitute natural interactions between raters or interviewers and interviewees. A need was expressed for future researchers to consider the stimuli used and whether video recordings (instead of face-to-face interviews) could influence accuracy results (De Kock et al., 2015).

External Validity

Most studies that have considered the relationship between intelligence and rating accuracy, collected data from university students rather than practitioners in an HRM setting (see Table 1). For this reason, De Kock et al. (2020) argued that the overuse of university or college samples could be distorting the observed correlations between rating accuracy and intelligence. The original study was therefore unique in examining data from managers in the field but similarly presented issues around external validity. Given that the sample was comprised of South African managers working for the same employer, further research would be required to determine whether results can be generalized to other South African organisations or international populations. For instance, Baranski et al. (2017) found that cultural variations exist between countries which could influence participants’ ability to detect and utilize behavioural cues. Study results should also not be generalized to nonmanagerial judges like psychologists and trained assessment centre assessors as suggested by De Kock et al. (2015). This is because the original study collected data from managers rating interviewees and there is evidence that rater types differ in the ratings they produce (Sagie & Magnezy, 1997). It is therefore recommended that future studies consider intelligence and rater types when choosing an appropriate sample.

Sample Size

Although there were several benefits to reanalysing the data from a peer-reviewed study, the sample size was out of the researcher’s control. There might have been an adequate number of participants for the primary study, but it did pose a problem for the present study that reported lower statistical power than ideal. The models containing differential elevation, stereotype accuracy and Cronbach's differential accuracy are therefore not considered definitive as they did not have sufficient power to detect any effects represented in the study population. Any conclusions drawn from these models are inconclusive until further research is conducted with a larger sample size.

Recommendations for Future Research

Although the present study confirmed the existence of a linear relationship between intelligence and accuracy, it also highlighted the need for additional research to advance our understanding of these constructs. Firstly, replicating this study with intelligence and accuracy as variables using a substantially larger sample size would allow for better detection of potential nonlinear effects. A larger sample size is expected to increase the statistical power of both accuracy measures and include raters with more varied IQ scores (Borman & Hallam, 1991; Simonton, 1985).

Secondly, there is value in exploring moderator effects like rater type or stimuli that influence the relationship between intelligence and accuracy. Since most studies have sampled student populations, there is a need to examine how intelligence impacts accuracy among different rater groups (De Kock et al., 2020). It is possible that interviewers, managers, and assessment centre assessors may differ in their levels of intelligence and further studies could examine how stable IQ is across each sample. Researchers should also vary the complexity of rating tasks to include stimuli that elicit poor and rich cues. This would indicate whether highly complex tasks eliciting poor cues from targets require more advanced intellect to detect and utilize cues. Compared to less complex tasks in cue rich environments where less intelligent individuals can produce highly accurate assessments (De Kock et al., 2020).

Conclusion

The purpose of this study was to examine how increases in general intelligence (IQ) relate to judgement accuracy among raters. More specifically, it set out to determine whether a nonlinear model could better predict the relationship between intelligence and rating accuracy than a linear model. This was a gap identified in the person-perception literature where previous studies assumed that the interaction was always linear such that more intelligent raters would always produce more accurate ratings (Borman, 1979; Christiansen et al., 2005; Lippa & Dietz, 2000). Nonlinear findings from Hauenstein and Alexander (1991) and Smither and Reilly (1987) highlighted the need for renewed attention given that intelligence remains the best predictor of accuracy among individual differences between raters (De Kock et al., 2020).

Using ordinary least squares regression analyses, data were reanalysed from a study conducted by De Kock et al. (2015) with 146 South African managers. The predictiveness of a baseline linear model was computed and compared with quadratic and cubic models to determine which statistical approach explained the most variance in rating accuracy scores. Results did not provide support for quadratic or cubic models but did validate the existence of a linear relationship.

Study findings and limitations were discussed along with directions for future research. Implications include further evidence to support our current understanding of how IQ and accuracy interact in practice. Findings should also guide how employers screen and select employees responsible for producing accurate ratings.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Allport, G. W. (1937). *Personality: A psychological interpretation*. Henry Holt & Co.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*(2), 256. https://doi.org/10.1037/0033-2909.111.2.256

American Psychological Association. (2017). *Ethical Principles of Psychologists and Code of Conduct* (2002, amended effective June 1, 2010, and January 1, 2017), Issue. https://www.apa.org/ethics/code/

Andersen, J. P., Prause, J., & Silver, R. C. (2011). A step-by-step guide to using secondary data for psychological research. *Social and Personality Psychology Compass*, *5*(1), 56-75. https://doi.org/10.1111/j.1751-9004.2010.00329.x

Antonakis, J., House, R. J., & Simonton, D. K. (2017). Can super smart leaders suffer from too much of a good thing? The curvilinear effect of intelligence on perceived leadership behavior. *Journal of Applied Psychology, 102*(7), 1003. https://doi.org/10.1037/apl0000221

Antwi, S. K., & Hamza, K. (2015). Qualitative and quantitative research paradigms in business research: A philosophical reflection. *European Journal of Business and Management*, *7*(3), 217-225.

Arnold, B. F., Hogan, D. R., Colford, J. M., & Hubbard, A. E. (2011). Simulation methods to estimate design power: An overview for applied research. *BMC Medical Research Methodology*, *11*(1), 1-10. https://doi.org/10.1186/1471-2288-11-94

Baranski, E. N., Gardiner, G., Guillaume, E., Aveyard, M., Bastian, B., Bronin, I., Ivanova, C., Cheng, J. T., Kock, F. S. d., & Denissen, J. J. (2017). Comparisons of daily behavior across 21 countries. *Social Psychological and Personality Science, 8*(3), 252-266. https://doi.org/10.1177/1948550616676879

Baron, R. A. (1986). Self-presentation in job interviews: When there can be “too much of a good thing. *Journal of Applied Social Psychology, 16*(1), 16-28. https://doi.org/10.1111/j.1559-1816.1986.tb02275.x

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, *19*(1), 3-11. https://doi.org/10.2466/pr0.1966.19.1.3

Becker, B. E., & Cardy, R. L. (1986). Influence of halo error on appraisal effectiveness: A conceptual and empirical reconsideration. *Journal of Applied Psychology, 71*(4), 662-671.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence*, *20*(3), 309-328. https://doi.org/10.1016/0160-2896(95)90013-6

Bobko, P. (2001). Expanding the regression repertoire: polynomial and interaction terms. In *Correlation and regression* (2nd ed., pp. 207-238). SAGE Publications, Inc. https://doi.org/10.4135/9781412983815.n9

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.

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

Borman, W. C., & Hallam, G. L. (1991). Observation accuracy for assessors of work-sample performance: Consistency across task and individual-differences correlates. *Journal of Applied Psychology*, *76*(1), 11-18. https://doi.org/10.1037/0021-9010.76.1.11

Brecker, N. (1989). *The effects of rater training, environmental complexity, cognitive complexity and rater intelligence on performance appraisal accuracy* [Thesis, Stevens Institute of Technology].

Brooks, J. O., & Brooks, L. L. (1990). A BASIC program for computing Cronbach’s accuracy components. *Behavior Research Methods, Instruments, & Computers*, *22*(4), 413-416. https://doi.org/10.3758/BF03203184

Brown, M., Wai, J., & Chabris, C. (2020). Can You Ever Be Too Smart for Your Own Good? Linear and Nonlinear Effects of Cognitive Ability. *Perspectives on Psychological Science*, *16*(6), 1337 –1359.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.

Campbell, B. A., Coff, R., & Kryscynski, D. (2012). Rethinking sustained competitive advantage from human capital. *Academy of Management Review*, *37*(3), 376-395. 10.5465/amr.2010.0276

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22. https://doi.org/10.1037/h0046743

Chen, R., Rafaeli, E., Bar-Kalifa, E., Gilboa-Schechtman, E., Lutz, W., & Atzil-Slonim, D. (2018). Moderators of congruent alliance between therapists and clients: A realistic accuracy model. *Journal of Counseling Psychology*, *65*(6), 1-12. https://doi.org/10.1037/cou0000285

Chirico, K. E., Buckley, M. R., Wheeler, A. R., Facteau, J. D., Bernardin, H. J., & Beu, D. S. (2004). A note on the need for true scores in frame-of-reference (FOR) training research. *Journal of Managerial Issues, 16*(3), 382-395.

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

Chung, A. S., Shah, K. H., Bond, M., Ardolic, B., Husain, A., Li, I., Cygan, L., Caputo, W., Shoenberger, J., & van Dermark, J. (2019). How well does the standardized video interview score correlate with traditional interview performance? *Western Journal of Emergency Medicine*, *20*(5), 726-730. https://doi.org/10.5811/westjem.2019.7.42731

Cohen, J. (1988). *Statistical power analysis for the social sciences*. Lawrence Erlbaum Associates.

Cohen, J. (1994). The earth is round (p<. 05). *American Psychologist*, *49*(12), 997-1003. https://doi.org/10.1037/0003-066X.49.12.997

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). Quantitative scales, curvilinear relationships, and transformations. In L. E. Associates (Ed.), *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed., pp. 193-254). Taylor & Francis Group.

Coleman, G. D., Koelling, C. P., & Geller, E. S. (2001). Training and scoring accuracy of organisational self-assessments. *International Journal of Quality & Reliability Management, 18*(5), 512-527. https://doi.org/10.1108/02656710110392827

Colvin, C. R., & Bundick, M. J. (2001). In search of the good judge of personality: Some methodological and theoretical concerns. In J. A. Hall & F. J. Bernieri (Eds.), *The LEA series in personality and clinical psychology* (pp. 47–65). Lawrence Erlbaum Associates Publishers.

Conway, A. R., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in cognitive sciences, 7*(12), 547-552. https://doi.org/10.1016/j.tics.2003.10.005

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

Daoud, J. I. (2017). Multicollinearity and regression analysis. Journal of Physics: Conference Series, Kuala Lumpur, Malaysia.

Davis, M. E. (1999). *Influence of assessor individual differences on rating errors and rating accuracy in assessment centers* [Dissertation, University of Nebraska].

Davis, M. H., & Kraus, L. A. (1997). Personality and empathic accuracy. In W. J. Ickes (Ed.), *Empathic accuracy* (pp. 144-168). The Guilford Press.

De Kock, F. (2015). *Individual differences in judgment accuracy in personnel selection: What makes the 'good judge'?* [Dissertation, Erasmus University Rotterdam]. Netherlands. hdl.handle.net/1765/79493

De Kock, F., Lievens, F., & Born, M. (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 Kock, F., 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

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13-21. https://doi.org/10.1016/j.intell.2006.02.001

Deffenbacher, K. A., & Hamm, N. H. (1972). An application of Brunswik's lens model to developmental changes in probability learning. *Developmental Psychology, 6*(3), 508-519.

Disatnik, D., & Sivan, L. (2016). The multicollinearity illusion in moderated regression analysis. *Marketing Letters, 27*(2), 403-408. https://doi.org/10.1007/s11002-014-9339-5

Dodrill, C. B. (1981). An economical method for the evaluation of general intelligence in adults. *Journal of Consulting and Clinical Psychology, 49*(5), 668-673. https://doi.org/10.1037/0022-006X.49.5.668

Du Prel, J.-B., Hommel, G., Röhrig, B., & Blettner, M. (2009). Confidence interval or p-value? Part 4 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International, 106*(19), 335-339. https://doi.org/10.3238/arztebl.2009.0335

Elliott, E., Barton, B., & Peat, J. (2008). Correlation and Regression. In *Statistics Workbook for Evidence-Based Health Care* (pp. 93-109). Wiley-Blackwell.

Embretson, S. E., & McCollam, K. M. S. (2000). Psychometric approaches to understanding and measuring intelligence. In R. Sternberg (Ed.), *Handbook of intelligence* (pp. 423–444). Cambridge University Press.

Engelhard Jr, G. (1996). Evaluating rater accuracy in performance assessments. *Journal of Educational Measurement, 33*(1), 56-70. https://doi.org/10.1111/j.1745-3984.1996.tb00479.x

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*(2), 175-191. https://doi.org/10.3758/BF03193146

Ferguson, L. (2004). External validity, generalizability, and knowledge utilization. *Journal of Nursing Scholarship, 36*(1), 16-22.

Field, A. (2009). *Discovering statistics using SPSS (3rd ed.)*. SAGE publications.

Flatt, C., & Jacobs, R. L. (2019). Principle assumptions of regression analysis: Testing, techniques, and statistical reporting of imperfect data sets. *Advances in Developing Human Resources, 21*(4), 484-502.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*(4), 652-670. https://doi.org/10.1037/0033-295X.102.4.652

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. Academic Press.

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

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods, 2*(3), 235-247. https://doi.org/10.1037/1082-989X.2.3.235

George, E. (2006). *Interviewer accuracy across levels of structure in the employment interview* [Dissertation, Colorado State University]. Fort Collins, Colorado.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism, 10*(2), 486-489. https://doi.org/10.5812/ijem.3505

Gill, A., & Brajer, V. (2012). Wonderlic, race, and the NFL draft. *Journal of Sports Economics, 13*(6), 642-653. https://doi.org/10.1177/1527002511429575

Gilmore, R. O., Kennedy, J. L., & Adolph, K. E. (2018). Practical solutions for sharing data and materials from psychological research. *Advances in Methods and Practices in Psychological Science*, *1*(1), 121-130. https://doi.org/10.1177/2515245917746500

Greiff, S., & Neubert, J. C. (2014). On the relation of complex problem solving, personality, fluid intelligence, and academic achievement. *Learning and Individual Differences, 36*, 37-48. https://doi.org/10.1016/j.lindif.2014.08.003

Guilford, J. P. (1967). *The nature of human intelligence*. McGraw-Hill.

Guion, R. M., & Gibson, W. M. (1988). Personnel selection and placement. *Annual Review of Psychology*, *39*(1), 349-374.

Hauenstein, N. M., & Alexander, R. A. (1991). Rating ability in performance judgments: The joint influence of implicit theories and intelligence. *Organizational Behavior and Human Decision Processes*, *50*(2), 300-323. https://doi.org/10.1016/0749-5978(91)90024-N

Henry, P. J. (2008). Student sampling as a theoretical problem. *Psychological Inquiry*, *19*(2), 114-126. https://doi.org/10.1080/10478400802049951

Hernández-Orallo, J., Dowe, D. L., & Hernández-Lloreda, M. V. (2014). Universal psychometrics: Measuring cognitive abilities in the machine kingdom. *Cognitive Systems Research*, *27*, 50-74. https://doi.org/10.1016/j.cogsys.2013.06.001

Horn, J. L., & Cattell, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta psychologica*, *26*, 107-129. https://doi.org/10.1016/0001-6918(67)90011-X

Hox, J. J., & Boeije, H. R. (2005). Data collection, primary versus secondary. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (pp. 593-599). Elsevier Science.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence, 3*(2), 105-120. https://doi.org/10.1016/0160-2896(79)90009-6

Iacobucci, D., Schneider, M. J., Popovich, D. L., & Bakamitsos, G. A. (2016). Mean centering helps alleviate "micro" but not "macro" multicollinearity. *Behavior Research Methods, 48*(4), 1308-1317.

Janovics, J. E. (2003). *Knowing thyself: The influence of dispositional intelligence on self-rating accuracy* [Dissertation, Central Michigan University].

Jensen, A. R. (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence, 17*(1), 47-77. https://doi.org/10.1016/0160-2896(93)90039-8

Johnson, W., Bouchard Jr, T. J., Krueger, R. F., McGue, M., & Gottesman, I. I. (2004). Just one g: Consistent results from three test batteries. *Intelligence, 32*(1), 95-107. https://doi.org/10.1016/S0160-2896(03)00062-X

Johnston, M. P. (2017). Secondary data analysis: A method of which the time has come. *Qualitative and Quantitative Methods in Libraries, 3*(3), 619-626.

Kassim, N. L. A. (2011). Judging behaviour and rater errors: An application of the many-facet Rasch model. *GEMA Online® Journal of Language Studies, 11*(3), 179-197.

Kelley, K., & Maxwell, S. E. (2010). Multiple regression. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 282-297). Routledge.

Kline, P. (1999). *The Handbook of Psychological Testing* (2nd ed.). Routledge.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121.

Kuncel, N. R., & Highhouse, S. (2011). Complex predictions and assessor mystique. *Industrial and Organizational Psychology, 4*(3), 302-306. https://doi.org/10.1111/j.1754-9434.2011.01343.x

Kyllonen, P. C. (1993). Aptitude testing inspired by information processing: A test of the four-sources model. *The Journal of General Psychology, 120*(3), 375-405. https://doi.org/10.1080/00221309.1993.9711154

Kyllonen, P. C. (1996). Is working memory capacity Spearman's g? In I. Dennis & P. Tapsfield (Eds.), *Human abilities: Their nature and measurement* (pp. 49-75). Lawrence Erlbaum Associates, Inc.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*(4), 389-433. https://doi.org/10.1016/S0160-2896(05)80012-1

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, *11*(4), 815-852. https://doi.org/10.1177/1094428106296642

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

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, *91*(1), 111-123. https://doi.org/10.1037/0022-3514.91.1.111

Levin, K. A. (2006). Study design III: Cross-sectional studies. *Evidence-Based Dentistry, 7*(1), 24-25. https://doi.org/10.1038/sj.ebd.6400375

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology*, *91*(2), 247. https://doi.org/10.1037/0021-9010.91.2.247

Linden, W., Ellis, A. T., & Millman, R. (2010). Deception in stress reactivity and recovery research. *International Journal of Psychophysiology, 75*(1), 33-38. https://doi.org/10.1016/j.ijpsycho.2009.10.012

Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior, 24*(1), 25-43.

Long, R. G. (2008). The crux of the method: Assumptions in ordinary least squares and logistic regression. *Psychological Reports*, *103*(2), 431-434. https://doi.org/10.2466/pr0.103.2.431-434

Matthews, T. D., & Lassiter, K. S. (2007). What does the wonderlic personnel test measure? *Psychological Reports, 100*(3), 707-712. https://doi.org/10.2466/pr0.100.3.707-712

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10. https://doi.org/10.1016/j.intell.2008.08.004

Mendoza, J. L. (1993). Fisher transformations for correlations corrected for selection and missing data. *Psychometrika*, *58*(4), 601-615. https://doi.org/10.1007/BF02294830

Miller, S., & Kirlik, A. (2006). Modeling the task environment: Act-R and the lens model. Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting, Los Angeles.

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., & Davidshofer, C. O. (1988). *Psychological testing: Principles and application* (4 ed.). Prentice Hall.

Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzer, W. K. (1982). Relationship between observational accuracy and accuracy in evaluating performance. *Journal of Applied Psychology*, *67*(3), 320-325. https://doi.org/10.1037/0021-9010.67.3.320

Neumayer, E., & Pluemper, T. (2020). The concept and measurement of robustness. *Available at SSRN 3513805*. https://doi.org/10.2139/ssrn.3513805

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Osborne, J. W. (2017). *Regression & linear modeling: Best practices and modern methods*. Sage Publications.

Osborne, J. W., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research, and Evaluation*, *8*(1), 1-5.

Parsons, M. A., Duerr, R., & Minster, J. B. (2010). Data citation and peer review. *Eos, Transactions American Geophysical Union*, *91*(34), 297-304.

Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management*, *39*(2), 313-338. https://doi.org/10.1177/0149206311410060

Powell, D. M. (2007). *Assessing personality in the employment interview: The impact of rater training and individual differences in rating accuracy* [Monograph, The University of Western Ontario]. London, Ontario, Canada.

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

Randall, R., & Sharples, D. (2012). The impact of rater agreeableness and rating context on the evaluation of poor performance. *Journal of Occupational and Organizational Psychology, 85*(1), 42-59. 10.1348/2044-8325.002002

Robie, C., Christiansen, N. D., Bourdage, J. S., Powell, D. M., & Roulin, N. (2020). Nonlinearity in the relationship between impression management tactics and interview performance. *International Journal of Selection and Assessment, 28*(4), 522-530. https://doi.org/10.1111/ijsa.12307

Robie, C., & Ryan, A. M. (1999). Effects of nonlinearity and heteroscedasticity on the validity of conscientiousness in predicting overall job performance. *International Journal of Selection and Assessment, 7*(3), 157-169. https://doi.org/10.1111/1468-2389.00115

Sagie, A., & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology, 70*(1), 103-108. https://doi.org/10.1111/j.2044-8325.1997.tb00634.x

Salgado, J. F. (2017). Using ability tests in selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell Handbook of the Psychology of Recruitment, Selection and Employee Retention* (pp. 115–150). Wiley-Blackwell. https://doi.org/10.1002/9781118972472.ch7

Salvemini, N. J., Reilly, R. R., & Smither, J. W. (1993). The influence of rater motivation on assimilation effects and accuracy in performance ratings. *Organizational Behavior and Human Decision Processes*, *55*(1), 41-60. https://doi.org/10.1006/obhd.1993.1023

Schmidt, F. L., & Hunter, J. E. (1993). Tacit knowledge, practical intelligence, general mental ability, and job knowledge. *Current Directions in Psychological Science*, *2*(1), 8-9. https://doi.org/10.1111/1467-8721.ep10770456

Seber, G., & Wild, C. (1989). *Nonlinear regression*. John Wiley & Sons, Inc. https://doi.org/10.1002/0471725315

Sharpe, D., & Faye, C. (2009). A second look at debriefing practices: Madness in our method? *Ethics & Behavior*, *19*(5), 432-447. https://doi.org/10.1080/10508420903035455

Simonton, D. K. (1985). Intelligence and personal influence in groups: Four nonlinear models. *Psychological Review, 92*(4), 532.

Smither, J. W., & Reilly, R. R. (1987). True intercorrelation among job components, time delay in rating, and rater intelligence as determinants of accuracy in performance ratings. *Organizational Behavior and Human Decision Processes, 40*(3), 369-391. https://doi.org/10.1016/0749-5978(87)90022-7

Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*, 88-101. https://doi.org/10.2307/1412159

Speer, A. B., Christiansen, N. D., & Laginess, A. J. (2019). Social intelligence and interview accuracy: Individual differences in the ability to construct interviews and rate accurately. *International Journal of Selection and Assessment, 27*(2), 104-128. https://doi.org/10.1111/ijsa.12237

Stadler, M., Becker, N., Gödker, M., Leutner, D., & Greiff, S. (2015). Complex problem solving and intelligence: A meta-analysis. *Intelligence, 53*, 92-101.

Sullivan, G. M., & Feinn, R. (2012). Using effect size—or why the P value is not enough. *Journal of graduate medical education, 4*(3), 279-282. https://doi.org/10.4300/JGME-D-12-00156.1

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

Sulsky, L. M., & Keown, J. L. (1998). Performance appraisal in the changing world of work: Implications for the meaning and measurement of work performance. *Canadian Psychology, 39*(52), 497-506. https://doi.org/10.1037/h0086794

Tabachnick, B., & Fidell, L. (2013). *Using multivariate statistics (6th ed.)*. Pearson.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2019). *Using multivariate statistics* (7 ed.). Pearson.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric monographs, 1*.

Vicente, K. J. (2003). Beyond the lens model and direct perception: Toward a broader ecological psychology. *Ecological Psychology, 15*(3), 241-267. https://doi.org/10.1207/S15326969ECO1503_4

Walfish, S. (2006). A review of statistical outlier methods. *Pharmaceutical Technology, 30*(11), 82-86.

Wiggins, J. S. (1973). *Personality and prediction: Principles of personality assessment*. Addison-Wesley.

Williams, C. (2007). Research methods. *Journal of Business & Economics Research, 5*(3), 65-72.

Wonderlic Inc. (2002). *Wonderlic Personnel Test & Scholastic Level Exam User's Manual*. Wonderlic, Inc.

Wood, R. E., & Marshall, V. (2008). Accuracy and effectiveness in appraisal outcomes: the influence of self-efficacy, personal factors and organisational variables. *Human Resource Management Journal, 18*(3), 295-313. https://doi.org/10.1111/j.1748-8583.2008.00067.x

Woodley, M. A. (2010). Are high-IQ individuals deficient in common sense? A critical examination of the 'clever sillies' hypothesis. *Intelligence, 38*(5), 471-480. https://doi.org/10.1016/j.intell.2010.06.002

Young, D. S. (2017). Multiple Linear Regression. In *Handbook of regression methods* (1st ed., pp. 85-108). CRC Press. https://doi.org/10.1201/9781315154701

Zizai, C. (2016). *The Two Sigma Conjecture: The Meaning of IQ and the Reason Why Ultra-High IQ Leads to Communication Barriers*. Hwa Chong Institution.

Appendix A

Approval from the Faculty of Commerce Research Ethics Committee

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: jacques.rousseau@uct.ac.za
Internet: www.uct.ac.za
@Commerce UCT **UCT Commerce Faculty Office**

30 08 2021

Marizanne Schade

School of Management Studies

University of Cape Town

REF: REC 2021/08/028

When More is not Better: Understanding the Potential Curvilinear Relationship between Intelligence and Rating Accuracy

We are pleased to inform you that your ethics application has been approved. Unless otherwise specified this ethical clearance is valid until 31-Dec-2022 .

Your clearance 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.

2021.08.30
15:32:21 +02'00'

Jacques Rousseau
Commerce Research Ethics Chair
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: jacques.rousseau@uct.ac.za
Website: http://www.commerce.uct.ac.za/com/Ethics-in-Research

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

Appendix B

SPSS Steps and Syntax for Centering Predictor Variable Around Mean

The following steps were executed in SPSS to center the predictor variable (general intelligence) around the mean as guided by Cohen et al. (2003).

Table D1

SPSS Steps and Syntax for Centering Predictor Variable Around Mean

Step	SPSS Syntax
1. Calculate the mean of the predictor value	`DESCRIPTIVES VARIABLES=gma /STATISTICS=MEAN STDDEV MIN MAX.`
2. Create a variable where the previously calculated mean is subtracted from the predictor value	`COMPUTE GMA_Centered=gma - 12.354610. EXECUTE.`
3. Check that the mean of the new variable is exactly equal to zero	`DESCRIPTIVES VARIABLES=GMA_centered /STATISTICS=MEAN STDDEV MIN MAX.`

Appendix C

Scatter Plots of Residuals and Predicted Values

Figure C1

IQ and Borman's Differential Accuracy

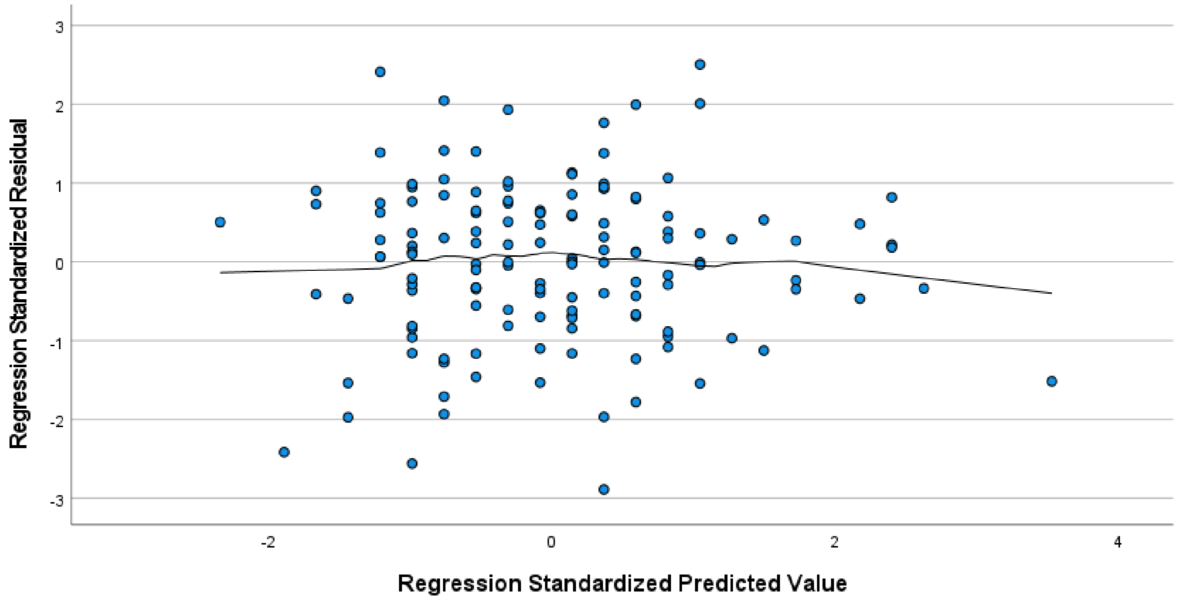


Figure C2

IQ and Elevation

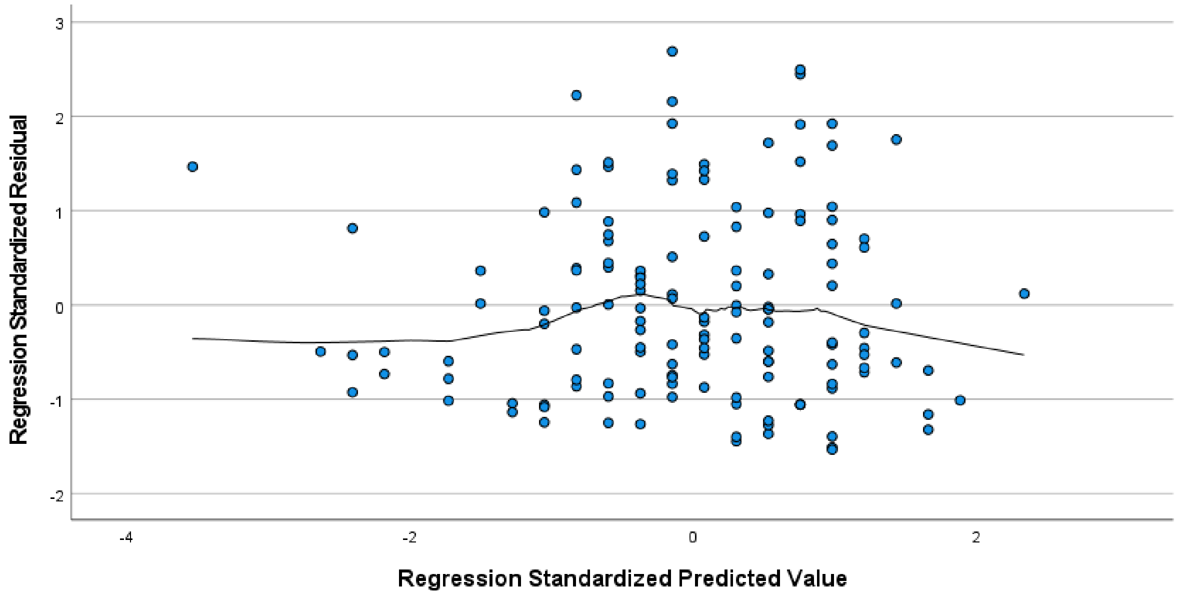


Figure C3

IQ and Differential Elevation

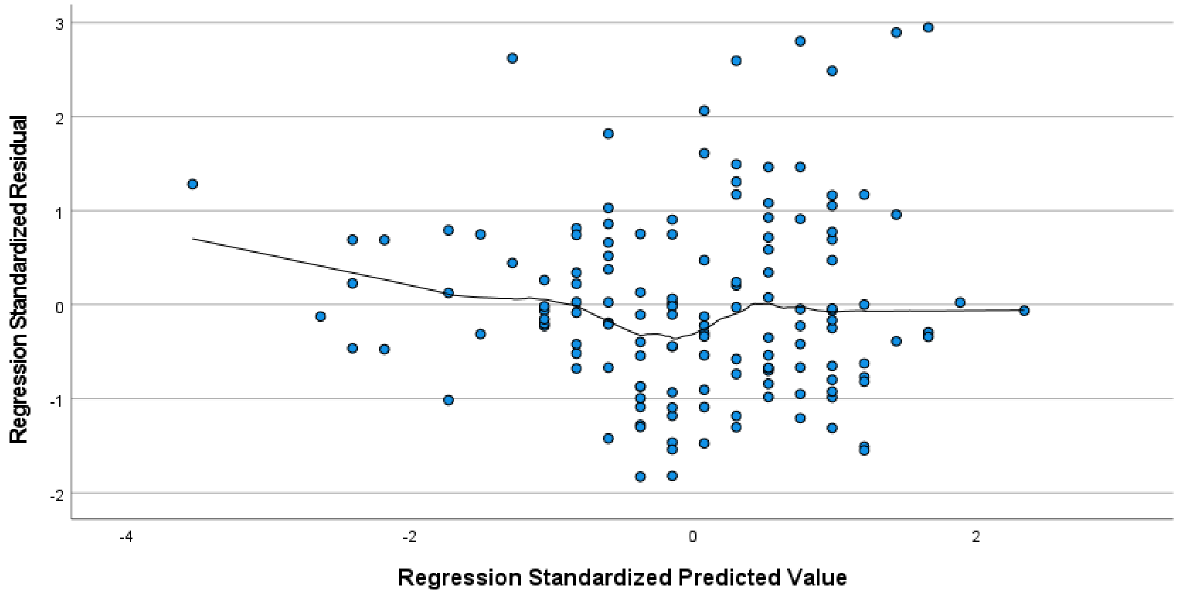


Figure C4

IQ and Stereotype Accuracy

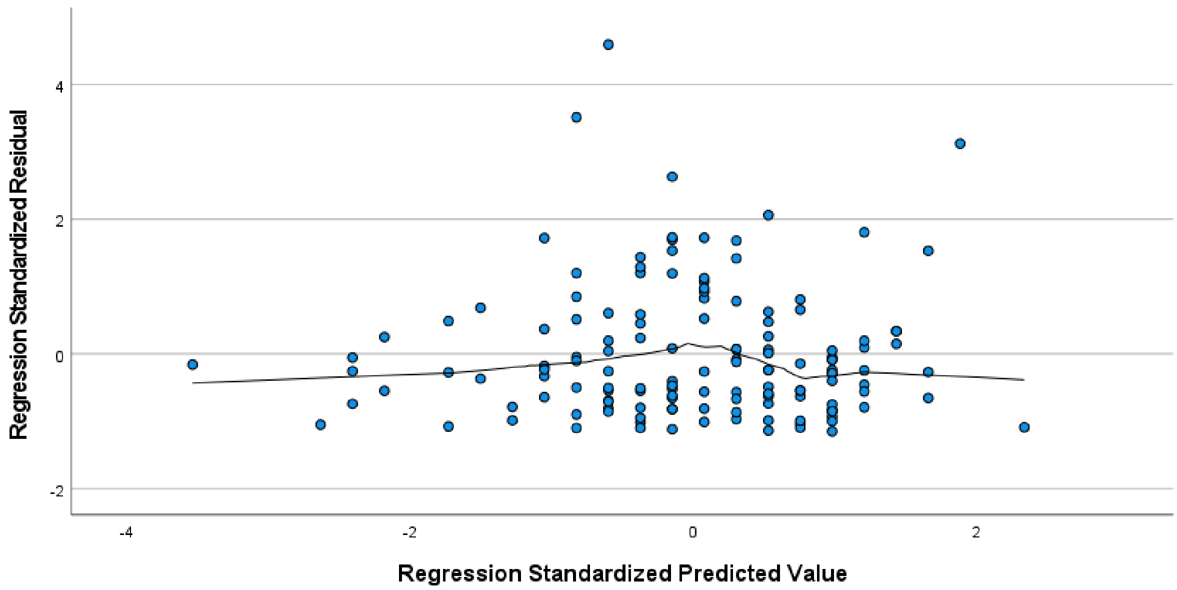
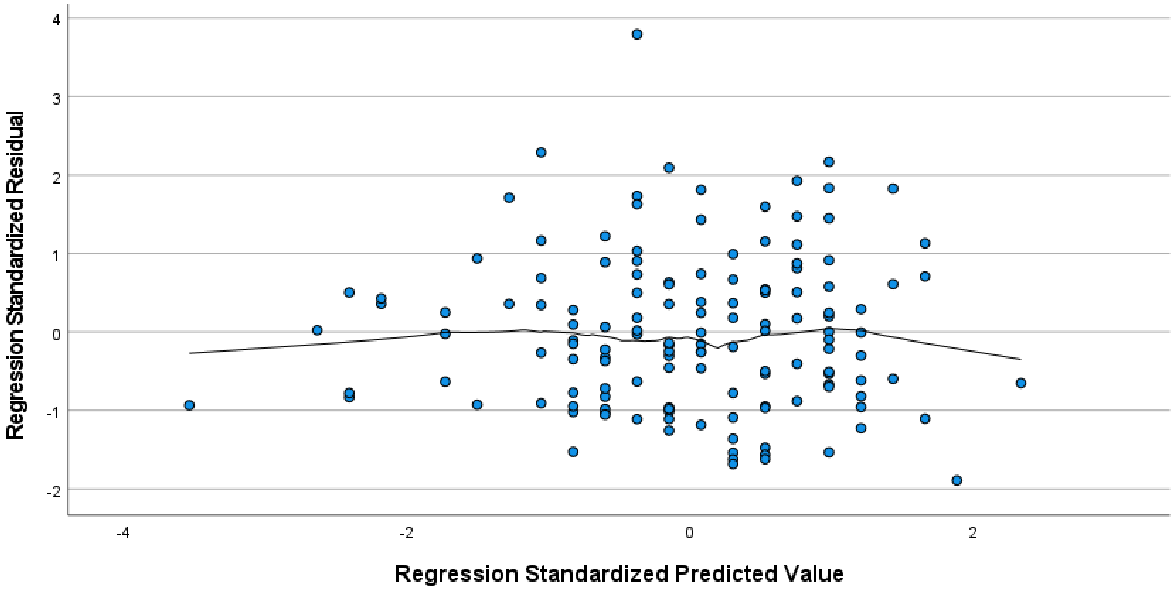


Figure C5

IQ and Cronbach's Differential Accuracy



Appendix D

Normal P-P Plot of Regression Standardized Residual Errors

Figure D1

IQ and Borman's Differential Accuracy

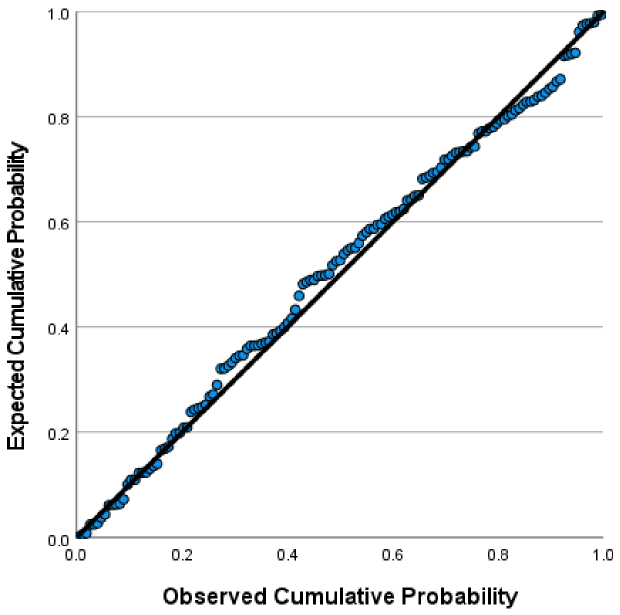


Figure D2

IQ and Elevation

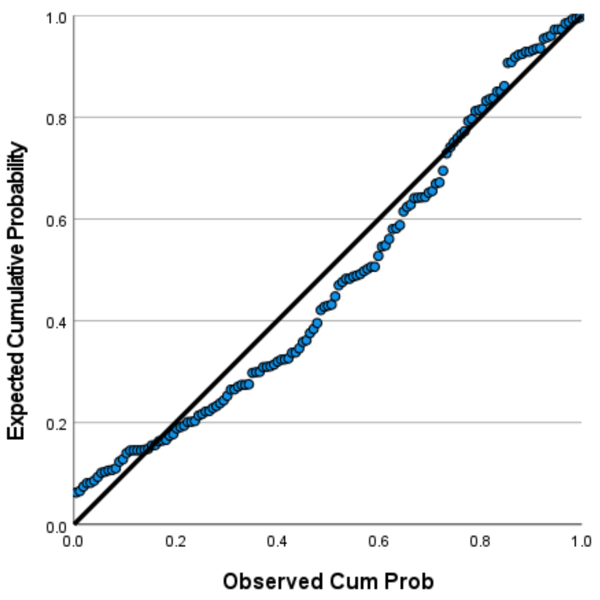


Figure D3

IQ and Differential Elevation

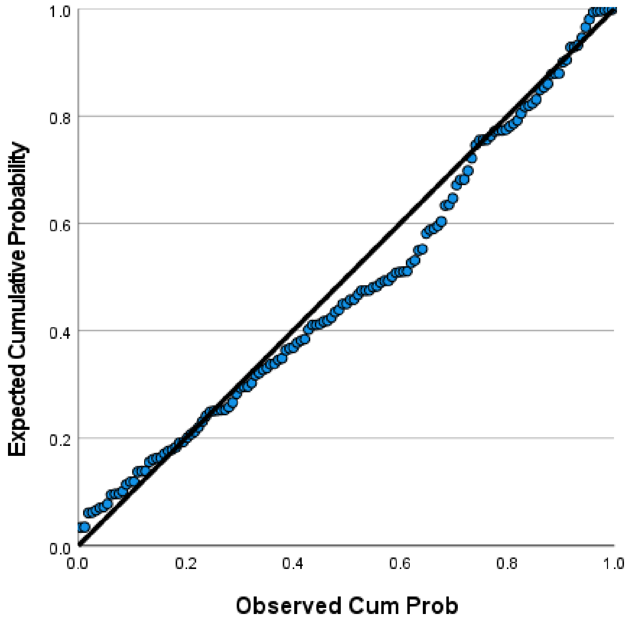


Figure D4

IQ and Stereotype Accuracy

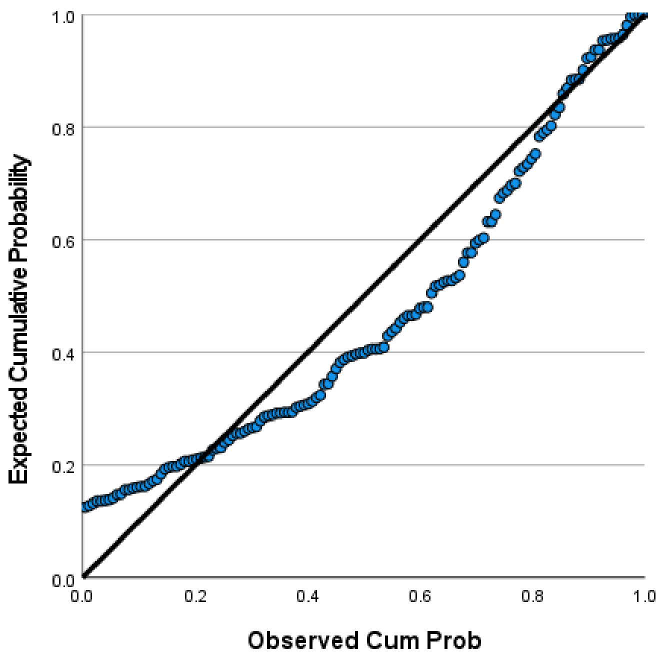
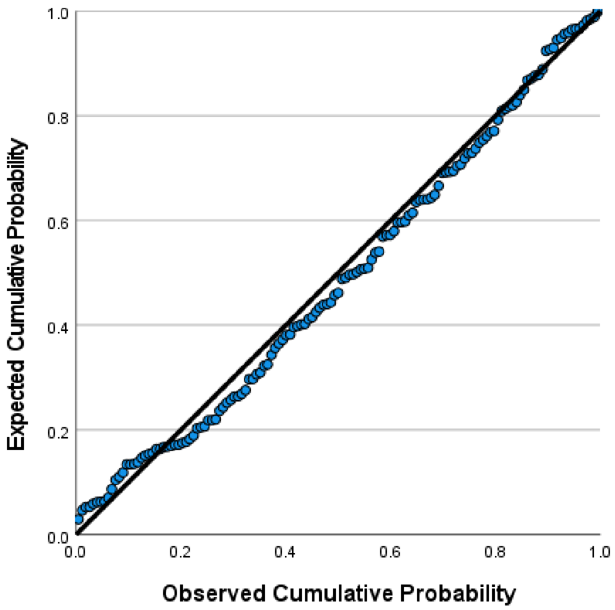


Figure D5

IQ and Cronbach's Differential Accuracy



Appendix E

Linear Regression Model Results: Borman's Differential Accuracy (DV)

Table E1

Linear Regression Model Output: Borman's Differential Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.20	.04	.03	.64	.04	5.92	1	139	.02	1.66

Table E2

ANOVA Results of Linear Model: Borman's Differential Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.41	1	2.41	5.92	.02
	Residual	56.54	139	.41		
	Total	58.95	140			

Table E3

Regression Coefficients of Linear Model: Borman's Differential Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.99	.05		18.34	.00	.89	1.09
	General Intelligence	.03	.01	.20	2.43	.02	.01	.05

Appendix F

Linear Regression Model Results: Cronbach's Accuracy Components (DV)

Table F1

Linear Regression Model Output: Elevation Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.16	.03	.02	.45	.03	3.85	1	139	.05	1.66

Table F2

ANOVA Results of Linear Model: Elevation Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.78	1	.78	3.85	.05
	Residual	28.01	139	.20		
	Total	28.79	140			

Table F3

Regression Coefficients of Linear Model: Elevation Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.64	.04		16.83	.00	.56	.71
	General Intelligence	-.02	.01	-.16	-1.96	.05	-.03	.00

Table F4

Linear Regression Model Output: Differential Elevation Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.12	.01	.01	.28	.01	2.02	1	139	.16	1.9

Table F5

ANOVA Results of Linear Model: Differential Elevation Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.16	1	.16	2.02	.16
	Residual	11.20	139	.08		
	Total	11.36	140			

Table F6

Regression Coefficients of Linear Model: Differential Elevation Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.66	.02		27.69	.00	.62	.71
	General Intelligence	-.01	.01	-.12	-1.42	-.16	-.02	.00

Table F7

Linear Regression Model Output: Stereotype Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.03	.001	-.01	.21	.00	.11	1	139	.74	2.19

Table F8

ANOVA Results of Linear Model: Stereotype Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.005	1	.01	.11	.74
	Residual	6.03	139	.04		
	Total	6.04	140			

Table F9

Regression Coefficients of Linear Model: Stereotype Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.66	.02		27.69	.00	.62	.71
	General Intelligence	-.01	.01	-.12	-1.42	.74	-.02	.00

Table F10

Linear Regression Model Output: Cronbach's Differential Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.12	.01	.01	.22	.01	2.07	1	139	.15	1.87

Table F11

ANOVA Results of Linear Model: Cronbach's Differential Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.10	1	.10	2.07	.15
	Residual	6.88	139	.05		
	Total	6.98	140			

Table F12

Regression Coefficients of Linear Model: Cronbach's Differential Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.46	.02		24.56	.00	.42	.50
	General Intelligence	-.01	.00	-.12	-1.4	.15	-.01	.00

Appendix G

Quadratic Regression Model Results: Borman's Differential Accuracy (DV)

Table G1

Quadratic Regression Model Output: Borman's Differential Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.20	.04	.03	.64	.04	5.92	1	139	.02	
2	.22	.05	.04	.64	.01	1.36	1	138	.25	1.68

Table G2

ANOVA Results of Quadratic Model: Borman's Differential Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.41	1.00	2.41	5.92	.02
	Residual	56.54	139.00	.41		
	Total	58.95	140.00			
2	Regression	2.96	2.00	1.48	3.65	.03
	Residual	55.99	138.00	.41		
	Total	58.95	140.00			

Table G3

Regression Coefficients of Quadratic Model: Borman's Differential Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.99	.05		18.34	.00	.88	1.09
	General Intelligence	.03	.01	.20	2.43	.02	.01	.05
2	(Constant)	1.03	.06		15.96	.00	.90	1.15
	General Intelligence	.04	.01	.24	2.70	.01	.01	.06
	General Intelligence Squared	.00	.00	-.11	-1.16	.25	-.01	.00

Appendix H

Quadratic Regression Model Results: Cronbach's Accuracy Components (DV)

Table H1

Quadratic Regression Model Output: Elevation Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.16	.03	.02	.45	.03	3.85	1	139	.05	
2	.18	.03	.02	.45	.00	.56	1	138	.45	

Table H2

ANOVA Results of Quadratic Model: Elevation Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.78	1	.78	3.85	.05
	Residual	28.01	139	.20		
	Total	28.79	140			
2	Regression	.89	2	.45	2.20	.11
	Residual	27.90	138	.20		
	Total	28.79	140			

Table H3

Regression Coefficients of Quadratic Model: Elevation Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.64	.04		16.83	.00	.56	.71
	General Intelligence	-.02	.01	-.16	-1.96	.05	-.03	.00
2	(Constant)	.65	.05		14.43	.00	.57	.74
	General Intelligence	-.01	.01	-.14	-1.50	.14	-.03	.00
	General Intelligence Squared	.00	.00	-.07	-.75	.45	.00	.00

Table H4

Quadratic Regression Model Output: Differential Elevation Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.12	.01	.01	.28	.01	2.02	1	139	.16	
2	.20	.04	.03	.28	.03	3.74	1	139	.06	

Table H5

ANOVA Results of Quadratic Model: Differential Elevation Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.16	1	.16	2.02	.16
	Residual	11.20	139	.08		
	Total	11.36	140			
2	Regression	.46	2	.23	2.90	.06
	Residual	10.90	138	.08		
	Total	11.36	140			

Table H6

Regression Coefficients of Quadratic Model: Differential Elevation Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.66	.02		27.69	.00	.61	.71
	General Intelligence	-.01	.01	-.12	-1.42	.16	-.02	.00
2	(Constant)	.63	.03		22.25	.00	.58	.69
	General Intelligence	-.01	.01	-.19	-2.08	.04	-.02	.00
	General Intelligence Squared	.00	.00	.18	1.93	.06	.00	.00

Table H7

Quadratic Regression Model Output: Stereotype Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.03	.00	-.01	.21	.00	.11	1	139	.74	
2	.10	.01	-.01	.21	.01	1.16	1	139	.28	

Table H8

ANOVA Results of Quadratic Model: Stereotype Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.00	1	.00	.11	.74
	Residual	6.03	139	.04		
	Total	6.04	140			
2	Regression	.05	2	.03	.63	.53
	Residual	5.98	138	.04		
	Total	6.04	140			

Table H9

Regression Coefficients of Quadratic Model: Stereotype Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.24	.02		13.55	.00	.20	.27
	General Intelligence	.00	.00	-.03	-.33	.74	-.01	.01
2	(Constant)	.25	.02		11.90	.00	.21	.29
	General Intelligence	.00	.00	.01	.12	.90	-.01	.01
	General Intelligence Squared	.00	.00	-.10	-1.08	.28	.00	.00

Table H10

Quadratic Regression Model Output: Cronbach's Differential Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.12	.01	.01	.22	.01	2.06	1	139	.15	
2	.14	.02	.01	.22	.01	.84	1	139	.36	

Table H11

ANOVA Results of Quadratic Model: Cronbach's Differential Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.10	1	.10	2.06	.15
	Residual	6.88	139	.05		
	Total	6.98	140			
2	Regression	.14	2	.07	1.45	.24
	Residual	6.84	138	.05		
	Total	6.98	140			

Table H12

Regression Coefficients of Quadratic Model: Cronbach's Differential Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.46	.02		24.56	.00	.42	.50
	General Intelligence	-.01	.00	-.12	-1.44	.15	-.01	.00
2	(Constant)	.47	.02		20.98	.00	.43	.52
	General Intelligence	.00	.00	-.09	-.96	.34	-.01	.00
	General Intelligence Squared	.00	.00	-.08	-.92	.36	.00	.00

Appendix I

Cubic Regression Model Results: Borman's Differential Accuracy (DV)

Table I1

Cubic Regression Model Output: Borman's Differential Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.20	.04	.03	.64	.04	5.92	1	139	.02	
2	.22	.05	.04	.64	.01	1.36	1	138	.25	
3	.23	.05	.03	.64	.00	.33	1	137	.57	1.67

Table I2

ANOVA Results of Cubic Model: Borman's Differential Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2.41	1.00	2.41	5.92	.02
	Residual	56.54	139.00	.41		
	Total	58.95	140.00			
2	Regression	2.96	2.00	1.48	3.65	.03
	Residual	55.99	138.00	.41		
	Total	58.95	140.00			
3	Regression	3.09	3.00	1.03	2.53	.06
	Residual	55.85	137.00	.41		
	Total	58.95	140.00			

Table I3

Regression Coefficients of Cubic Model: Borman's Differential Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.99	.05		18.34	.00	.88	1.09
	General Intelligence	.03	.01	.20	2.43	.02	.01	.05
2	(Constant)	1.03	.06		15.96	.00	.90	1.15
	General Intelligence	.04	.01	.24	2.70	.01	.01	.06
	General Intelligence Squared	.00	.00	-.11	-1.16	.25	-.01	.00
3	(Constant)	1.02	.07		15.08	.00	.88	1.15
	General Intelligence	.04	.02	.30	2.30	.02	.01	.08
	General Intelligence Squared	.00	.00	-.06	-.45	.66	-.01	.00
	General Intelligence Cubed	.00	.00	-.10	-.57	.57	.00	.00

Appendix J

Cubic Regression Model Results: Cronbach's Accuracy Components (DV)

Table J1

Cubic Regression Model Output: Elevation Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.16	.03	.02	.45	.03	3.85	1	139	.05	
2	.18	.03	.02	.45	.00	.56	1	138	.45	
3	.24	.06	.04	.45	.03	3.78	1	138	.05	1.72

Table J2

ANOVA Results of Cubic Model: Elevation Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.78	1	.78	3.85	.05
	Residual	28.01	139	.20		
	Total	28.79	140			
2	Regression	.89	2	.45	2.20	.11
	Residual	27.90	138	.20		
	Total	28.79	140			
3	Regression	1.64	3.00	.55	2.76	.04
	Residual	27.15	137.00	.20		
	Total	28.79	140.00			

Table J3

Regression Coefficients of Cubic Model: Elevation Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.64	.04		16.83	.00	.56	.71
	General Intelligence	-.02	.01	-.16	-1.96	.05	-.03	.00
2	(Constant)	.65	.05		14.43	.00	.57	.74
	General Intelligence	-.01	.01	-.14	-1.50	.14	-.03	.00
	General Intelligence Squared	.00	.00	-.07	-.75	.45	.00	.00
3	(Constant)	.68	.05		14.51	.00	.59	.77
	General Intelligence	-.03	.01	-.32	-2.45	.02	-.06	-.01
	General Intelligence Squared	.00	.00	-.24	-1.89	.06	-.01	.00
	General Intelligence Cubed	.00	.00	.33	1.94	.05	.00	.00

Table J4

Cubic Regression Model Output: Differential Elevation Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.12	.01	.01	.28	.01	2.02	1	139	.16	
2	.20	.04	.03	.28	.03	3.74	1	139	.06	
3	.20	.04	.02	.28	.00	.01	1	139	.93	1.90

Table J5

ANOVA Results of Cubic Model: Differential Elevation Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.16	1	.16	2.02	.16
	Residual	11.20	139	.08		
	Total	11.36	140			
2	Regression	.46	2	.23	2.90	.06
	Residual	10.90	138	.08		
	Total	11.36	140			
3	Regression	.46	3	.15	1.92	.13
	Residual	10.90	137	.08		
	Total	11.36	140			

Table J6

Regression Coefficients of Cubic Model: Differential Elevation Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.66	.02		27.69	.00	.61	.71
	General Intelligence	-.01	.01	-.12	-1.42	.16	-.02	.00
2	(Constant)	.63	.03		22.25	.00	.58	.69
	General Intelligence	-.01	.01	-.19	-2.08	.04	-.02	.00
	General Intelligence Squared	.00	.00	.18	1.93	.06	.00	.00
3	(Constant)	.63	.03		21.26	.00	.57	.69
	General Intelligence	-.01	.01	-.20	-1.52	.13	-.03	.00
	General Intelligence Squared	.00	.00	.17	1.33	.18	.00	.00
	General Intelligence Cubed	.00	.00	.02	.09	.93	.00	.00

Table J7

Cubic Regression Model Output: Stereotype Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.03	.00	-.01	.21	.00	.11	1	139	.74	
2	.10	.01	-.01	.21	.01	1.16	1	139	.28	
3	.11	.01	-.01	.21	.00	.57	1	139	.45	2.17

Table J8

ANOVA Results of Cubic Model: Stereotype Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.00	1	.00	.11	.74
	Residual	6.03	139	.04		
	Total	6.04	140			
2	Regression	.05	2	.03	.63	.53
	Residual	5.98	138	.04		
	Total	6.04	140			
3	Regression	.08	3	.03	.61	.61
	Residual	5.96	137	.04		
	Total	6.04	140			

Table J9

Regression Coefficients of Cubic Model: Stereotype Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.24	.02		13.55	.00	.20	.27
	General Intelligence	.00	.00	-.03	-.33	.74	-.01	.01
2	(Constant)	.25	.02		11.90	.00	.21	.29
	General Intelligence	.00	.00	.01	.12	.90	-.01	.01
	General Intelligence Squared	.00	.00	-.10	-1.08	.28	.00	.00
3	(Constant)	.25	.02		11.16	.00	.20	.29
	General Intelligence	.00	.01	.08	.62	.53	-.01	.02
	General Intelligence Squared	.00	.00	-.03	-.26	.80	.00	.00
	General Intelligence Cubed	-.00	.00	-.13	-.76	.45	.00	.00

Table J10

Cubic Regression Model Output: Cronbach's Differential Accuracy (DV)

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.12	.01	.01	.22	.01	2.06	1	139	.15	
2	.14	.02	.01	.22	.01	.84	1	139	.36	
3	.14	.02	.00	.22	.00	.04	1	139	.84	1.88

Table J11

ANOVA Results of Cubic Model: Cronbach's Differential Accuracy (DV)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.10	1	.10	2.06	.15
	Residual	6.88	139	.05		
	Total	6.98	140			
2	Regression	.14	2	.07	1.45	.24
	Residual	6.84	138	.05		
	Total	6.98	140			
3	Regression	.15	3	.05	.97	.41
	Residual	6.84	137	.05		
	Total	6.98	140			

Table J12

Regression Coefficients of Cubic Model: Cronbach's Differential Accuracy (DV)

		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	.46	.02		24.56	.00	.42	.50
	General Intelligence	-.01	.00	-.12	-1.44	.15	-.01	.00
2	(Constant)	.47	.02		20.98	.00	.43	.52
	General Intelligence	.00	.00	-.09	-.96	.34	-.01	.00
	General Intelligence Squared	.00	.00	-.08	-.92	.36	.00	.00
3	(Constant)	.47	.02		19.96	.00	.42	.52
	General Intelligence	.00	.01	-.07	-.53	.60	-.02	.01
	General Intelligence Squared	.00	.00	-.07	-.53	.60	.00	.00
	General Intelligence Cubed	.00	.00	-.03	-.20	.84	.00	.00

Appendix K

Statistical Power Analyses of Regression Models

Table K1

*General Input for G*Power 3 Calculation (Linear Multiple Regression: Random model)*

Input	Value
Options:	Exact distribution
Analysis:	Post hoc: Compute achieved power
Input:	
Tail(s)	One
H0 ρ^2	0
α err prob	.05
Total sample size	146
Number of predictors	1

Table K2

*Input and Output Data for G*Power 3 Calculation for Each Model*

Model	Input		Output
	Observed R^2	H_1 p^2	Power (1 – β err prob)
Linear Models			
Borman's Differential Accuracy	.04	.06	.86
Elevation	.03	.05	.80
Differential Elevation	.01	.03	.60
Stereotype Accuracy	.001	.02	.36
Cronbach's Differential Accuracy	.01	.03	.60
Quadratic Models			
Borman's Differential Accuracy	.05	.07	.91
Elevation	.03	.05	.80
Differential Elevation	.04	.06	.86
Stereotype Accuracy	.01	.03	.60
Cronbach's Differential Accuracy	.02	.04	.72
Cubic Models			
Borman's Differential Accuracy	.05	.07	.91
Elevation	.06	.08	.94
Differential Elevation	.04	.06	.86
Stereotype Accuracy	.01	.03	.60
Cronbach's Differential Accuracy	.02	.04	.72