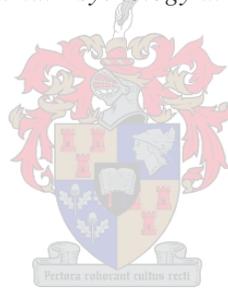


The moderating effect of interview structure on race-group
similarity effects in simulated interview ratings

by
Daniel Benjamin Hauptfleisch

*Thesis presented in partial fulfilment of the requirements for the degree
Master of Commerce in Industrial Psychology at Stellenbosch University*



Supervisor: Mr Francois Servaas de Kock
Faculty of Economic and Management Sciences
Department of Industrial Psychology

March 2012

Declaration

By submitting this thesis electronically, I, D.B. Hauptfleisch, declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2012

ABSTRACT

This study investigated race-group similarity effects as a form of interviewer bias in selection interview ratings. Social Identity Theory predicts that interviewers would assign higher ratings to interviewees of the same social group (the so-called in-group) primarily through the mechanism of similarity attraction. Research findings up to now have lent only partial support to this hypothesis. This study argues that interview structure may help to explain inconsistent research findings since structure could inhibit the functioning of the similarity-attraction mechanism. The present research pursued two objectives, namely (1) to determine the degree to which race-group similarity (between interviewer and interviewee) exerts a biasing effect on selection interview dimension ratings, (2) to determine whether same-group bias increases when interview structure is experimentally diminished. This experimental study manipulated the degree of structure in interviews (high- and low-structured conditions) and compared the degree to which race group similarity effects were evident under each condition. Interviews were simulated by showing video-taped interview segments to a sample of participants and asking them to rate interview dimensions on rating scales that had been compiled to reflect the degree of structure in each condition. The data were analysed using Hierarchical Linear Modelling (HLM) and multiple regression analysis to determine whether similarity effects were present in the interview rating data. The results support the hypothesis that racial similarity effects are found under low-structured conditions, as well as the hypothesis that interview structure moderates the influence of similarity effects. However, racial similarity effects were also found with the highly structured condition. Although these effects were smaller than in the low-structured condition, they were statistically significant. Future research should attempt to replicate this study as a field study to test the generalisability of the findings.

OPSOMMING

Hierdie studie ondersoek onderhoudvoerdersydigheid in die vorm van rasgroepsoortgelykheidseffekte in seleksie-onderhoudbeoordelings. Sosiale Identiteitsteorie voorspel dat onderhoudvoerders diegene van dieselfde rasgroep (die sogenaamde ingroep) met hoër beoordelingstellings sal aanslaan, primêr deur die werking van die soortgelykheid-aangetrokkenheidsmeganisme. Navorsingsresultate tot op hede leen slegs gedeeltelike steun aan hierdie hipotese. Hierdie studie argumenteer dat die rede vir teenstrydige navorsingbevindinge moontlik die gevolg van die bemiddelende effek van onderhoudstruktuur kan wees, aangesien struktuur moontlik die funksionering van die soortgelykheid-aangetrokkenheidsmeganisme kan inperk. Die studie streef dus twee doelwitte na, nl. (1) om die mate waartoe rasgroeppooreenstemming tussen die onderhoudvoerder en onderhoudnemer 'n sydig invloed op onderhouddimensietellings uitoefen te bepaal en (2) om te bepaal of soortgelykheidseffekte toeneem namate onderhoudstruktuur eksperimenteel verlaag word. 'n Eksperimentele ontwerp is gebruik waarbinne onderhoudstruktuur (hoog- en laag gestruktureerde toestande) in video-opnames van onderhoude nageboots is. 'n Groep beoordelaars het hierdie stimilusmateriaal beoordeel aan die hand van beoordelingskriteria wat opgestel is om die mate van struktuur binne elke toestand te reflekteer. Gevolglik is die mate van rasgroepsoortgelykheidseffekte binne elke struktuurtoestand vergelyk. Die navorsingsdata is met gebruik van Hiërargiese Lineêre Modelling (HLM) en veelvoudige regressie ontleed om die teenwoordigheid van soortgelykheidsydigheid te bepaal. Die resultate steun die hipotese dat rassoortgelykheidseffekte onder laaggestruktureerde toestande voorkom, asook dat onderhoudstruktuur 'n modererende rol speel. Nietemin is soortgelyke effekte ook onder die hoog gestruktureerde toestand gevind. Alhoewel hierdie effekte kleiner as onder die laaggestruktureerde toestand was, was dit steeds statisties beduidend. Toekomstige navorsing kan poog om 'n soortgelyke ondersoek as 'n veldstudie te onderneem om die moontlikheid van veralgemening van die resultate te bepaal.

Acknowledgements

First and foremost I want to thank my Lord and Saviour, Jesus Christ, for giving me the hope, courage, favour and wisdom to complete this thesis. You are my strength, my song and my salvation! I also thank

- My Mother and Father, for all the love, text messages, encouragement and prayers
- The most beautiful girl in the world, Ronelle, for all your love, support and prayers
- Francois de Kock, for motivation, direction and wisdom
- My friends Pieter Bouwer and Ockert Augustyn, for friendship and encouragement
- The A-team, Antonette, Christine, Marthine and Nadia, for supporting one another to the end. We made it!
- Ayanda Vabaza and Terzel Rasmus who did wonders in helping me with data collection
- The department of Industrial Psychology for the bursary that supported me in my final year.

TABLE OF CONTENTS

DECLARATION.....	i
ABSTRACT.....	ii
OPSOMMING.....	iii
ACKNOWLEDGEMENTS.....	iv
1. INTRODUCTION.....	1
1.1 Interviewing in Employee Selection	1
1.2 Legal Implications of the Selection Interview	3
1.3 Judgement and Rater Error in the Selection Interview.....	5
1.4 Demographic Similarity as a Cause of Interview Bias.....	6
1.5 Research Problem.....	7
1.6 Value of the Study.....	8
1.7 Research Objectives	9
1.8 Summary	9
1.9 Delimitations	10
1.10 Overview of Thesis Structure	10
2. LITERATURE REVIEW	12
2.1 Introduction.....	12
2.2 The Selection Interview and its Research Niche.....	13
2.3 Interviewer Judgement and Decision Making	14
2.3.1 Interviewer Judgement and Decision Making.....	16
2.3.2 Influences to Interviewer Decision Making.....	18
2.3.3 Summary.	20

2.4 Psychometric Perspectives on Bias	20
2.4.1 Analysis of Bias.....	22
2.4.1.1 Bias in Personnel Selection	23
2.4.1.2 Measuring and Determining Bias	26
2.5 Legal Perspectives on Bias	27
2.6 Psychological Perspectives on Bias	29
2.6.1 Social Identity Theory (SIT).	30
2.6.2 Similarity Attraction Paradigm.	31
2.6.3 Social Cognition Theory.	32
2.6.4 In-group Theory.....	33
2.6.5 Summary.	34
2.7 Demographic Similarity Effects in Interviews	35
2.7.1 Research Findings.....	36
2.7.2 Choice of Race as Similarity Variable.....	40
2.8 Reducing Bias in Interviews	41
2.8.1 Defining Interview Structure.....	42
2.8.2 To Structure or not to Structure.....	43
2.8.3 SIT and Interview Structure.	47
2.9 The South African Context	47
2.10 Research Problem.....	48
2.11 Hypotheses.....	49
2.12 Implications of Expected Results.....	51

CHAPTER 3: RESEARCH METHODOLOGY 52

 3.1 Introduction..... 52

 3.2 Research Design..... 52

 3.2.1 MAXMINCON. 53

 3.3 Conceptual Model 55

 3.4 Stimulus Development 56

 3.5 Measuring Instruments 57

 3.6 Sample 59

 3.7 Procedure 61

 3.7.1 Low-structured Data Collection. 62

 3.7.2 High-structure Data Collection. 63

 3.8 Data Analysis 64

 3.8.1 HLM Analysis. 64

 3.8.2 Regression Analysis..... 69

 3.9 Statistical Hypothesis 70

 3.10 Summary 72

CHAPTER 4: RESULTS 73

 4.1 Introduction..... 73

 4.2 Data Cleaning..... 73

 4.3 Testing for Assumptions..... 75

 4.3.1 Normality, Linearity, Homoscedasticity..... 76

 4.4 Descriptive Statistics and Inter-correlations of Study Variables 79

4.4.1 Descriptive Statistics of the Low-structured Condition.....	80
4.4.2 Descriptive Statistics of the Highly Structured Condition.....	81
4.4.3 Descriptive Statistics on Applicant Level.....	82
4.4.3.1 Low-structured Condition.	83
4.4.3.2 Highly structured Condition.	84
4.4.4 Descriptive Statistics on Item Level.....	89
4.5 Results of HLM and Multiple Regression Analyses.....	90
4.5.1 Low-structured Condition Results.....	90
4.5.1.1 HLM Analysis Results.	90
4.5.1.2 Regression Analysis.....	93
4.5.2 Highly Structured Condition Results.....	95
4.5.2.1 HLM Analysis.	95
4.5.2.1.1 Additional HLM Analyses.....	98
4.5.3 Regression Analysis	100
4.6 Comparative Analysis	103
4.7 Summary.....	105
CHAPTER 5: DISCUSSION OF RESULTS AND RECOMMENDATIONS FOR FUTURE RESEARCH	106
5.1 Introduction.....	106
5.2 Background.....	106
5.3 Summary of Results	108
5.3.1 Low-structured Condition.	108

5.3.2 Highly structured Condition.....	111
5.3.3 Main Hypotheses Testing.....	115
5.4 Discussion of Results	116
5.5 Limitations	121
5.6 Recommendations for Future Research.....	122
5.7 Practical Applications.....	123
5.8 Concluding Remarks	124
REFERENCES	126

ADDENDUM A: Low-structure rating sheet and consent form

ADDENDUM B: High-structure rating sheet and consent form

ADDENDUM C: Ethical clearance form

LIST OF TABLES

Table 3.1: <i>Level-1 descriptive statistics for the low-structured condition</i>	60
Table 3.2: <i>Level-2 descriptive statistics for the low-structured condition</i>	60
Table 3.3: <i>Level-1 descriptive statistics for the high-structured condition</i>	61
Table 3.4: <i>Level-2 descriptive statistics for the high-structured condition</i>	61
Table 4.1: <i>Missing value analysis for low-structured condition</i>	74
Table 4.2: <i>Missing value analysis for the highly structured condition</i>	75
Table 4.3: <i>Means, standard deviations, sample sizes, and interview ratings for each applicant-interviewer race in the low-structured condition</i>	80
Table 4.4: <i>Means, standard deviations, sample sizes, and interview ratings for each applicant-interviewer race in the highly structured condition</i>	82
Table 4.5: <i>Descriptive statistics on applicant level for applicant 1</i>	83
Table 4.6: <i>Descriptive statistics on applicant level for applicant 2</i>	83
Table 4.7: <i>Descriptive statistics on applicant level for applicant 3</i>	83
Table 4.8: <i>Descriptive statistics on applicant level for applicant 4</i>	83
Table 4.9: <i>Descriptive statistics on applicant level for applicant 5</i>	83
Table 4.10: <i>Descriptive statistics on applicant level for applicant 6</i>	83
Table 4.11: <i>Descriptive statistics on applicant level for applicant 7</i>	83
Table 4.12: <i>Descriptive statistics on applicant level for applicant 8</i>	83
Table 4.13: <i>Descriptive statistics on applicant level for applicant 1</i>	85
Table 4.14: <i>Descriptive statistics on applicant level for applicant 2</i>	85
Table 4.15: <i>Descriptive statistics on applicant level for applicant 3</i>	86
Table 4.16: <i>Descriptive statistics on applicant level for applicant 4</i>	86
Table 4.17: <i>Descriptive statistics on applicant level for applicant 5</i>	87

Table 4.18: <i>Descriptive statistics on applicant level for applicant 6</i>	87
Table 4.19: <i>Descriptive statistics on applicant level for applicant 7</i>	88
Table 4.20: <i>Descriptive statistics on applicant level for applicant 8</i>	88
Table 4.21: <i>Descriptive statistics on item level for the high-structure data</i>	89
Table 4.22: <i>Final estimation of fixed effects for results in HLM: Random coefficient regression model</i>	91
Table 4.23: <i>Final estimation of fixed effects for results in HLM: Intercepts as outcomes model</i>	92
Table 4.24: <i>Final estimation of fixed effects for results in HLM: Slopes as outcomes model</i>	92
Table 4.25: <i>Model summary of the multiple regression analysis of the low-structure condition</i>	94
Table 4.26: <i>ANOVA model of the multiple regression analysis of the low-structure condition</i>	94
Table 4.27: <i>Coefficients model of the multiple regression analysis of the low-structure condition</i>	94
Table 4.28: <i>Final estimation of fixed effects for results in HLM: Random coefficient regression model</i>	96
Table 4.29: <i>Final estimation of fixed effects for results in HLM: Intercepts as outcomes model</i>	97
Table 4.30: <i>Final estimation of fixed effects for results in HLM: Slopes as outcomes model</i>	97
Table 4.31: <i>Final estimation of fixed effects for results in HLM (Overall rating – no MVI)</i>	98
Table 4.32: <i>Final estimation of fixed effects for results in HLM (Communication)</i>	99
Table 4.33: <i>Final estimation of fixed effects for results in HLM (People Management)</i>	100
Table 4.34: <i>Coefficients model of the multiple regression analysis of the low-structure condition</i>	102
Table 4.35: <i>ANOVA model of the multiple regression analysis of the highly structured condition</i>	102
Table 4.36: <i>Coefficients model of the multiple regression analysis of the highly structured condition</i>	102
Table 5.1: <i>Sample sizes and summary of key HLM results for racial similarity in the low-structure condition</i>	110
Table 5.2: <i>Sample sizes and summary of key HLM results for racial similarity in the highly structured condition</i>	113

LIST OF FIGURES

Figure 1.1: Prohibition of unfair discrimination in the Employment Equity Act (Republic of South Africa, 1998, p. 27).	4
Figure 1.2: Prohibition in the Employment Equity Act of the use of any personnel selection procedure that “is (...) biased against any employee or group” (Republic of South Africa, 1998, p. 27).	4
<i>Figure 2.1:</i> The context and core processes of the employment interview.	15
<i>Figure 2.2:</i> Interviewer decision-making processes.	18
<i>Figure 2.3:</i> Explanatory diagram of the different levels and different types of bias.	22
<i>Figure 2.4:</i> Prohibiting unfair discriminations in the Employment Equity Act (Republic of South Africa, 1998, p. 27)	29
<i>Figure 2.5:</i> Psychometric testing regulations in the Employment Equity Act (Republic of South Africa, 1998, p. 27)	30
<i>Figure 2.6:</i> The perceived vs actual effectiveness of selection instruments.	44
<i>Figure 3.1:</i> Conceptual model of the interview judgement process influenced by racial similarity and interview structure.	55
<i>Figure 4.1:</i> Histogram depicting the normality of distribution for the ‘Rating’ in the low-structure data set.	77
<i>Figure 4.2:</i> Histogram depicting the normality of distribution for the overall score in the high-structure data set.	78
<i>Figure. 4.3:</i> Histogram depicting the normality of distribution for the communication competency in the high-structure data set.	78
<i>Figure. 4.4:</i> Histogram depicting the normality of distribution for the people management competency in the high-structure data set.	78
	xii

1. INTRODUCTION

1.1 Interviewing in Employee Selection

Employee selection is one of the central functions in human resource management, as well as a primary concern of industrial and organisational psychology research (Guion, 1998). A meta-analysis investigating the relationship between human capital and firm performance by Crook, Todd, Combs, Woehr, and Ketchen (2011) found that human capital related strongly to operational and firm performance, supporting the notion that people are the most valuable asset in any organisation ($r_{\text{operational}} = .25, p < .10; r_{\text{firm}} = .17, p < .10$). Tepstra and Rozell (1993) also recorded a positive relationship between the use of formal and validated selection procedures and organisational profitability. Van Iddekinge, Perres, Verrewe, Perryman, Blass and Heetderks (2009) formally established the relationship between selection and training and unit level performance by using a longitudinal design. Their results show a significant relationship between selection practices and unit performance ($r = .23, p < .05$). Moreover, the effect that accurate selection and good person-job fit has on organisational welfare indicators, such as employee turnover (McCulloch & Turban, 2007), absenteeism (Ones, Viswesvaran, & Schmidt, 2003) and bottom-line profit (Hough & Oswald, 2000) have been shown to be substantial. In this light, a primary focus of human resource management research is to develop valid and reliable selection procedures and to incorporate them into standard human resource management best practices.

Interviews are one of the most popular assessment tools in employee selection (Posthuma, Morgenson, & Campion, 2002). In a recent survey (Wilk & Cappelli, 2003), a large number ($N > 3000$) of employers reported how regularly they use a wide range of selection methods.

On a five-point Likert scale, selection interviews received the highest frequency of usage of any of the selection methods ($x = 4.61$) compared to other popular tools such as *resumes* ($x = 3.52$) and *references from previous employers* ($x = 3.84$). In another large-scale employer survey, only one of the participating companies rejected the use of the interview as a selection tool (Robertson & Makin, 1986). Data from the field suggests that interviews remain the flagship of most organisations' selection programmes.

The reasons for the ubiquitousness of interviews probably reside in the perceptions users have of their usefulness. Bevan and Fryatt (1988) showed that only 2% of personnel managers in their sample survey thought interviews to be a poor predictor of future job performance. These usage patterns have resulted from research evidence that show that interviews can be useful to predict important work outcomes (Conway, Jako & Goodman, 1995).

The predictive validity of interviews depends on interview structure. Structured interviews have been shown to provide greater reliability and better overall ability to accurately predict future performance than unstructured interviews (Campion, Palmer, & Campion, 1997; Campion et al., 1988; Schmidt & Hunter, 1998; Wiesner & Cronshaw., 1988). In their meta-analysis, Conway et al. (1995) found that the upper limits of validity of structured interviews were estimated at .67, as opposed to only .34 for unstructured interviews.

In practice, the use of unstructured interviews is favoured over its structured counterpart. From the previously mentioned research and from other research on interviews (e.g., Macan, 2009; McCarthy, Iddekinge, & Campion, 2010; Postuma et al., 2002) it logically follows that the use of structured interviews should be common practice due their clear benefit, i.e.,

predictive validity. This assumption, however, is not supported by field surveys of interview usage patterns. In a comprehensive study by Ryan, McFarland, Barron, and Page (1999), which included 959 companies from 20 different nations, it was found that only 34,7% of the companies used structured interviews instead of unstructured interviews. In the same study, it was found that 50% of companies in South Africa used structured interviews. Though the SA figure compares relatively well with the international trend toward using unstructured interviews, it is alarming to note that unstructured interviews are still so widely practised – given all the advocating research that has been done on the benefits of using the structured interview, as well as the possible dangers of using unstructured interviews, e.g. interviewer subjectivity and bias (Guion, 1998).

1.2 Legal Implications of the Selection Interview

The Employment Equity Act No. 55 of 1998 provides clear stipulations regarding the use of 1) psychometric assessment tools and 2) selection and discrimination practices. The selection interview is used in selection as a psychometric assessment tool (Guion, 1998) and should therefore be aligned to adhere to the laws that govern the above-mentioned areas.

Figure 1.1.

Prohibition of unfair discrimination in the Employment Equity Act (Republic of South Africa, 1998, p. 27)

CHAPTER 2

Prohibition of unfair discrimination

5. Elimination of unfair discrimination

Every employer must take steps to promote equal opportunity in the workplace by eliminating unfair discrimination in any employment policy or practice.

6. Prohibition of unfair discrimination

- (1) No person may unfairly discriminate, directly or indirectly, against an employee, in any employment policy or practice, on one or more grounds, including race, gender, sex, pregnancy, marital status, family responsibility, ethnic or social origin, colour, sexual orientation, age, disability, religion, HIV status, conscience, belief, political opinion, culture, language and birth.

Since interviews are so widely used and trusted in personnel selection, establishing their reliability and validity remains paramount. More than the financial implications of accurate and effective selection decision making, there are also legal implications. The Employment Equity Act (No. 55 of 1998) established clear guidelines for the use of personnel selection procedures in Chapter 2 (Prohibition of Unfair Discrimination) (see Figure 1.1), prohibiting unfair discrimination on any grounds that are not job-related, including demographic group membership. Personnel selection procedures, more specifically, should therefore be shown to be free from bias that systematically disadvantages any subgroup members that does not carry a 'protected' status (see Figure 1.2).

Figure 1.2.

Prohibition in the Employment Equity Act of the use of any personnel selection procedure that "is ... biased against any employee or group" (Republic of South Africa, 1998, p. 27).

8. Psychometric testing

Psychometric testing and other similar assessments of an employee are prohibited unless the test or assessment being used-

- (a) has been scientifically shown to be valid and reliable;
- (b) can be applied fairly to employees; and
- (c) is not biased against any employee or group.

From a legal perspective it would be to the benefit of users of selection interviews to scrutinise their interview procedures for the presence of these prohibitions. Using validated, reliable, and fair selection interviews that are free from bias should be legally defensible and less prone to litigation.

1.3 Judgement and Rater Error in the Selection Interview

Interviewers play a major role in interviews, since they interact with interviewees and produce ratings as a result of this interaction (Macan, 2009). The interview process makes use of people as judges, called interviewers, and not mechanical answering and scoring sheets. This complication in assessment forces the imperative of understanding the subjectivity of human judgment as well as finding the best way to manage its flaws in order to avoid the prohibited practices outlined in the previous section.

The judgement process that interviewers follow to assign scores to applicants plays a pivotal role in employment interviews, since the validity of the employment interview depends largely on the degree to which rating error can be removed from the judgement process before, during and after the interview (Macan, 2009). Rating error or bias can be defined as any construct-irrelevant source of variance in ratings (Schmitt, Pulakos, Nason, & Whitney, 1996). Rater error specifically exists when actual or perceived differences between applicants cause variance unrelated to the measured constructs in judgements and subsequent ratings (Schmitt et al., 1996).

Rater error (bias) has been found to account for substantial portions of variability in scores. Hoffman, Lance, Bynam, and Gentry (2010), for instance, report that idiosyncratic rater effects accounted for an average of 55% of the variance in multisource job performance ratings in their study, similar to earlier studies (e.g., Scullen, Mount & Goff 2000: 58%; Mount, Judge, Scullen, Sytsma & Hezlett (1998: 71%). Similar effects have been found in other judgement contexts. For instance, raters accounted for between 20% (Kenny, 1991) and 37% (Hoyt & Kerns, 1999) of the variance in ratings in social perception tasks and observer

ratings, respectively. This evidence points to the possibility that rater judgements are systematically influenced by diverse, potentially irrelevant factors.

1.4 Demographic Similarity as a Cause of Interview Bias

The reasons for these persistent rater source effects have been heavily researched. For instance, there is considerable evidence that demographic similarity between raters and those who are rated can influence various work outcomes (Riordan, 2000) and the influence that demographic similarity has on interviewer judgement and ratings has been extensively researched (Buckley, Jackson, Bolino, Veres, & Feild, 2007; Goldberg, 2005; Graves & Powell, 1996; Harris, 1989; Lin, Dobbins, & Fahr, 1992; McCarthy, Van Iddekinge, & Campion, 2010; Prewett-Livingston, Veres III, field, & Lewis, 1996; Sacco, Scheu, Ryan, & Schmitt, 2003; Schmitt, 1976) and continues to attract further research attention. The majority of these research investigations, as is the case with this proposed study, draw on Social Identity Theory (SIT) as theoretical basis. SIT has been prompted to be the reason why individuals show a preference for similar others (Goldberg, 2005). SIT is relevant to this field of study due to its derived assumption that raters will more favourably perceive and rate those similar to themselves.

Although strongly supported by its foundation in theory, the aggregate of research findings on demographic similarity effects in interviews surprisingly tend to be inconclusive (Huffcutt, 2011). In addressing the inconclusive nature of the findings, Posthuma et al. (2002, p. 5) state that "...future research should articulate the underlying psychological mechanisms through which similarity may influence interviewer judgments". In other words: why interviewer-interviewee similarity could affect interviewer judgment is not yet fully

understood. Much of the prior research investigating similarity effects has, apart from the overarching SIT paradigm, not yet explicated how similarity effects develop and are influenced by external factors.

It could be argued that the functioning of the similarity attraction mechanism (within the SIT paradigm) is constrained by interview design factors such as interview format (Lin, Dobbins, & Fahr, 1992) and/or interview design and rating scale format. The prevalence of racial similarity effects in interviews of differing structure have, however, only been investigated indirectly. Sacco et al. (2003) more directly suggested that future research might evaluate the hypothesis that similarity effects would be found under conditions when the interview is unstructured. Not having found evidence of similarity effects in the structured interviews used in their research study, Sacco et al. suggested that "...[they] strongly suspect that less-structured recruiting interviews would be more susceptible to demographic similarity effects" (p. 860). This suspicion has not been proven to date and leaves the HR practitioner and Industrial Psychologist with reason for debate, food for thought and opportunities for research.

1.5 Research Problem

The research problem of this study is: To what extent does interview structure moderate the prevalence of racial similarity effects in selection interview ratings within a South African context?

1.6 Value of the Study

The extent to which demographic variables play a role in selection decisions can have important consequences for those being evaluated (and for organisations that make use of these ratings) with respect to fairness, diversity, and legal defensibility (McCarthy, Van Iddekinge, & Campion, 2010). Moreover, systematic sources of variance such as demographic similarity effects in interviews have important consequences for construct- and criterion-related validity (McCarthy et al., 2010).

If systematic, irrelevant factors do account for variance in the judgement process, as suggested by the research evidence cited above, it would have imminent implications that provide utility to the study of this phenomenon. In South Africa, the Employment Equity Act (No. 55 of 1998) clearly prohibits psychometric or similar assessments, such as interviews, unless the assessment tool: 1) has been proven valid and reliable; 2) can be applied fairly to any employee and 3) is not biased against any group or individual. When construct-irrelevant sources of variance are present in interview ratings, employers using these ratings are open to litigation. Furthermore, identifying and removing sources of interviewer bias should increase the probability that employers place the right individual in the right job, preventing job-person misfit and high employee turnover. In other words, employers stand to gain from determining whether interviewer-interviewee similarity acts as a bias in interview ratings, especially those employers that make use of unstructured interviews.

The individual and society at large also stand to gain. Employment practices should promote the wellbeing of society at large and, as such, should not adversely affect the employment outcomes of certain subgroups of society for reasons that are not job-related or, stated otherwise, inherent requirements of the job, or for reasons that do not promote the sound use

of affirmative action measures. Individual applicants stand to gain from interview procedures that are free from interviewer bias, since the probability of being appointed to a position would be a direct function of their probability of success in the position (cf. Guion & Highhouse, 2006).

1.7 Research Objectives

In an attempt to address the previously mentioned research needs and the purpose of this study, the proposed research will pursue the following research objectives:

- 1) To establish the extent to which race-group similarity between interviewers and interviewees influence interviewer ratings in employment interviews.
- 2) In doing so, to investigate the generalisability of similarity research conducted elsewhere to the South African context.
- 3) To determine whether interview design can influence the prevalence of possible race-group similarity effects in interview ratings.
- 4) To determine whether there might be other variables influencing the prevalence of similarity effects in interview ratings.
- 5) To make recommendations for future research, as well as to highlight practical applications that might sprout from the research results.

1.8 Summary

With the high price placed on employment equity and affirmative development initiatives and policies, the selection of personnel in South Africa has increased in complexity and therefore

needs clear and mechanical principles that practitioners can use with confidence. The objective of the employment interview is to assist in making a fair and unbiased judgement with regard to predicting future job performance. There is ample evidence suggesting that demographic similarity effects may be a cause of systematic bias in interview ratings in the literature. The literature also suggests that the influence of interview structure also plays a part in eradicating or catalysing such bias. This study lends itself to future research in order to further determine and specify the most effective interview design for the general South African context.

1.9 Delimitations

Although the need to investigate perceived similarity effects in the same way that the current study aimed to investigate actual similarity effects was recognised, it falls beyond the scope of this study to investigate the influence of perceived similarity effects in the employment interview. The current study also recognised that there are numerous interview types that may yield different results in accordance with the objective of this study, but, this study being the first of its kind in South Africa, it was decided to focus on only two types of interviews (Campion, Pursell, & Brown, 1988). In the interest of future research, demographic factors other than race group, like age and sex, can be used to formulate hypotheses similar to those formulated here. Moderators other than interview design can be also tested for their ability to influence the prevalence of similarity bias.

1.10 Overview of Thesis Structure

A thorough literature review on the elements influencing and those central to the employment interview judgement and decision making process was undertaken for this research project.

Hypotheses central to the determination of the research objectives were formulated from an understanding gained from the literature. The statement of the hypotheses is followed by a detailed explanation of the method whereby these hypotheses were tested and statistical hypotheses were formulated. The next chapter reports the results obtained from the analyses conducted on the captured data and reports on the acceptance or rejection of the stated hypotheses. The final chapter presents a discussion of the results in the context of the literature reviewed, previous studies conducted and the expectations of this study. It concludes by highlighting recommendations for future research, practical applications and a discussion of the limitations of the study.

The next chapter will provide an overview and discussion of relevant research and literature that underlies the research problem and objectives of this study.

2. LITERATURE REVIEW

2.1 Introduction

The selection interview is an effective, but complex and controversial, tool for personnel selection (Wiesner & Cronshaw, 1988). It is unique among other assessment techniques like paper-and-pen and situational judgement tests due to the fact that the interview judgement process, at its core, comprises the interaction between two or more people. Even though the employment interview has proven to be a good predictor of future job performance, biased ratings still torment the practice (Macan, 2009). The challenge to the researcher and the practitioner is to pin-point areas where biases occur and to develop methods to reduce the probability of these biases influencing the judgement and subsequent rating processes.

This literature review comprises: (1) a discussion of the selection interview and its research niche; (2) discussion of the interview judgement process and a brief explanation of its core processes; (3) discussion of bias in interviews, focusing on interviewer bias; (4) an investigation of the legal considerations with regard to the use of the selection interview; (5) presentation of the theoretical base and mechanisms that underlie the concept of similarity effects; (6) presentation and discussion of results from research on demographic similarity effects in interviews; (7) a debate on the use of structure as a moderating variable for possible racial bias in interviews, while also referring to relevant research results; and (8) concludes with the formulation of hypotheses and implications of expected results.

2.2 The Selection Interview and its Research Niche

Employee selection is one of the central functions in human resource management, as well as a primary concern of industrial and organisational psychology research (Guion, 1998). In a meta-analysis investigating the relationship between human capital and firm performance, Crook et al. (2011) found human capital to relate very strongly to firm performance, supporting the notion that people are the most valuable asset in any organisation. Moreover, the effect that accurate selection and good person-job fit has on organisational welfare indicators such as employee turnover (McCulloch & Turban, 2007), absenteeism (Ones et al., 2003) and bottom-line profit (Hough & Oswald, 2000) have been shown to be substantial. In this light, a primary focus of human resource management research is to develop valid and reliable selection procedures and to incorporate them into standard human resource management best practices.

Interviews are one of the most popular assessment tools in employee selection (Posthuma et al., 2002). A recent survey (Wilk & Cappelli, 2003) recorded a large number ($N > 3000$) of employers reporting how regularly they use a wide range of selection methods. On a five-point Likert scale, selection interviews received the highest frequency of usage of any of the selection methods, with a mean score of 4.61, compared to other popular tools such as *resumes* (3.52) and *references from previous employers* (3.84). In another large-scale employer survey, only one of the participating companies rejected the use of the interview as a selection tool (Robertson & Makin, 1986). Data from the field seem to suggest that interviews remain the flagship of selection programmes in most organisations.

The selection interview is a rich source of interactional behaviour between applicants and interviewers (Guion, 1998). While such a setting provides much of the interest for behavioural science, the concern and focus of the industrial psychologist primarily involves organisational welfare and success by means of scientifically anchored human capital management (Theron, 2010b). Therefore, the most important behaviours in the selection interview, from the vantage point of the industrial psychologist, would be interviewer judgement and the subsequent decision making that interviewers engage in. Judgement and decision-making behaviours eventually impact final selection decisions, talent concentration, and employee turnover figures (McCulloch & Turban, 2007) and should therefore be studied thoroughly to be understood well.

2.3 Interviewer Judgement and Decision Making

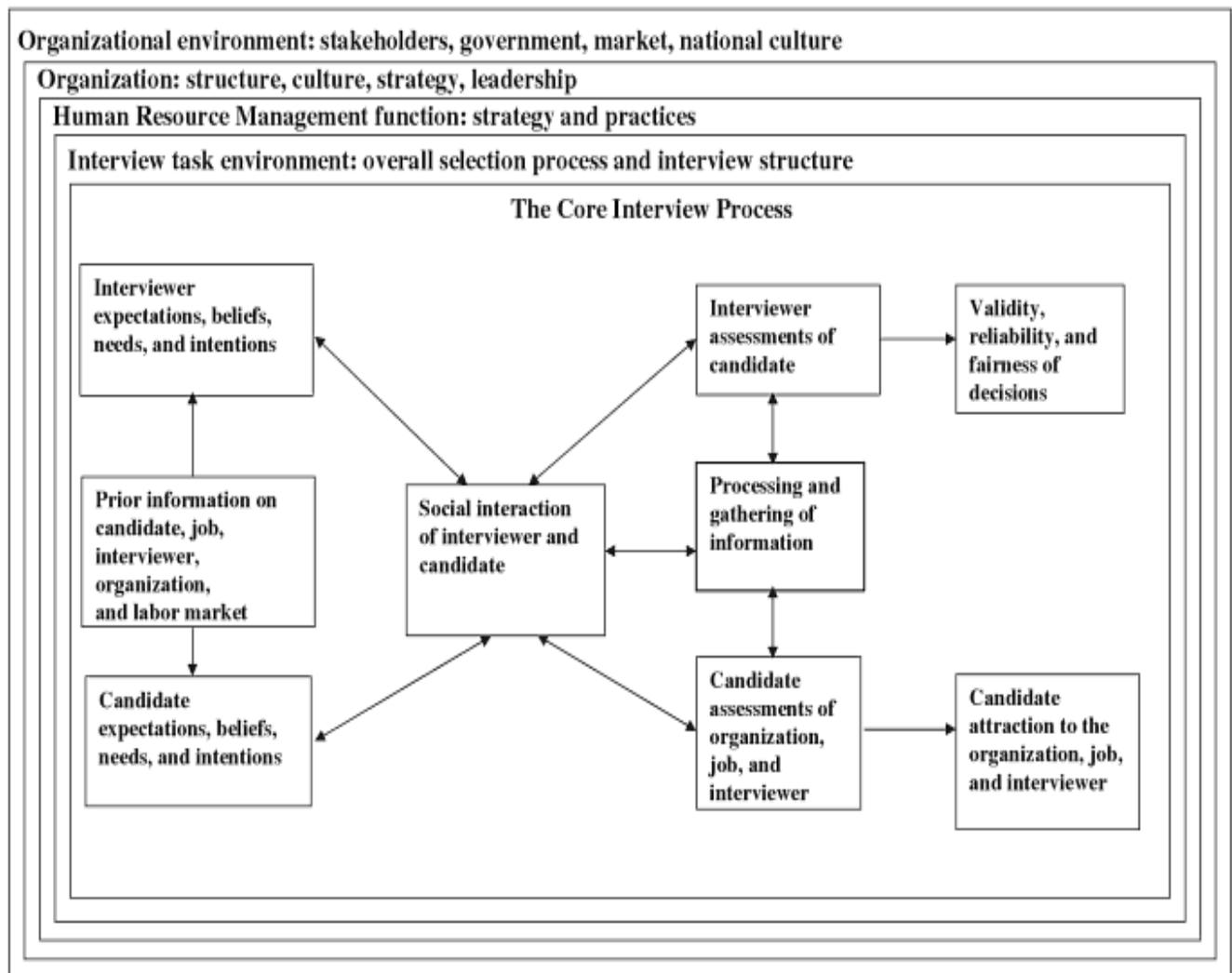
“Judgements are made during interviews, whether formally recorded as ratings or not, and judgements include assessments, predictions and decisions” (Guion, 1998). The interviewer, the applicant and the larger organisational and social environment are integral parts of the judgement and decision-making process in the employment interview. Though not all of these factors are actively involved in the actual interview process, their influence on the applicant and interviewer is significant (Dipboye, 2005).

For understanding the core processes of the employment interview, the following model is of great help in explaining the context and environment that surrounds any employment interview (Dipboye, 2005). Dipboye points out that the intentions, expectations, needs and beliefs of both the interviewer and the applicant are taken into consideration, together with the interaction between these two sets of variables and the

decisions that are made by both parties as a result of the interview. This model, as a starting point, provides a framework for interview investigation since it provides a comprehensive overview of all the relevant factors that could and do influence the selection interview, process – from the need therefore to the outcome thereof.

Figure 2.1.

The context and core processes of the employment interview (Dipboye, 2005)



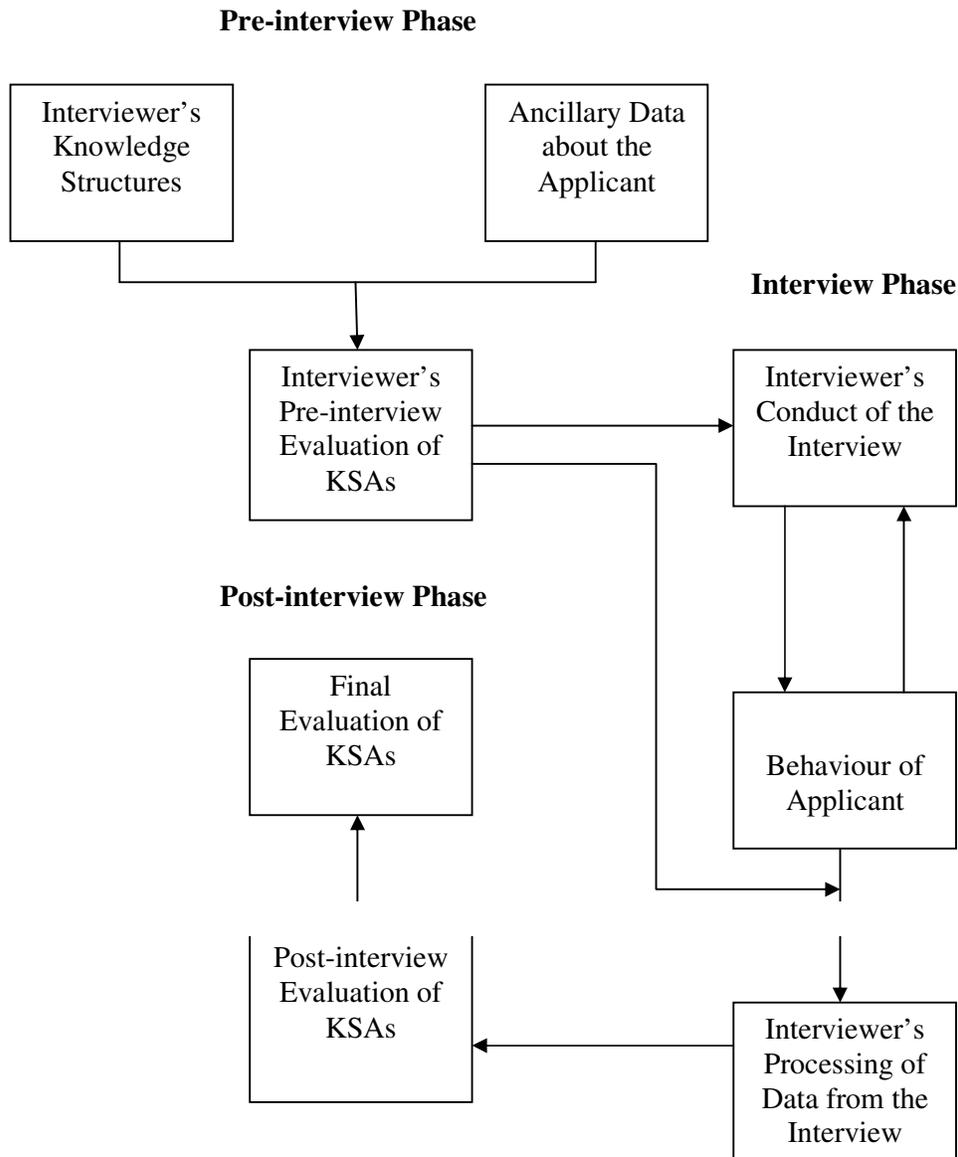
2.3.1 Interviewer Judgement and Decision Making

In moving from this broad overview of the core interview processes to a more direct investigation of the decision-making process applicable to this study, a related framework on interviewer information processing, judgement and decision making by Fiske and Neuberg (1990) follows. This framework models the cognitive and sociological processes that the interviewer typically engages in whilst forming judgements and making decisions. First, the interviewer would, almost subconsciously, *categorise* the applicant. This categorisation can take place in accordance with social cognition theory (2.6.3), as a result of the interviewer's cognitive schemes, stereotypes and prototypes formed from previous experiences (Kulik & Bainbridge, 2006). Second, the interviewer will *characterise* the applicant within the categorised framework. This is done in reaction to the responses the individual gives to questions asked and on the trait-levels that the interviewer derives from the responses. The characterisation phase is limited to the boundaries that the previously chosen category cognitively (and largely subconsciously) imposes on the interviewer. With a reasonably fixed perception of the individual, the interviewer will then, in the last part of the process, alter the formed schema about the person by means of *correction* of previously held ideas that are proven false by new information from and reactions on the part of the applicant (Fiske & Neuberg, 1990).

In another model, further 'zooming in' on interviewer behaviour, by Dipboye (2002) (Figure 2.2) the emphasis is placed more directly on the decision-making process that the interviewer typically engages in. Central to the model is the construct of *knowledge structures*. Knowledge structures refer to the aggregate of a person's previous education, training and

experience (Gatewood & Feild, 2001) that would impact on the three phases of the interview. The model describes '*pre-interview*', '*interview*' and '*post-interview*' phases of the selection interview. The '*pre-interview*' phase comprises a precipitate evaluation by the interviewer – judging the ancillary data of the applicant within the framework of the interviewers' current knowledge structures. The '*interview*' phase is concerned with the interaction between the interviewer's conduct and the response of the applicant and resolves as the interviewer's processing of the interview information. During the '*post-interview*' phase the interviewer will evaluate the knowledge, skills and abilities (KSAs) of the applicant, from information gathered before and during the interview, and will conclude with a final evaluation of the applicant's KSAs (Gatewood & Feild, 2001).

Figure 2.2.
Interviewer decision-making processes



2.3.2 Influences to Interviewer Decision Making

As mentioned earlier, the broader social, cognitive and environmental context of the interview, interviewer and applicant influences the judgement of the interviewer and the subsequent decisions encouraged by the prompts of the earlier judgement. This section provides insight to a number of such influences.

In an experimental study investigating variables that influence interviewer decision making by Webster, as cited in Guion (1998), the following conclusions, among others, were reached (a brief, intriguing remark follows each conclusion):

Interviewers with the same background develop stereotypes of a ‘good candidate’ based on their own background and subsequently try and match applicants to their favoured stereotype (Webster, 1964).

It can logically be concluded that the ideal match with the stereotype will probably share the same social background as that which led the interviewer to develop the favoured stereotype.

Most of the interviewer judgement and assessment decisions are formed within the first four minutes of the interview and final decisions tend to be consistent with it.

If this is the case, then, according to the Fiske and Neuberg (1990) theory, there would not be sufficient time or energy for the last of the three decision-making processes, namely, the *correction* of previously held ideas that are proven false by new information. It can be argued that the bulk of the decision making is made within the *categorisation* and *characterisation* stages – which are grounded in interviewer stereotypes.

Research has also shown other factors that might influence interviewer judgements towards non-criterion orientated decisions. One such variable, namely interviewer experience, might seem to be an asset to an interview panel, but Gehrlein, Dipboye, and Shahani (1993) argue that experience breeds confidence, even if it is unwarranted. In their study, higher validity coefficients were found among inexperienced interviewers than the experienced.

Nonverbal cues are also known to influence interviewer judgements. Behaviours like leaning back after answering, giving a firm handshake or sitting with arms folded are seen by many

interviewers as indicators of character and predictions of future behaviour (Guion, 1998). Although generally accepted by many interviewers, there is no empirical evidence to confirm that any non-verbal cues can be used to predict the character trait they are attributed to, less even the job-relevant criterion (Guion, 1998).

2.3.3 Summary

The interviewer decision-making process, as a whole, seems to be a very fallible source for obtaining valid and reliable information, with many factors that might provoke biased judgement and decision making (Guion, 1998). The following sections dig deeper into the concept of bias in order to gain some insight as to how biased decision making might be understood and subsequently limited in or removed from the employment interview.

2.4 Psychometric Perspectives on Bias

In any interview judgement context there is an observed score (rating) (X), a true score (T) and an error score (e) that can be written in the equation: $X = T + e$ (Gatewood & Feild, 1995). This is known as the true score model of classic reliability theory (Hoyt, 2000). It would be ideal to only use the true score for decision making, but with innumerable variables present in the human condition it is impossible not to have error variance as a part of the observed score, which is inevitably the score by which to discriminate (Guion, 1998). When an assessment tool yields an observed score of which a small percentage is contributed by an error score, the effect thereof can be ignored, but too often the error score proves very directive in final judgements and decisions (Guion, 1998).

Research into selection interviews should focus the spotlight on sources of error variance and provide ways of removing them to the greatest extent possible in order that observed scores may provide a less polluted reflection of the true score – a proposed measure of the criterion. Guion indicates that error variance can be dissected into 1) systematic error variance (e_s) and 2) random error variance (e_r). Systematic errors are errors that, within a specific judgment context, produce inaccurate ratings repeatedly and predictively. Random errors, on the other hand, are errors that seem to vary randomly across repeated measures within the same context.

Bias would be one factor that produces systematic measurement errors. Murphy and Davidshofer (2005) explain that bias in measurement is any systematic error in judging a specific characteristic or attribute. Any irrelevant factors that cause judgement to sway to either a positive or negatively bias side would account for error variance and imply immediate unfair discrimination (Foxcroft & Roodt, 2005). An analysis of bias would therefore be fruitful in the process of removing error variance from the judgement process.

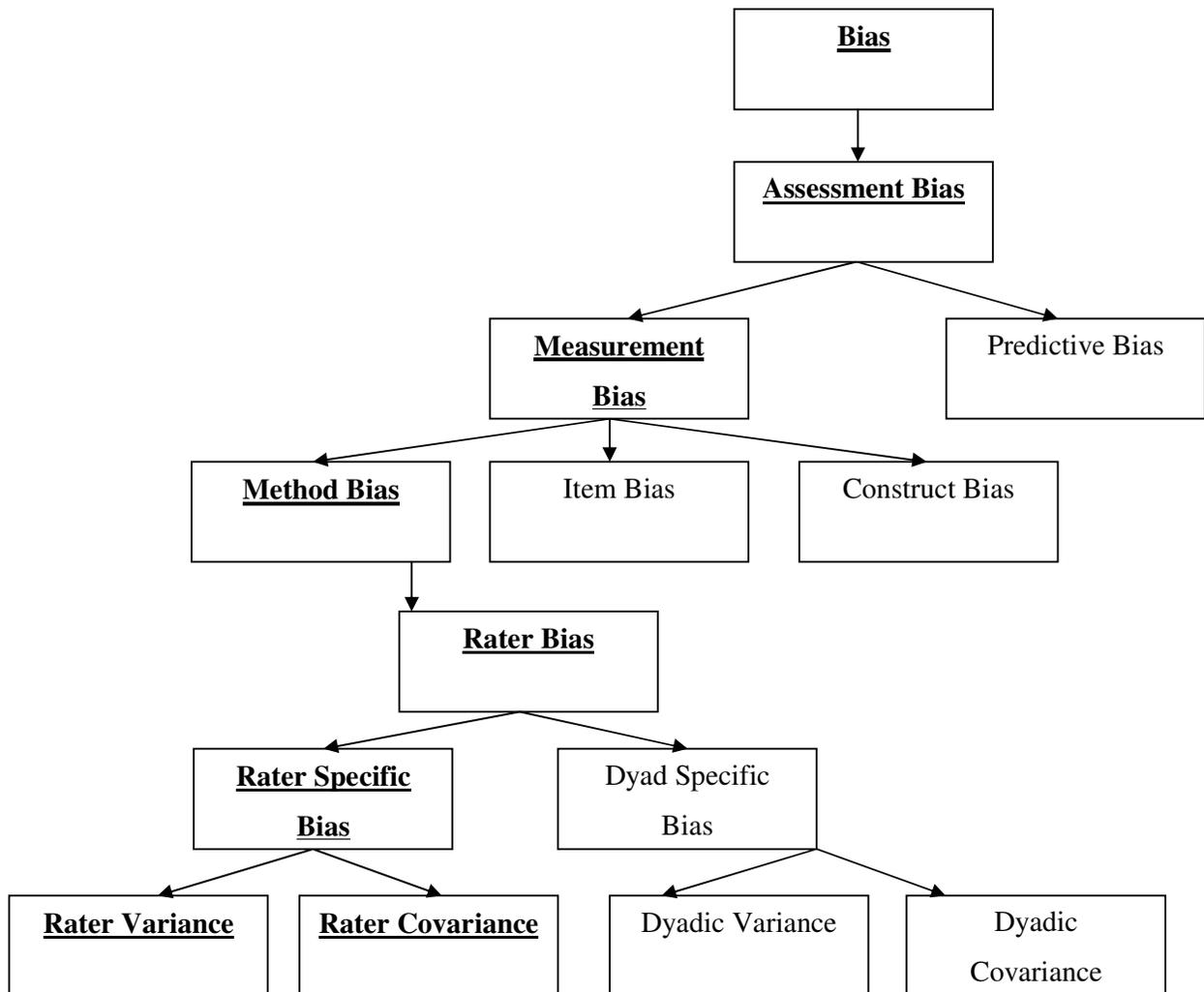
Bias is defined by Guion (1998, p. 433) as:

...systematic group differences in item responses, test scores, or other assessments for reasons unrelated to the trait being assessed – a form of the more general third variable problem in which one or more sources of unwanted systematic variance function differently in the groups compared.

2.4.1 Analysis of Bias

Figure 2.3 provides insight into the different types of bias with specific reference to the relevant type of bias (in bold) that this study investigates. The model was composed from various sources that are authoritative in the area of bias (Gatewood & Feild, 1995; Guion, 1998; Murphy & Davidshofer, 2005; Theron, 2010). Bias is known to have an impact in many areas of modern life not at all relevant to this study. The paths and details of the model, as well as its relevance to this study, will be discussed in more detail in the next section.

Figure 2.3.
Explanatory diagram of the different levels and types of bias



2.4.1.1 Bias in Personnel Selection

Bias in the personnel management and, more specifically, the employee selection field, is referred to as assessment bias (Hoyt, 2000). Assessment bias can be divided into two major types – predictive bias and measurement bias (Murphy & Davidshofer, 2005). Predictive bias exists when consistent non-zero prediction errors are found to be made for members of a specific subgroup (SIOP, 2003). For the purpose of this study we will look more closely at measurement bias. Murphy and Davidshofer (2005, p. 318) define measurement bias when “...the test makes systematic errors in measuring a specific characteristic or attitude.”

Theron (2010, p. 123) explains it as:

Measurement bias refers to all systematic factors that could account for variance in observed test scores that cannot be accounted for in terms of the latent variable of interest.... Other systematic but non-relevant factors and non-systematic, random factors [also] play a role in determining the response to the test stimulus set. These systematic nuisance factors essentially refer to any systematic source of unique variance in the test scores that cannot be explained in terms of variance in the latent variable of interest.

Essentially there are three types of measurement bias:

- Item bias
- Construct bias
- Method bias

Item bias is also known as differential item functioning (DIF). This form of measurement bias focuses on bias at the item level and is present when group membership can explain the variance in an observed item response that cannot be explained by the latent variable. In other words, if the probability of observing a score on a specific item will be different for individuals from different groups, even though they have the same standing on the latent variable that is being measured (Theron, 2010a). Item bias will therefore cause the regression of the observed scores to differ across groups in terms of either intercept or slope. When item bias affects the slope it is referred to as non-uniform measurement bias and when it affects the intercept it is referred to uniform measurement bias (Theron, 2010a). To summarise, similar scores of individuals from different groups can only be seen to reflect an equal standing on the latent variable if the corresponding regression models are equivalent in terms of slope and intercept.

Construct bias can be defined as bias that occurs when observed scores do not reflect the same construct across different groups (Theron, 2010).

Construct bias exists if the construct that is measured by the test in different groups differ in terms of; 1) the number of factors it comprises, 2) how these factors are related, 3) the pattern with which the items load on the factors and 4) how the construct is embedded in a larger nomological network (Theron, 2010, p. 127).

Method bias focuses on group-related factors that cause members from different groups to respond in different ways to various test stimuli. Method bias, unlike item and construct bias, does not describe a specific facet of the latent variable test-testee response relationship, but rather serves as a way to better explore and explain item and construct bias. According to

Theron (2010) there are four major sources that can cause method bias: a) social desirability of individual responses; b) item familiarity of different groups; c) different item response styles; and, lastly, d) various group differences that can affect individuals' responses to test stimuli.

For the purposes of this study, the focus is more specifically on method bias as is clear from the definitive discussion above. The more specific type of bias relevant to this study within the scope of method bias is referred to as rater bias. Hoyt (2000) explains that "Rater bias refers to disagreements among raters due to either (a) their differential interpretations of the rating scale or (b) their unique (and divergent) perceptions of individual targets". He refers to the two types of rater bias identified in the definition of rater bias as (a) rater specific bias and (b) dyad specific bias, explaining that dyadic variance refers to the extent to which ratings by raters will vary on the grounds of unique, non-relevant perceptions about certain applicants while dyadic covariance reflects the way in which these dyadic effects seen on one item of assessment will also be seen in another. In other words, it determines the extent to which a rater will be more lenient towards a particular candidate, or candidates, across items. In literature this is also known as a leniency effect. Rater specific bias, or rater effects, concerns the degree to which raters differ in their generalised perceptions of targets. This study focused on the latter of these two, of which there is again two types – rater variance and rater covariance (Hoyt, 2000).

Rater variance and covariance jointly refer to the effect that research often refers to as interaction effects. An interaction effect would be one where two independent variables in joint existence or operation, create an effect that the variables on their own do not create. An example would be racial interaction in the context of an interview, where the interaction

between the race of the rater and the race of the applicant proves to have a significant influencing effect on the rating that the rater gives the applicant.

2.4.1.2 Measuring and Determining Bias

From the above definitions of bias, it can be assumed that a test is free from measurement bias or is measurement invariant if different groups have the same probability of scoring any random score for a specific test. In the context of this study, the test concerned whether different race groups rate the similar-to-them and different-to-them applicant groups systematically as different or consistently as the same. The prior rating would indicate a racial similarity effect as a form of rater bias.

Measurement bias is a potential concern for both predictor- and criteria-related validity and should be tested for to know its impact. Testing for measurement bias necessitates the comparison of observed scores and true scores (Society for Industrial and Organizational Psychology 2003). The method of testing for or determining measurement bias involves examining the external correlations of performance for different individuals on a test or assessment (Society for Industrial and Organizational Psychology, 2003; Theron, 2010). One can also utilise interval evidence to determine whether an assessment measures different constructs differently in different groups (Murphy & Davidshofer, 2005).

2.5 Legal Perspectives on Bias

The aim of the selection interview is to provide information that facilitates fair discrimination between applicants (Macan, 2009). The Employment Equity Act (No. 55 of 1998) Section 20(3) lists the legal grounds for discrimination in order to make a fair selection decision, while also condemning any unfair selection processes that do not comply with these criteria for fair discrimination:

- a) Formal qualifications;
- b) Prior learning;
- c) Relevant experience;
- d) Capacity to acquire, within a reasonable time, the ability to do the job. (Republic of South Africa, 1998)

Discrimination on any other grounds would count as unfair discrimination. This should be read with Section 6(2) b which explains that it is also not unfair to exclude any individual on the basis of an *inherent job requirement*. If there is proof that the criteria used to discriminate is indeed directly related to an inherent job requirement, discrimination on such grounds would be deemed fair. Section 6(2) therefore provides a void in which the employer can create criteria for discrimination that specifically fits the organisations' needs. For example, a beauty salon may argue that having slim, beautiful lady working at reception, will boost the company image and is therefore an inherent requirement to the job. Would this be fair, however? This leads to further questions: (1) what is fairness and (2) does the law contradict itself in creating room for 'unfair' selection practices?

To attain legal compliance, it is important to gain insight into a definition of fairness, since fairness is what the law requires of selection practices. The problem with a definition of fairness is that fairness is a value-laden concept that might not have the same meaning for everyone – depending on ethical and social factors (Theron, 2007). The most widely accepted definition of fairness, though, is the Cleary definition. Cleary (1968) explained that fairness implies equal regression lines for different groups. Unpacked further, it would imply that any measure that systematically over or under predicts a certain group's performance, would be an unfair measure.

From these definitions it is clear that measurement bias would probably carry through to unfair discrimination. It is further interesting to note that, even though the predictor (test, assessment, and interview) can be declared free from measurement bias, as defined earlier, it does not necessarily indemnify inferences made from the predictor data from being unfair (Theron, 2007). The purpose of this study, however, was focused on removing measurement bias from the employment interview – a possible source of unfair discrimination – and did not necessitate an in-depth discussion of predictive bias.

The Employment Equity Act (no. 55 of 1998) stipulates clear requirements when it comes to the use of 1) psychometric assessment tools and 2) selection and discrimination practices. The selection interview is used in selection as a psychometric assessment tool (Guion, 1998) and should therefore be aligned to adhere to the laws that govern the above mentioned areas.

8. Psychometric testing

Psychometric testing and other similar assessments of an employee are prohibited unless the test or assessment being used-

- (a) has been scientifically shown to be valid and reliable;
- (b) can be applied fairly to employees; and
- (c) is not biased against any employee or group.

Since interviews are so widely used and trusted in personnel selection, establishing their reliability and validity remains paramount. Beyond the financial implications of accurate and effective selection decision making, there also are legal implications. The Employment Equity Act (No. 55 of 1998) established clear guidelines for the use of personnel selection procedures in Chapter 2 (Prohibition of Unfair Discrimination) (Figure 1.1), prohibiting unfair discrimination on any non job-related grounds, including demographic group membership. More specifically, personnel selection procedures should, therefore, be shown to be free from any bias that systematically disadvantages any subgroup members that do not carry 'protected' status (see Figure 1.2).

From a legal perspective it would be to the benefit of personnel practitioners to scrutinise their interview procedures for the presence of these prohibitions. Using validated, reliable, and fair selection interviews that are largely deemed free from bias should be legally defensible and less prone to litigation.

2.6 Psychological Perspectives on Bias

The following section provides insight into the prevalence of similarity effects as proposed by an aggregation of psychological and social theory. Social Identity Theory provides the 'backbone' to this section, with many overlapping and corresponding insights from similar

psychological perspectives. The probability of interviewers making biased judgements and decisions again are highlighted, with reference to ‘tried and tested’ psychological theory.

People organise their complex worlds by organising information, classifying people and judging situations and decisions according to their cognitive ability. When a situation presents itself (as an event, person or other stimuli), individuals use formed schemas to make sense thereof and to categorise the stimuli ‘appropriately’ (Fiske & Taylor, 2008; Sacco et al., 2003). Sacco et al. (2003) propose that these schemas change over time as life is experienced more thoroughly and specifically. In the context a judgement, these schemas are unconsciously used to categorise the judgement outcome in terms of a perception an interviewer holds of the applicants’ demographic group or context. It is important to note that these schemas tend to change and be adapted as the context changes (Barlsalou, 1982 as cited in Sacco et al., 2003).

2.6.1 Social Identity Theory (SIT)

Social Identity Theory was conceptualised by Henry Tajfel. This theory is used as a framework for understanding the prevalence of similarity effects in selection interviews. SIT, as conceptualised by Tajfel, refers to the way in which (1) individuals categorise their world (things and people) and the fact that (2) individuals choose to associate themselves with something or someone (friend, sports team, ideology) (Jenkins, 2003). SIT is concerned with, among other things, the way in which individuals attach value to certain cognitive categories that they have established and how these schemas influence behaviour interpersonally,

socially and professionally (Capozza & Brown, 2000; Jenkins, 2003). Because of the inconsistency of how humans attach value to the same constructs, people and events, SIT introduces the possibility of similarity bias in judgement contexts – highlighting the possibility of one individual judging another through a lens of personal social identity that, to some extent, influences objectivity.

Various theoretical propositions, most of which come from SIT, could be put forward for why the similarity between interviewers and applicants could affect the ratings that are produced by interviewers. SIT is cited forward to explain why individuals show a preference for similar others (Goldberg, 2005). Uncovered intentions, if taken as fact, drive suspicion in terms of the motives with which interviewers judge applicants for selection purposes. It also confronts us with the idea that applicants, due to social identity, might not always be interested in a job where they realise they do not fit (socially), and might subconsciously portray a worse image of themselves that might be falsely attributed to rater bias. SIT provides many insights into the similarity effect framework, but has to be understood from its roots to gain the full perspective.

2.6.2 Similarity Attraction Paradigm

In order to comprehend the impacts and outcomes of SIT, it is insightful to take note of some theoretical developments that stem from the premise set by SIT. The similarity-attraction paradigm (Byrne, 1971) suggests that people feel attracted to similar others and, hence, could be expected to favour individuals that resemble themselves in various characteristics (Winter & Kjørlien, 2000). Similarity attraction proposes that actual (demographic) similarity

between people will lead to perceived similarity in terms of values and attitudes and that this will, in turn, lead to interpersonal attraction between people (Graves & Powell, 1996). Interpersonal attraction has been shown to lead to more favourable judgement (Dipboye & Macan, 1988).

2.6.3 Social Cognition Theory

According to this theory, people make sense of the world around them by storing information in appropriate cognitive categories or *bins*. These authors explain that, just as bins are used in a storehouse to store different substances within the same storehouse, the mind uses cognitive bins to store similar information separately. Since it is not possible for the human mind to have a complete and objective perception about all stimuli in the world, it tends to sort stimuli and add information that falls within specific categories to our memory bins (Kulik & Bainbridge, 2006).

Social cognition further proposes three types of cognitive categories, or memory bins, namely *schemas* (e.g. Kalin & Hodgins, 1984), *stereotypes* (e.g. Glick, Zion & Nelson, 1988) and *prototypes* (e.g. Fiske & Taylor, 2008). Though overlapping, each of these explains a unique part of human social cognition. Fiske and Taylor (2008) and Kulik and Bainbridge (2006) explain these cognitive categories, indicating that *schema* is the overarching term that represents knowledge of certain stimuli categorised to fit the same cognitive bin, while a *prototype* would be an individual unit of a particular categorised group that serves well as representative of that cognitive category. For instance, a dove might be a prototype for the 'birds bin', since it might represent all the features that you would expect to be typical of

birds. Stereotypes typically refer to people in groups and would organise information on the grounds of group membership. A stereotype would create expectations of a new member of a group on the grounds of previously aggregated perceptions of such group member (Fiske & Taylor, 2008; Kulik & Bainbridge, 2006).

The danger and opportunity for biased decision making is highlighted by social cognition in that ‘schemas’, ‘prototypes and ‘stereotypes’ are ‘works in progress’ that can hardly be expected to contribute to accurate and objective judgements (Fiske & Taylor, 2008). For example, when an individual perceives someone who has grown up in the same neighbourhood as himself, went to the same school and did the same sport, it would account for a much better understanding and acceptance of that individual since he fits a well-known stereotype. The same individual might be much less comfortable with someone from a different race group who comes from a completely different area, since the appropriate cognitive ‘bin’ is still very underdeveloped and might even be distorted due to a few random encounters with similar stimuli. According to social cognition theory, an individual might make judgements of another person not based on individual characteristics, but on the stereotype held with regard to the individual’s proposed group membership (Kulik & Bainbridge, 2006).

2.6.4 In-group Theory

In a further development, stereotypes can be explained as having three distinct properties, i.e., intergroup differentiation, in-group favouritism, and differential accuracy (DiDonato, Ullrich & Krueger, 2011). *Intergroup differentiation* refers to the phenomenon of people

perceiving groups as different when they are similar (Jetten, Spears & Manstead, 1998). This might result from an underdeveloped scheme or prototype that could ‘blind’ the observer to overt intergroup differences. *In-group favouritism* explains that people attribute more positive characteristics to groups to which they belong than to groups to which they do not belong (Hewstone, Rubin & Willis, 2002). *Differential accuracy* explains that individual perceptions of in-groups tend to be more accurate than their perceptions of out-groups, since they have more in-depth knowledge of the in-group that they are a part of themselves (Judd & Park, 1993). DiDonato et al. (2011) proposed an integrative reasoning model to tie the concepts of intergroup differentiation, in-group favouritism and differential accuracy together. The model assumes, from empirical evidence, that most people have a positive self-image and that they project this positive image towards in-groups more than towards out-groups (Didonato et al., 2011). This implies that intergroup differentiation and differential accuracy is prone to favour the in-group more than the out-group – due to the influence of in-group favouritism.

2.6.5 Summary

Aforementioned psychological perspectives lead to a few general conclusions that could typically be used to explain sources of idiosyncratic rater effects. From studying each of the above theories, paradigms and ideas, one can conclude that it is central to human behaviour to subconsciously (and sometimes even consciously) perceive those who are similar to the observer in a more favourable light than others less similar to the observer. In turn, this might play out in any judgement process as a bias enhancing factor – leaving ratees that are less similar to the rater in an immediately less favourable position than those more similar. Within

the human condition it is fairly common to find classification and association of people in terms of race. Neighbourhoods, countries, social classes and friend-groups mostly comprise racially homogenous units. Though exceptions exist and are promoted in socio-political contexts, race seem to be a major cause of people either regarding themselves as similar or different to those around them – race can therefore be expected to yield results, as an predictor variable, within a ‘similar-to-me’, ‘not-similar-to-me’ study.

2.7 Demographic Similarity Effects in Interviews

“Similar-to-me is a bias. Similar-to-the-ideal-candidate seems a useful match to an ideal prototype; if the prototype is valid, matching it should imply valid assessment as well” (Guion, 1998, p. 623). Demographic similarity is known as one factor that can produce error variance in judgement, rating and decision-making processes (Riordan, 2000). Similarity effects refer to the notion that an individual would hold another individual who is perceived to be similar to him/herself in a more favourable light than a less similar individual (Sacco et al., 2003). Similarity effects have been found to impact on various work outcomes and always imply unfair discrimination against individuals less similar to those who are in decision-making roles (Riordan, 2000; Sacco et al., 2003). In the following section, research findings on demographic similarity effects in interviews are discussed, compared and summarised to present an argument for the selection of race as demographic similarity variable for investigation in this study.

2.7.1 Research Findings: Demographic Similarity Effects

This section provides an overview of the research that has been conducted on demographic similarity effects in selection interviews. The studies are presented chronologically and more attention is given to the more recent studies that have the largest bearing on this study.

Similarity effects in the selection interview context have been widely researched, although findings have been inconclusive. Schmitt (1976) found evidence that racial and attitudinal similarity were related to higher interview ratings. Harris (1989) reported a rating bias towards applicants of the same sex. In a study on race and age similarity effects in conventional and situational interviews Lin, Dobbins, and Fahr (1992) found that ratings of Black and Hispanic applicants were higher when the interview panel consisted of their racial counterparts. They also concluded that fewer racial similarity effects were found with the use of the situational interview than with the conventional structured interview. No age similarity effects were found. Graves and Powell (1995) conducted a study on the effect that sex similarity has on recruiters' evaluations of applicants. They found that interviewers found members of the opposite sex more similar to themselves, although it effected only marginally higher ratings. Prewett-Livingston, Field, Veres, and Lewis (1996) investigated the effect that interviewer race, applicant race and racial composition of interview panels has on interview ratings. They reported that unbalanced interview panels rated racially similar others more favourably and that interviewers in the minority also conformed to the majority perception. From these results, support is found for the psychological and sociological theories discussed earlier, since some traces of demographic similarity effects were found. Interestingly, racial

similarity effects are found to exist as a direct effect, whereas sex similarity seems to be more prevalent as a dissimilarity effect.

Sacco et al. (2003) tested the effect of racial similarity effects on interview ratings. In their study, they used ratings from 708 interviewers of 12 203 applicants in highly structured one-on-one interviews. Using hierarchical linear modelling (HLM) (as opposed to ANOVA approaches used by earlier studies), they found no significant race or sex similarity effects in the highly structured one-on-one interviews. In HLM terms, there was no significant Y11 value ($p < 0.05$) which would be evident in a cross-level or interaction effect that would reveal demographic similarity effects. Small similarity effects, however, were found when interaction approaches based on *D*-score and analysis-of-variance (ANOVA) were conducted at the individual level of analysis. The effects that were found were not significant enough to invoke practical concern and highlighted the importance of accommodating nested data, as HLM does, above the ANOVA and *D*-score approach (Sacco et al.). The results of their study suggest that organisations using highly structured interviews for selection purposes need not be overly concerned about bias due to interviewer and applicant race and/or sex mismatch.

Goldberg (2005) investigated the effect that demographic similarity between interviewer and applicant has on selection decisions. The results showed sex dissimilarity effects (males rating females higher) related to interpersonal attraction and applicant appearance. Significant racial similarity effects were found for white recruiters only, while no age similarity effects could be traced. In a study examining implicit sources of bias in selection interviews Segrest-Purkiss, Perrewe, Gillespie, Mayes, and Ferris (2006) found that recruiters responded less favourably towards applicants who had an ethnic name and spoke with an accent. Buckley et al., (2007) conducted a study on the influence of demographic similarity

effects in the structured interview setting. They used twenty assessors, divided into racially equal panels, and showed them interview videos from which they had to rate the applicants in the videos. The researchers found small, but significant racial similarity effects related to the panel interviews utilising structured interviews. At this point, some conflicting results started to creep in, with Sacco et al., (2003) not finding any racial similarity effects in their large-scale field study while Buckley et al. (2007) did find effects in their more controlled laboratory study. Sex similarity still seemed to operate with a dissimilarity effect.

In a semi-replication study of the Sacco et al. (2003) investigation, De Meijer (2007) investigated for similarity effects within the Dutch police force, but this yielded no evidence of similarity effects. The hypothesis was that assessors of majority groups and minority groups would record higher ratings for similar-to-them applicants. Psychologists were engaged to conduct the structured interviews. Using HLM analysis, as promoted by Sacco et al., no significant Y11 value ($p < 0.05$) was found and the conclusion was that there were no similarity effects as was hypothesised, as in the case of the Sacco et al. (2003) study.

In the latest large-scale study of interviewer-applicant demographic similarity effects, McCarthy et al. (2010) found that highly structured interviews are highly resistant to demographic similarity effects in interviews, probably because interviewers are guided by job-relevant assessment criteria that make it difficult for them to have their judgement swayed by social judgement heuristics such as similarity effects. The study also used the HLM approach to analyse the data and yielded similar results (though even smaller effect sizes) to the de Meijer (2008) and Sacco et al. (2003) studies. Over and above the HLM analysis, McCarthy et al. (2010) also computed pseudo R^2 value and Cohen's d . These analyses yielded confirmatory results, with pseudo R^2 values of .00 and insignificant d values

throughout all relationships tested. The findings of the earlier studies supported the existence of demographic similarity effects in interviews to a greater degree than these latest large-scale studies by de Meijer et al. (2008), McCarthy et al. (2010) and Sacco et al. (2003) – all failing to find any significant effects.

The McCarthy et al. (2010) study proposes that highly structured interviews are resistant to demographic similarity effects, but can be criticised on the fact that there is no evidence that similarity effects would be found if they tested for it in low-structured interviews as well. The assumption that highly structured interviews are resistant to demographic similarity effects assumes the presence of effects with the use of low-structured interviews. In studies reported by Elliot (1981) and Wiley and Eskilson (1985) no effects were found with the use of low structured interviews. This can imply that similarity effects should never be a concern in the employment interview – whether the interview is structured or unstructured. If McCarty et al. (2010) had experimentally varied the interview structure in their sample and from such comparative data proved the resisting effect that structured interviews have over the less structured formats, it would have strengthened their case.

In summary, the magnitude of observed similarity effects varies, with some reporting no effects (e.g. Graves & Powell, 1995; McCarthy et al., 2010; Sacco et al., 2003), while others report small to moderate effects (e.g., Lin, Dobbins, & Fahr, 1992). Various reasons for the inconclusive nature of the body of empirical evidence that has addressed demographic similarity effects in interviews have been suggested, including the use of simulated interviews in some studies, sampling error, and as resulting from variations in the degree of interview structure (McCarthy et al., 2010).

2.7.2 Choice of Race as Similarity Variable

In similar studies to this one, a variety of demographic factors have been used to operationalise actual similarity. In most cases, though, race and/or sex (gender) were used. In order to contribute to the current vein of research, this study has not attempted any groundbreaking investigation into new demographic variables, but rather attempted to voids in the existing and reasonably inconclusive body of research.

In ten studies on racial similarity effects, five showed significant effects, although small in magnitude (Buckley et al., 2007; Goldberg, 2005; Lin et al., 1992; Prewett-Livingston et al., 1996; Rand & Wexley, 1975) while five showed no effects (De Meijer, 2008; Fiedler, 2001; McCarthy et al., 2010; McFarland, Ryan, Sacco, & Kriska, 2004; Sacco et al., 2003). It is interesting to note that four of the five studies that showed no effects were conducted using structured interviews (de Meijer, 2008; McCarthy et al., 2010; McFarland et al., 2004; Sacco et al., 2003), while the fifth one does not specify the structure condition used. (Fiedler, 2001). This strengthens the notion that highly structured interviews tend to be resistant to demographic similarity effects.

In eleven studies investigating gender effects, only one showed significant effects, small in magnitude (Walsh, Weinberg, & Fairfield, 1987), while ten showed no effects (Elliot, 1981; Gallois, Callan, & Palmer, 1992; Goldberg, 2005; Graves & Powell 1995; Graves & Powell 1996; McCarthy et al., 2010; Reid, Kleinman, & Travis, 2001; Sacco et al., 2003; Simas & McCarrey, 1979; Wiley & Eskilson, 1985). Dissimilarity effects have also been hypothesised

and found in gender studies (e.g. Goldberg, 2005). These findings cast a dim light on the fruitfulness of future study into gender effects, at least in the way that it has been approached in the past.

There is much clearer evidence supporting the existence of racial similarity effects than gender effects in interview ratings. In this light, the study proposed to focus only on racial similarity as the existence of gender effects can, to a large extent, be questioned and might be more complex than has been hypothesised to date. Taking a more in-depth look at the working of sex-similarity and dissimilarity effects in interviews could be considered for future research, to perhaps find a fresh theoretical approach to understanding its influence in personnel selection. Demographic similarity effects, if and where they exist, should be understood thoroughly and measures should be put in place to remove such bias from the selection interview.

2.8 Reducing Bias in Interviews

This section presents theory about ways to remove as much bias as possible from the selection interview by means of interview design. The previous section presented the prevalence of possible demographic similarity effects, which immediately put the organisational psychologist to work to manage this unwanted source of human behaviour.

In a meta-analysis of interview validity, Marchese and Muchinsky (1993) hypothesised six factors to be moderators of interview validity and found interview structure to be the strongest moderator, with highly structured interviews proven more valid. The following

section presents an argument for the difference between basic interview structures and formats and concludes with a probable answer to the question: Which type of interview will assist in removing the most bias, in terms of racial similarity effects, from the selection interview?

2.8.1 Defining Interview Structure

There are three basic ways to distinguish between a structured and an unstructured interview (Blackman, 2002). First, in structured interviews, all interviewers are to put the same set of questions to all applicants, with little follow-up or additional probing questions. Second, structured interviews are based on a thorough job analysis from which the questions are formed. Lastly, with structured interviews, interviewers are more often than not trained to use standardised rating sheets to assess applicant answers with (Blackman, 2002; Campion et al., 1997; Campion et al., 1988).

Furthermore, and with more detail, Campion et al. (1988) proposed that the highly structured interview has the following six characteristics: (1) It is developed from a thorough job analysis; (2) it is set up to ask the same set of questions from each candidate; (3) the answering is anchored with example answers and ratings, therefore; (4) an interview panel should rate applicant answers as opposed to a single interviewer; (5) the process is consistently administered to all candidates to ensure standardisation and (6) special attention is given to the job-relatedness and fairness of interview processes and questions.

Unstructured interviews, on the other hand, can be defined as being void of any of the above-mentioned elements (Guion, 1998). Unstructured interviews will allow interviewers the freedom to make-and-shape each interview as they think appropriate without using any standardised rating sheet (Wiesner & Cronshaw, 1988). An interview can furthermore be classified as unstructured if the interviewer is expected to make one subjective overall rating of each applicant, instead of behaviourally anchored ratings on each question asked (Wiesner & Cronshaw, 1988).

Kohn and Dipboye (1998) also presented an argument for semi-structured interviews, challenging the often dichotomous view of interviews either being structured or unstructured. They labelled semi-structured interviews as those that employ some of the elements of structure (referring to the framework proposed by Campion et al. (1988)), but not all of them. It therefore is important to understand and clarify which type or structure of interview is most appropriate for effective, error-reduced, selection decision making.

2.8.2 To Structure or not to Structure

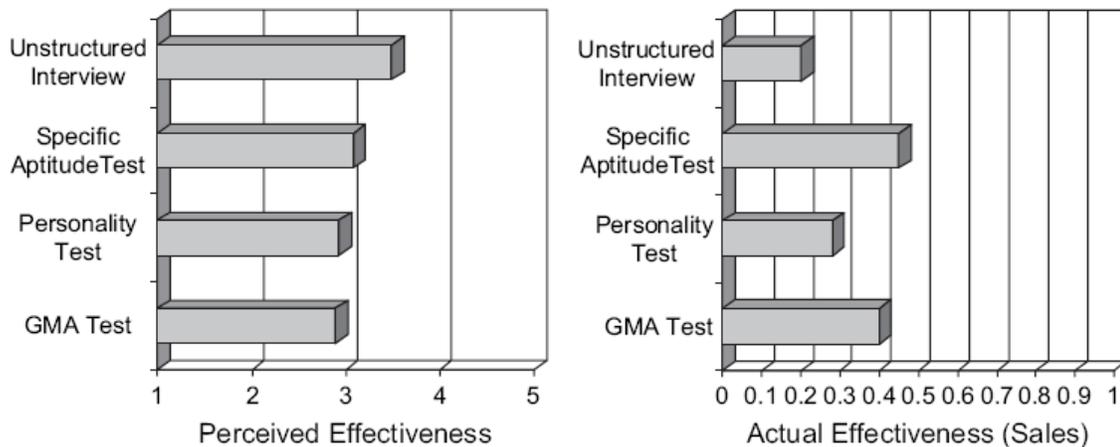
In the debate between promoters of differing interview designs, many more battles have been won by the highly structured faithful. Structured interviews have been shown to provide greater validity, reliability and better overall ability to accurately predict future job performance than unstructured interviews (Campion et al., 1988; Campion et al., 1997; Cronshaw & Wiesner, 1989; Schmidt & Hunter, 1998). In a meta-analysis undertaken by Conway et al. (1995) it was found that the upper limits of validity of structured interviews were estimated at .67 while at only .34 for unstructured interviews. Further it is important to

note that structured interviews are proposed to be resistant to the influence of demographic similarity effects (McCarthy et al., 2010; Sacco et al., 2003).

In a study by Highhouse (2008), the perceived effectiveness of common selection methods were correlated with the actual effectiveness thereof. Actual effectiveness was operationalised as a performance outcome ('number of sales') and is therefore a good representation of the future job performance criterion in personnel selection. Though minor discrepancies were found for most of the selection methods, the perceived effectiveness of the unstructured interview significantly superseded the actual effectiveness thereof (Figure 2.4). The unstructured interview was deemed by practitioners to be more effective than specific aptitude tests, personality tests and GMA tests, but was proven to be the least effective, as indicated by the graph on the right hand side of Figure 2.4.

Figure 2.6.

The perceived vs actual effectiveness of selection instruments



An argument in favour of unstructured interviews as a better measure of personality dimensions than the structured interview has been presented by Blackman (2002). Even if this can be accepted as fact, which is debatable, it seems unfruitful to fully neglect the predictive validity of the structured interview to obtain a better personality indication, which can more accurately be obtained by proven personality questionnaires. Other reasons for using unstructured interviews include ignorance, laziness, intuitive appeal and convenience, all of which provide a dim outlook on the fairness and accuracy of general selection procedures (Blackman, 2002).

Aggregated findings seem to support the idea that structured interview settings provide more resistance to demographic similarity effects in selection interviews. From a meta-analysis, Huffcutt (2011) reported that racial similarity studies using structured interviews showed averaged similarity effect sizes of $d = .23$, while the use of unstructured interviews showed a mean effect size of $d = .32$, supporting the notion that interview structure moderates racial similarity effects across studies sampled in their meta-analysis.

Some researchers suggest that the use of situational interviews as a form of structured interviewing provides even greater validity coefficients than the normal structured interview (Latham, Saari, Pursell, & Campion, 1980); Lin et al., 1992). Structured situational interviews differ from normal structured interviews in that all the questions are in a “*What would you do if...*” format and are derived from preconducted job analysis (Lin et al., 1992). Structured, situational interviews have been found to be superior to normal structured interviews in terms of higher validity coefficients (Latham, 1989). Lin et al., (1992) found much stronger similarity effects in conventional structured interviews than in situational

interviews. Latham, Saari, Pursell, and Campion (1988) also found that situational interviews do not discriminate against blacks and women, which indicates a supposed ability to assist in eradicating demographical bias. Thus, it can be surmised that the structured situational interview would provide the most resistance to demographic similarity effects, whereas the unstructured interview could potentially be more vulnerable to such effects.

In agreement with the above research McCarthy et al. (2010) hypothesised that high interview structure would cause interviews to be resistant to demographic similarity effects. Their findings confirmed this hypothesis, as in a similar study done by Sacco et al. (2003). These findings can only be generalised tentatively since other studies that used the more unfavourable unstructured interview found only very small effects (Elliot, 1981), similar to the size of the effects reported in the McCarthy et al. (2010) and Sacco et al. (2003) studies. Furthermore, in a study by Buckley et al. (2007), conducted with the use of situational interviews (that have been proven by Lin, Dobbins and Fahr (1992) to yield fewer similarity effects than the conventional structured interview), significant race effects were indeed found. The moderating effect of structure, therefore, cannot be assumed or accurately derived from the aggregation of these results; it seems clear that many other variables influence interview ratings and that they were (obviously) not held constant throughout this range of studies. In light of this it is apparent that the only way in which the moderating effect of structure can accurately be determined is by varying interview structure experimentally and comparing results from the different structure options. The intention of this study therefore was to experimentally vary interview structure, holding other variables constant, in order to measure the moderating effect that low-, medium- and highly structured interviews have on

the influence of demographic similarity effects on interviewer judgements and subsequent interview ratings.

2.8.3 SIT and Interview Structure

As previously explained, SIT proposes that underlying cognitive schemas of individuals might influence their judgement. The structuring of interviews might have a limiting effect on this social phenomenon by focusing the interview on specific questions and providing a behaviourally anchored rating sheet that prompts ratings of predicted future performance and not intuition- and experience-based ratings (McCarthy et al., 2010). With the use of highly structured interviews, all applicants are presented with the same set of questions, thereby limiting social influence that might otherwise direct the interview in an unstructured setting. These are the reasons for having hypothesised and partly proven that structured interviews provide resistance to demographic similarity effects (McCarthy et al., 2010; Sacco et al., 2003).

2.9 The South African Context

The present study addresses a gap in existing literature on demographic similarity in interviews, since it directly tests the role of interview structure on the existence of interviewer-interview similarity bias. As a corollary, it provides a much needed replication of earlier studies that have sought to investigate similarity effects in interviews that have exclusively been published in the USA in Europe. There are compelling reasons to expect that these findings would not be generalisable to the South African labour force context,

which is characterised by a socio-political history and diverse race group composition that differs markedly from the American and European contexts.

2.10 Research Problem

Some confusion around the existence and prevalence of demographic similarity effects in interviews has definitely been created by discordance between theory and practical research and also because of conflicting findings in research. It is clear from the theory that the social mind of a human being is unconsciously very susceptible to making similarity judgements. The powerful theory does, however, not translate into research results of corresponding magnitude. This immediately raises the question of what other variables are mediating or moderating the influence of these similarity effects on interview ratings and, in the process, damping their flare; or rather their poison.

Interview structure is promoted as a moderating variable between demographic similarity effects and the validity of interview ratings. Existing studies have, however, not succeeded in modelling and testing for this moderating effect with regard to its influence on demographic similarity effects. To some extent, the case has been closed in research by claiming that highly structured interviews are resistant to demographic similarity effects (McCarthy et al., 2010). Even so, this does not declare every employment interview safe from these effects, since 50% of recruiters have been shown to use unstructured interviews (Ryan et al., 2004).

Racial similarity has clearly been shown in the literature as the (1) the most prevalent in similarity research and (2) the demographic variable that has shown the greatest moderating

results, i.e. danger in the employment interview. If this is the case, it will be fruitful to pursue the clarity of the influence of demographic similarity effects on interviews, with the variable that tends to show the most potential for (1) providing with notable effects and (2) having legal implication in terms of paragraph eight of the Employment Equity Act (no. 55 of 1998).

2.11 Hypotheses

The present study takes a critical stance on the way authors like McCarthy et al. (2010) and Sacco et al. (2003) assumed the ability of the structured interview to provide resistance against demographic similarity effects. In these studies, only structured interviews were used - whether or not they would have revealed demographic similarity bias with the use of less structured interviews cannot be known for sure. The size of the resistant effect that interview structure has on demographic similarity effects can therefore not be assumed and can only be determined by experimentally varying interview structure within the same sample and by keeping all other factors constant.

In agreement with most authors cited in the literature review, as well as the social logic derived from psychological theory, this study proposed the existence of demographic similarity effects to the extent that it would influence interview ratings. In other words, it was supposed that the race of the interviewer would have a moderating effect on the relationship between applicant race and applicant rating. This investigation agrees with the way that McCarthy et al. (2010) proposed that highly structured interviews would be resistant to

demographic similarity effects, but experimentally varied interview structure to prove the way in which structure does indeed provide resistance to racial similarity bias.

The following general hypotheses were formulated:

Hypothesis 1

Interviewers using low-structured (unstructured) interviews will award higher ratings to same-race candidates compared to different-race candidates.

Hypothesis 2

No race-similarity or dissimilarity effect will be found in ratings by interviewers of same-race candidates compared to different-race candidates when using highly structured, situational interviews.

In literature reviews on interview studies where interviewer and applicant main effects have been hypothesised and tested, results are shown to be mostly inconclusive (Macan, 2009; Posthuma et al., 2002). Main effects refer to systematic differences in average group scores, in this case, with groups defined in terms of some demographic characteristic (Sacco et al., 2003). In this light, this study did not attempt to hypothesise main effects, but was designed to test for and report on them in the process of determining interaction, or cross-level effects, as will be explained more comprehensively in Chapter 3.

2.12 Implications of Expected Results

If similarity effects were found, as hypothesised, it would imply that they contribute to the error variance of the employment interview process and should be minimised to the extent possible. Since the South African workforce is characterised by high heterogeneity at all levels except for senior management, as found by the latest Employment Equity Report (2010), the results of this study are expected, if the hypotheses were confirmed, to lead to some practice and policy recommendations towards minimising possible similarity effects. In order to select and develop previously disadvantaged individuals in the South African workforce, organisations need objective ratings of interview dimension performance that are unbiased by demographic similarity effects. The possibility exists that this research study could contribute to affirmative development in the sense that it could suggest design and training interventions that are required to improve the quality of interviewer ratings. If the hypotheses were confirmed concerning the moderating influence of structure, this also could invoke practical considerations in terms of using either structured or unstructured interviews to produce error-reduced selection decisions. The validity of selection procedures such as interviews could be improved by eliminating possible interviewer bias caused by similarity effects.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Introduction

This chapter introduces and explains the methodology used to determine the effect that interview structure has on the influence of racial similarity in simulated selection interview judgements.

The objectives that have been set in the preceding chapters include determining the size of possible racial similarity effects on interview ratings; determining the moderating effect that interview structure can have on such effects, and also the power it has to remove the influence of similarity effects and testing the generalisation of results of this South African study to similar studies done abroad. Research was undertaken to determine the most effective way of reaching these objectives and testing the subsequently formed hypotheses. In summary, this chapter provides details of the research design, statistical hypotheses, sample information, data collection and analysis techniques.

3.2 Research Design

Research design refers to the plan and structure that is followed in order to gain the data and information the researcher needs to answer the relevant research questions (Theron, 2010b).

This study made use of an experimental design and was quantitative in nature. An experimental research design is used to compare two or more occurrences of a similar process (Kerlinger & Lee, 2000). These occurrences are differentiated by manipulated factors or variables, the impact of which could be determined when the difference in outcome was

observed and compared (Kerlinger, & Lee, 2000). In one sample (occurrence 1), interviews were unstructured and in the other sample (occurrence 2), they were highly structured. The extent to which racial similarity effects were found in the differing structures, were tested for, as hypothesised.

The notation for the design of this study can be represented as:

$$\begin{array}{rcc}
 & X_1 & Y_1 \\
 [R_m] & X_m & Y_2 \\
 & X_h & Y_3,
 \end{array}$$

where

[R_m] = Random assignment to different treatments

X₁ = Low structured interview condition (independent variable)

X_m = Medium structured interview condition (independent variable)

X_h = Highly structured interview condition (independent variable)

Y₁ = Cross-level effect in low-structured interview

Y₂ = Cross-level effect in medium structured interview

Y₃ = Cross-level effect in highly structured interview

3.2.1 MAXMINCON

Kerlinger and Lee (2000) maintain that any research design should have the ability to control variance through MAXMINCON. MAXMINCON refers to the maximisation of systematic variance, the minimisation of error variance and the control of extraneous variance (Theron,

2010). Each of these three requirements are briefly defined and then followed by an explanation of the way in which this study attempted to MAXMINCON effectively:

- MAX - The maximisation of systematic variance in order to increase the likelihood that H_0 will be rejected during statistical testing
- MIN - Minimisation of error variance in order to increase the likelihood that the effect of X on Y becomes “visible” amongst or discernable from the effect of other, non-relevant Xs on Y – which could be enhanced by ensuring the reliability of the measurements (Theron, 2010)
- CON - Control of extraneous variance. This can be addressed by standardising the procedure for both conditions and controlling what factors are to be kept constant and what is to be induced experimentally.

This study, with its experimental design, has been brought into thorough submission to healthy MAXMINCON controls. First, the sample of applicants in the video stimuli was the same in both condition groups, as was the case with the rater groups. Thus, the same raters rated the same applicants, with the only difference being the level of structure and the individual responses to the differing questions in the two structure conditions. Second, the same setting was used for all the data capturing opportunities. Lecture halls were used and the setting always was of an academic nature with the intention of transferring skill and knowledge in terms of the development and conducting of interviews. Third, the video stimuli provided additional control to a field study in that all raters had opportunity to rate the same responses by the same applicants – enriching the comparative value of the results. Fourth, all the raters received the same instructions, differing for the two conditions, but constant within each one, to ensure that all the raters had the same understanding and instruction on how to rate the applicants. Fifth, all the raters had exactly the same amount of

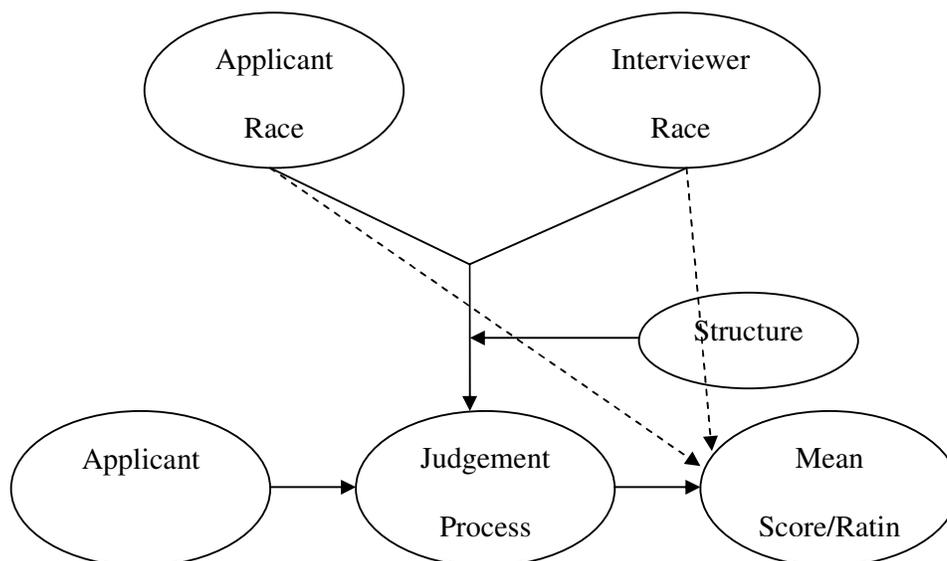
time to judge the applicants and decide on the appropriate ratings they should give to each applicant or each answer.

3.3 Conceptual Model

This model serves to clarify the processes and effects that are investigated, and present the way the variables interact with one another. The dashed lines indicate the main effects that were to be tested for. Main effects refer to variance in ratings of different race groups at the applicant or interviewer level only. The interaction effect between interviewer and applicant race was then hypothesised to influence the judgement process that would cause incremental variance in subgroup ratings. Lastly, interview structure was hypothesised to moderate the size and significance of the cross level effect (i.e., influence of racial similarity bias on interview ratings).

Figure 3.1.

Conceptual model of the interview judgement process influenced by racial similarity and interview structure



3.4 Stimulus Development

Video recordings of mock interviews differing in interview structure were made for the purposes of this study. Sixteen interviews of between three and ten minutes each were recorded. Eight interviews used a highly structured format and eight an unstructured or low-structured format. Only eight ‘applicants’ were used, thus each of them were interviewed in each of the structured conditions. The group of eight consisted of two white males, two black males, two white females and two black females. Although gender effects were not to be tested for in the proposed study, keeping the distribution even between males and females was a precaution against extraneous variance that might have been caused by possible gender effects.

Interviewees were sampled randomly from among under- and postgraduate students at Stellenbosch University. Two of the interviewees were already working and were not students, although they fell within the age category of the rest of the interviewees. It was explained to all participants that a mock job interview for a generic supervisory position at an imaginary company was being conducted. They, furthermore, were asked not to elaborate on their personal field of study or practice, but to keep their answers generic in order to promote the ideal of being able to use the videos as stimuli in a non job-specific environment.

All the interviews were conducted in the same setting, most of them within the same conference room at the J.S. Gericke Library at Stellenbosch University. Participants sat in front of a camera, at a desk opposite the researcher. Only the participants were recorded in the videos, as the demographics of the researcher were expected to account for extraneous

variance. The researcher conducted and recorded the two interviews back-to-back with each of the eight participants.

The recorded videos were uploaded onto the researchers' computer and edited using NCH's VideoPad Video Editor (2010) software. The videos were edited by muting the voice of the researcher and by displaying the questions asked as text on the screen, while the reactions of the participant could still be observed. The sound was activated once the participant began to answer the relevant question. The eight videos of each structured condition were linked together to form two video clips that were used as stimuli in the research.

3.5 Measuring Instruments

Measurement instruments used in the study comprised highly structured and unstructured interview rating sheets. The high-structure rating sheet was developed as a competency-based, situational interview that took the form (though not exclusively) of "*What would you do if...*" -type questions typical of a situational interview (Lin et al., 1992). (See Addendum B)

Situational interviews are found to be more resistant to demographic similarity effects than the conventional structured interview (Latham, Saari, Pursell, & Campion, 1988) and also produce higher predictive validity coefficients (Latham, 1989) than the conventional structured interview. It is for these reasons that the situational interview was chosen as representative of highly structured interviews for the purposes of this study.

The highly structured interview consists out of eight questions relating to two competencies – *communication* and *people management*. These competencies were shown by the CIPD (2008) survey to be among the most commonly used competencies by managers for interviewing. These two competencies were each assessed with four situational type questions for the purposes of the highly structured rating sheet. The questions were behaviourally anchored and were rated on a 7-point Likert scale. Each question had an accompanying *rating guide* which was used to guide raters with regard to the most ideal answer (for a score of 7), to an average answer is (a score of 4) and an example of a very bad answer (a score of 1). Raters were free to score any value between 1 and 7, as they saw fit.

The guidelines for developing and conducting a highly structured interview, as previously mentioned, and explained by Campion et al. (1988), were taken into consideration for the development of the highly structured interview. The only one of the six ‘prerequisites’ proposed by Campion et al. (1988) that was not practically feasible in this study was the rating by a panel of interviewers. It was the purpose of the current study to determine the effect of one-on-one racial similarity on interview ratings, therefore panels were not used. In an attempt to derive the effect to a panel, the instructions to the interviewers sketch the scenario of being part of a panel of diverse interviewers, to whom they were accountable in terms of their rating of each candidate.

The content of the unstructured interviews was based on five very broad and unspecific questions found, through various informal interview guides on the internet, to be frequently asked questions in selection interviews. (See Addendum A). The raters were given an opportunity to watch each of the ‘candidates’ in the videos and compare them by rating each ‘candidate’ on an overall score of 100. To best simulate the process and characteristics of an

unstructured interview, no rating guide, rating scale or competency-based questions were included.

3.6 Sample

A non-probability sample of convenience was used for the purposes of the research study. The sample group of approximately 200 mock interviewers consisted of undergraduate students at the University of Stellenbosch. The participants in the target sample were relatively similar with regard to educational background, academic interests, social class and age. They therefore comprised an excellent sample for experimental research, since many variables were more constant than it would have been in a sample of a more general population. This study was more concerned with the moderating effect of interview structure than the generalisability of the actual similarity effects to the general working population. The supposition was that, if there indeed had been similarity effects and they were found to be moderated by the degree of interview structure, there would be cause for replicating this study as a large field study. The researcher argued, as explained in Chapter 2, that structure, as a moderating variable, would have the same influence on any sample, since it removes the same amount of subjective judgement from any interview context, due to the standardising effect that structure brings. Generalisability would be a concern with the ratings in the unstructured interview, since the motives and incentives to judge would be different in an undergraduate class setup to those in an actual selection procedure. This limitation is discussed in more detail later, but, for this reason, the study should be replicated as a field study.

Participants were requested to participate in the research as part of interview training. Participants were randomly assigned to either to the high- or the low-structured conditions within a large-group context. Participation was voluntary. The following tables provide the descriptive statistics of the sample as computed for multilevel analyses, for the two conditions:

Table 3.1

Level-1 (applicant race) sample statistics for the low-structured condition

Variable name	N	Mean	Sd	Minimum	Maximum
Applrace	1560	0.50	0.50	0.00	1.00

N = 195 x 8 (each rater rated 8 applicants).

The mean of .5 indicates the equal amount of white and black applicants in the sample. Although there were only four white and four black applicants, they were rated by 195 interviewers and therefore the n = 195.

Table 3.2

Level-2 (interviewer race) sample statistics for the low-structured condition

Variable name	N	Mean	Sd	Minimum	Maximum
Raterrace	195	0.21	0.40	0.00	1.00

The mean of .21 indicates the ratio of white to black interviewers. There were 195 interviewers in the sample of which 41 were black and 154 were white.

Table 3.3

Level-1 (applicant race) sample statistics for the highly structured condition

Variable name	N	Mean	Sd	Minimum	Maximum
Applrace	1462	0.50	0.50	0.00	1.00

N = 185 x 8 (each rater rated 8 applicants).

The mean of .5 indicates the equal amount of white and black applicants in the sample. Although there were only four white and four black applicants, they were rated by 185 interviewers and therefore the n = 185.

Table 3.4

Level-2 (interviewer race) descriptive statistics for the highly structured condition

Variable name	N	Mean	Sd	Minimum	Maximum
Raterrace	185	0.13	0.34	0.00	1.00

The mean of .13 indicates the ratio of white to black interviewers. There were 185 interviewers in the sample of which 24 were black and 161 were white.

3.7 Procedure

The recorded videos were sorted and packaged on the basis of the degree of structure. As mentioned earlier, there were two video packages with eight recordings each - one of a highly structured format and one of low-structured format.

The sample of undergraduate students was used within the setting a formal academic lecture. The setting for the low-structured interview was a lecture hall with between 150 and 300 students and a data projector on which the videos were shown. The same students were used during their practical classes for the highly structured interview data collection. The structured interview data collection formed part of a class on interviewing and interview development and was preceded by brief interviewer training to establish a level of confidence and competence for the data collection phase.

3.7.1 Low-structured Data Collection

The data were collected during two theoretical classes of a second-year group of students. The videos of the unstructured interview were set up on a projector and the consent forms and rating sheets were distributed on the desks in preparation for the class. When the class was seated, the researcher was introduced and he briefed the class on the completion of the consent form and the process that they would be engaged in – the rating of the applicants shown in the unstructured interview videos.

The class was instructed to rate each of the eight applicants on a score of 100. The job the applicants were ‘applying’ for was not specified, as this would have required specific knowledge for judgement that all the raters could not be expected to have since all of them did not specialise in the same field. The position applied for was defined as a ‘general supervisory position at an imaginary company’ where part of the management that had to facilitate the selection process. They were allowed to compare applicants portrayed in the videos and to rate any candidate at any time during any of the interviews. They were instructed to present eight scores out of 100, one for each of the applicants, based purely on

their personal opinion the applicants. In addition, the respondents were required to provide basic demographic information at the end of the rating sheet.

The videos were started and were shown to and rated by the respondents without any interruption. After the rating process, a post-experiment debriefing was conducted and the consent forms and rating sheets were collected when the class was thanked and dismissed. The same procedure was followed exactly for both the classes, with the exception that there was insufficient time for a full debriefing with the last class. The debriefing was abbreviated to fit the time that was available.

3.7.2 High-structure Data Collection

The data were collected during several practical classes attended by a group of a second-year students – the same class group that was used for the low-structure data collection. The videos for the structured interview were set up on a projector and the consent forms and rating sheets were distributed on the desks in preparation for the class. When the class was seated, the researcher was introduced and he conducted a 15-minute training session on the development, use and utility of the structured interview. After this brief introduction to structured interviews, the raters were introduced to the rating sheets they would use for the rating process. All questions the class had were answered until the researcher was satisfied that they all understood the rating sheet and the process they were to engage in.

The participants were instructed that they had to rate each of the eight applicants on each question answered, according to the behaviourally anchored rating sheet, as explained earlier. The job that the applicants were ‘applying’ for was not specified as this would require

specific knowledge in judgement that all the raters could not be expected to have since all of them did not specialise in the same field. The position applied for was defined as a ‘general supervisory position at an imaginary company’ where they formed part of the management that had to facilitate the selection process. Each interview lasted for between 5 and 10 minutes and the raters had to complete most of the rating within that time, with about 30 seconds in between interviews to complete their ratings. In addition, the respondents were required to provide basic demographic information at the end of the rating sheet.

The videos were started and shown to and rated by the respondents without any interruption. After the rating process, a post-experiment debriefing was conducted and the consent forms and rating sheets were collected when the class was thanked and dismissed. The same procedure was followed exactly for both classes.

3.8 Data Analysis

In this section, the two methods of analysis used to test the hypotheses formulated for this study, are explained. The techniques that were used are HLM (Hierarchical Linear Modelling) and SPSSs’ Multiple Regression Analysis.

3.8.1 HLM Analysis

This study operationalised similarity effects in a way referred to in the literature as a *cross-level effect* or a *cross-level interaction* (Rousseau, 1985 as cited in Sacco et al., 2003). *Cross-level effects* occur when an independent variable from one level of analysis moderates relationships at another level of analysis. Even though interviewers are not a level of analysis

in the same way that, for instance, work units are, applicants in this study were nested within interviewers as each interviewer had to rate multiple candidates (Sacco et al., 2003). Therefore, the levels of analysis in this study were: L1 – applicant level and L2 – interviewer level. This study accordingly examined whether the race of the interviewer moderated the relationship between race and interview ratings at the applicant level, and also tested for moderating effect that differing interview structures has on this relationship.

Other ways of operationalising similarity effects have been proposed, but they, for the most part, fail the ideal on one or more methodological grounds (Sacco et al., 2003). The basic ANOVA approach that models the race of the rater as a third variable that may determine the relationship between applicant race and applicant rating has been presented. This is in line with the way in which this study has modelled similarity effects, but it can only be used if there are equal observations for each rater (Sacco et al., 2003). Since there were a quite a few instances of missing values that resulted in list-wise deletion of cases, each rater no longer had the same eight applicants nested – causing unequal applicant nesting within raters. Another approach to investigating similarity effects is by calculating D-scores that “...index the dissimilarity between an individual and another individual” (Sacco et al., 2003). D-scores have been criticised for a variety of conceptual and methodological reasons (Riordan & Shore, 1997; Sacco et al., 2003).

The data were analysed by using Hierarchical Linear Modelling (HLM), which is intended to accommodate nested or multilevel data structures (Raudenbush & Bryk, 2002). This technique of analysis was seen as appropriate for this study since a number of applicants were nested within each interviewer. A series of models are evaluated in order to test the

hypothesis when HLM is used as proposed (De Meijer, 2007; Raudenbush & Bryk, 2002; Sacco et al., 2003).

In order to effectively test for racial similarity effects, the HLM analysis procedure firstly determined the existence and size of any interviewer or applicant main effects. Main effects refer to whether the race of the interviewers has a direct influence on applicant ratings and whether the race of applicants has a direct effect on their ratings. When these effects are determined, the interaction effect of interviewer-applicant race can be determined.

Because HLM-nested data structures are similar to levels of analysis in an organisation, the applicant effect was termed Level 1 (L1), while the interviewer effect was referred to as Level 2 (L2). Part of the utility of using HLM has to do with the fact that it allows the researcher to assess whether the L2 variable (in the case of this study, interviewer race) impacts outcomes at L1 (applicant level) or moderates the relationship between the L1 predictor and L1 outcomes, respectively (Sacco et al., 2003). It is the latter moderation that this study was most interested in, in trying to determine whether the L2 independent variable, interviewer race, moderated the relationship between applicant race and interview rating. In other words, the question concerned whether the race of interviewers account for significant deviation in overall mean scores of different applicant race groups.

The first step of the analysis involves the testing of two basic regression equations. These equations yield results on within- and between-group variance (Sacco et al., 2003). It estimated and compared mean differences of interviewer scoring and applicant scores based on applicant and interviewer race. Because there are no predictor variables used in the model below, it is referred to as a *null model*:

$$\text{L1: } Rating = \beta_{0j} + r_{ij} \quad [1]$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad [2]$$

In the above equations, L1 predicts applicants' interview ratings based on the mean rating (i.e., intercept) within each of the j interviewers (β_{0j}) and the error for each of the applicants (r_{ij}). The L2 equation represents the interviewer intercept as the grand mean (γ_{00}), together with each interviewer's deviation (u_{0j}). Importantly, the associated variance components of these error terms are used to calculate an intraclass correlation coefficient (ICC). The ICC indexes the ratio of variance between interviewers to the total variance. A significant ICC value, therefore, grants the utility of further investigation with HLM, since it indicates a significant nesting effect. If an insignificant ICC value is obtained, HLM will provide with little additional utility compared to a basic OLS regression approach (Sacco et al., 2003).

If the first step of the analysis proved that there was significant intercept variability, the next step was to add a predictor to the L1 equation. The predictor, in line with the purpose of the study was applicant race. Equations such as L1 (equation 3) are called *random coefficient regression models*, since the coefficients β_{0j} and β_{1j} are modelled as random effects in equations 4 and 5.

$$\text{L1: } Rating = \beta_{0j} + \beta_{1j}(race_{App}) + r_{ij} \quad [3]$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + u_{0j} \quad [4]$$

$$\text{L2: } \beta_{1j} = \gamma_{10} + u_{1j} \quad [5]$$

These coefficients comprised the overall mean (γ_{00}) and the slope (γ_{10}) for each interviewer. Significant L2 parameters indicated that ratings were significantly different from zero and

that applicant race was related to ratings. The error variance components u_{0j} and u_{1j} indicate whether there was significant variability in the corresponding coefficients at L1. It is important, furthermore, to note that these models can only be estimated for interviewers who have rated applicants of both races. For this reason, videos with both black and white applicants were shown to each interviewer in this study.

The next step is taken if significant variability in the intercepts is found in the previous step. In this step, an L2 variable (interviewer race) was added as a predictor in the L2 equation. This variable was added as predictor of the variability of the intercepts at L1:

$$\text{L1: } Rating = \beta_{0j} + \beta_{1j}(race_{App}) + r_{ij} \quad [6]$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(race_{Int}) + u_{0j} \quad [7]$$

$$\text{L2: } \beta_{1j} = \gamma_{10} + u_{1j} \quad [8]$$

The above model is known as the *intercepts-as-outcomes* model and it tests for significant differences in mean ratings as a function of interviewer race. If this model shows significant variability in the applicant race slope, the following and final series of equations are estimated:

$$\text{L1: } Rating = \beta_{0j} + \beta_{1j}(race_{App}) + r_{ij} \quad [9]$$

$$\text{L2: } \beta_{0j} = \gamma_{00} + \gamma_{01}(race_{Int}) + u_{0j} \quad [10]$$

$$\text{L2: } \beta_{1j} = \gamma_{10} + \gamma_{11}(race_{Int}) + u_{1j} \quad [11]$$

This is called the *slopes-as-outcomes model*, since the race of the interviewer is used to predict the variability in the slope of L1 (applicant's race). In this model, a significant γ_{11} coefficient indicates that there is a cross-level interaction, or, in this case, significant racial similarity

effects. A significant γ_{11} coefficient then provides the evidence that interviewer race moderates the relationship between applicant race and mean interview rating. A significantly positive γ_{11} value would confirm hypothesis one in the low-structured sample and an insignificant ($\gamma_{11} \approx 0$) value would support hypothesis three in the highly structured interview sample.

One further consideration is the centring of the data. There are three approaches to centring in HLM; grand mean, group mean and no centring (Sacco et al., 2003). None of the approaches are better or more correct than the other, although they might yield different results which should be interpreted appropriately (Raudenbush & Bryk, 2002). The choice of an appropriate centring approach should be related to the specific research question (Sacco et al., 2003). In this research, the intercept at L1 is meaningful when it is not centred, since the intercept represents the mean rating for the race group coded as zero (Sacco et al., 2003). Even though this was not of primary concern in this study, the researcher believed that it would be easier to interpret in such a format.

3.8.2 Regression Analysis

Multiple regression analysis is a statistical technique used to observe the relationship between a dependent variable X and one or more independent variables Y_j . The regression coefficients in the regression equation are estimated using the method of least squares (Tabachnick & Fidell, 2007). In this way, the sum of squared residuals between the regression plane and the observed values of the dependent variable are minimized. The regression equation represents a plane in a $k+1$ dimensional space in which k is the number of independent variables, plus one dimension for the dependent variable (Tabachnick & Fidell, 2007).

The multiple regression technique was used in this study as a secondary method of analysis, only to confirm results in the HLM analysis and also to test the extent to which HLM did, in fact, model the hypothesised effects better, as proposed by Sacco et al. (2003).

3.9 Statistical Hypothesis

Two separate analyses were conducted for the data in each of the two structured conditions. The end of each HLM analysis was expected to yield a γ_{11} value that would indicate the cross-level effect size, i.e. the similarity effect size. These three γ_{11} values were to be compared and the relationships tested, as hypothesised. The analysis, as outlined above, would therefore be performed for each condition, in which the following hypotheses were tested for: (The omitted regression equations [1], [2], [4], [5], [8], did not include hypotheses, since they were modelling regressions only, without predictor variables.)

$$L1: Rating = \beta_{0j} + \beta_{1j}(race_{App}) + r_{ij} \quad [3]$$

$$H0_3L1: B1j < 0$$

$$Ha_3L1: B1j > 0$$

$$L1: Rating = \beta_{0j} + \beta_{1j}(race_{App}) + r_{ij} \quad [6]$$

$$H0_6L1: B1j < 0$$

$$Ha_6L1: B1j > 0$$

$$L2: \beta_{0j} = \gamma_{00} + \gamma_{01}(race_{Int}) + u_{0j} \quad [7]$$

H0₇L2: Y01 < 0

Ha₇L2: Y01 > 0

$$L1: Rating = \beta_{0j} + \beta_1(race_{App}) + r_{ij} \quad [9]$$

H0₉L1: B1 < 0

Ha₉L1: B1 > 0

$$L2: \beta_{0j} = \gamma_{00} + \gamma_{01}(race_{Int}) + u_{0j} \quad [10]$$

H0₁₀L2: Y01 < 0

Ha₁₀L2: Y01 > 0

$$L2: \beta_{1j} = \gamma_{10} + \gamma_{11}(race_{Int}) + u_{1j} \quad [11]$$

H0₁₁L2: Y11 < 0

Ha₁₁L2: Y11 > 0

The γ_{11} values obtained were used in the three aforementioned hypotheses to test for the moderating effect that interview structure has on the racial similarity effects:

Hypothesis 1: Interviewers using low-structured (unstructured) interviews will award higher ratings to same-race candidates compared to different-race candidates.

$$H0_1: \gamma_{11}^* \approx 0$$

$$Ha_1: \gamma_{11} > 0$$

Hypothesis 2: No race-similarity or dissimilarity effect will be found in the ratings by interviewers of same-race candidates compared to different-race candidates when using highly structured, situational interviews.

$$H0_2: \gamma_{31} > 0 \text{ or } \gamma_{31} < 0$$

$$Ha_2: \gamma_{31} \approx 0$$

$$Hb_2: \gamma_{31} < \gamma_{21} < \gamma_{11}$$

If Hb_2 is not rejected and the difference in the two γ values is significant, the study will be able to confirm the hypothesis that interview structure moderates the biasing impact of racial similarity effects in the selection interview.

3.10 Summary

This chapter conveyed the methods and techniques whereby racial similarity were operationalised and tested empirically. HLM was chosen as the primary method of analysis. The data of the two structured conditions were analysed by means of HLM and the results compared to observe the moderating effect that interview structure had on these effects. Multiple regression analysis was run as a secondary, supportive analysis to the HLM. The next chapter will present the results obtained from the analyses.

CHAPTER 4: RESULTS

4.1 Introduction

An experimental design was chosen to test the moderating effect that interview structure has on racial similarity effects in selection interviews. Highly structured interviews were hypothesised to be more resistant to racial similarity effects than the unstructured counterpart. Data were obtained in a laboratory study setting, using simulated interview video recordings with undergraduate students as raters. The data were analysed using HLM analysis as well as multiple regression analysis to provide comparative data to recent research in this niche area (De Meijer, 2008; McCarthy et al., 2010; Sacco et al., 2003).

This chapter presents the explanation of how the data was cleaned after capturing (in SPSS), how missing values were handled and assumptions for analysis tested. This is followed by descriptive statistics of the two samples, as well as the regression analysis and HLM analysis results. The chapter concludes with a comparative analysis between the results of the structured and unstructured samples, testing the critical hypothesis of the moderating role of structure, as well as a summative discussion of the results.

4.2 Data Cleaning

Both data sets were prepared for analyses by scanning and correcting for coding errors, missing values and outliers. This section will briefly explain the way the data was cleaned, as necessitated by the raw data sets, in both of the conditions.

No missing values or coding errors occurred in the low-structure data set, but a few outliers were identified (Table 4.1). The outliers mostly presented low extremes. This study recognises that extreme ratings are a reality indeed and a great concern in practice, since the size of a large diverse panel of interviewers even is not big enough in practice to protect an interviewee from the possible bias of one extreme rating. This study was especially concerned with the effect of such extreme bias ratings, therefore the outliers were not removed from the dataset. These considerations compelled the researcher to believe that leaving the outliers in the data set would provide a more accurate and realistic reflection of aggregated rater bias.

Table 4.1.

Missing value analysis for low-structured conditions

	N	Mean	Std. Deviation	Missing		No. of Extremes ^a	
				Count	Percent	Low	High
Rating	1560	60.87	14.111	0	.0	12	1

*a. Number of cases outside the range (Q1 - 1.5*IQR, Q3 + 1.5*IQR)*

Some missing values were detected in the high-structure data set (Table 4.2). These missing values amounted to less than 0.5% of the sample size in seven of the eight items, and were therefore deleted from the data set before the analyses were conducted in both SPSS and HLM. A significant amount of missing values amounting to 12.8% of the item scores were indicated in the eighth item (PEOP4). These missing values resulted from the omission of the PEOP4 question for one of the applicants in the video stimuli, with the effect that an aggregated score could not be calculated for the PEOPLE MANAGEMENT competency or the OVERALL SCORE and this left the analysis with only seven of the eight applicants. Eight applicants had been selected with the methodological intent that was central to the

objective of the study, namely to have equal representation of specific different groups – two black males, two white males, two black females and two white females. The loss of one of the applicants would therefore be less than ideal. In accordance with the proposed methodology, it was decided to impute the missing values by manually imputing the average value of the three other items measuring PEOPLE MANAGEMENT, namely, PEOP1, PEOP2 and PEOP3 at the PEOP4 field of candidate eight. The table below presents the situation before the missing values were imputed (Table 4.2). After the imputation, the 189 missing values at PEOP4 were reduced to 0.

Table 4.2.

Missing value analysis for the highly structured condition

	N	Mean	Std. Deviation	Missing		No. of Extremes ^a	
				Count	Percent	Low	High
COM1	1474	4.95	1.688	6	.4	63	0
COM2	1476	4.51	1.308	4	.3	91	98
COM3	1477	4.42	1.274	3	.2	114	66
COM4	1475	4.42	1.687	5	.3	0	0
PEOP1	1479	4.60	1.221	1	.1	54	95
PEOP2	1477	5.18	1.296	3	.2	8	0
PEOP3	1475	4.69	1.667	5	.3	84	0
PEOP4	1291	4.25	1.839	189	12.8	0	1

*a. Number of cases outside the range ($Q1 - 1.5*IQR$, $Q3 + 1.5*IQR$)*

4.3 Testing for Assumptions

In this section, the general assumptions for OLS regression and HLM required assumptions for each of the two data sets are explained and reported.

4.3.1 Normality, Linearity, Homoscedasticity

Normality refers to the degree to which data are normally distributed and can be tested for by skewness and kurtosis statistics, as well as the observation of a histogram of the distribution of the dependant variable (Tabachnick & Fidell, 2007). In the low-structure data set, skewness and kurtosis were tested for, using SPSS DESCRIPTIVES, and the data were found to be reasonably normal, though slightly skewed to the left (Skewness statistic: $-.448$, Kurtosis statistic: $.268$). Figure 4.1 depicts the skewness and kurtosis visually, as a histogram of the dependant variable, *Rating*. The degree of skewness and kurtosis were small enough to ignore since the robustness of regression analysis could ensure results to be accurate despite small deviations from normality (Kerlinger & Lee, 2000). In the high-structure data set, normality was tested for on the two competencies evaluated, as well as the overall score, since all three of these were used as dependent variables in the subsequent analyses. Similar, and even better distributional characteristics were observed in the high-structure data compared to the low-structured data (OVERALL SCORE: Skewness statistic: $-.478$, Kurtosis statistic: $.196$; COMMUNICATION: Skewness statistic: $-.373$ Kurtosis statistic: $-.407$; PEOPLE MANAGEMENT: Skewness statistic: $-.411$, Kurtosis statistic: $.355$) and therefore also did not require any transformation prior to the analyses. The normality of the data can be visually observed in Figures 4.2, 4.3 and 4.4

Figure 4.1.

Histogram depicting the normality of distribution for the 'Rating' in the low-structure dataset.

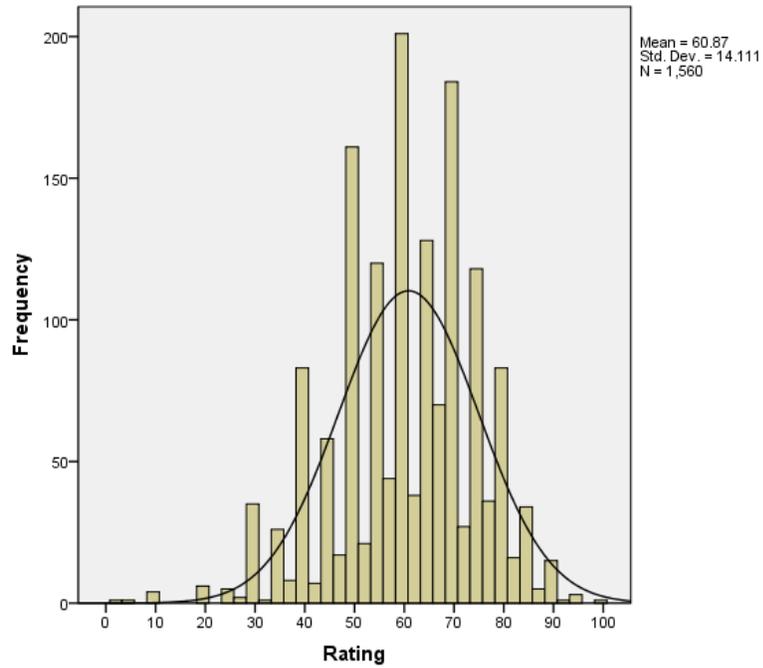


Figure 4.2.

Histogram depicting the normality of distribution for the COMMUNICATION competency in the high structure dataset.

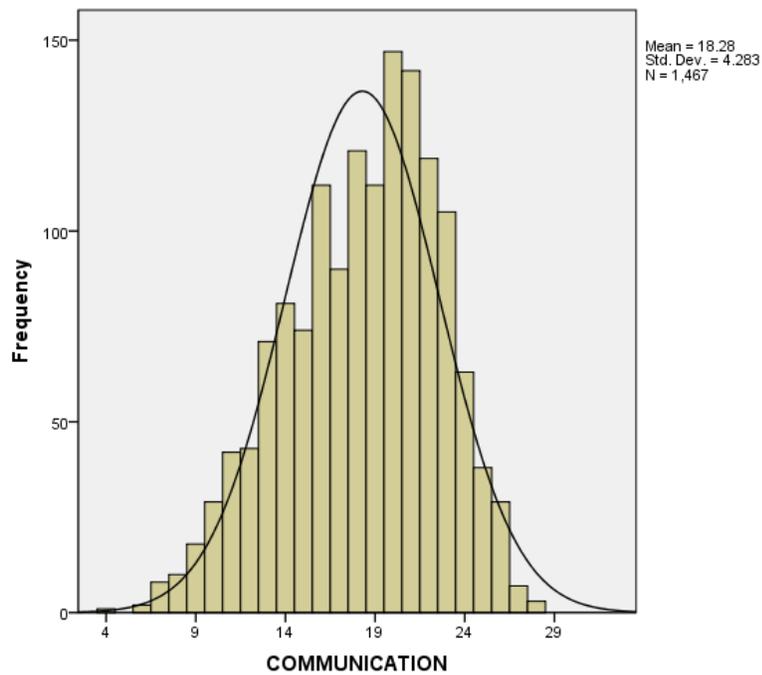


Figure 4.3.

Histogram depicting the normality of distribution for the PEOPLE MANAGEMENT competency in the high structure data set.

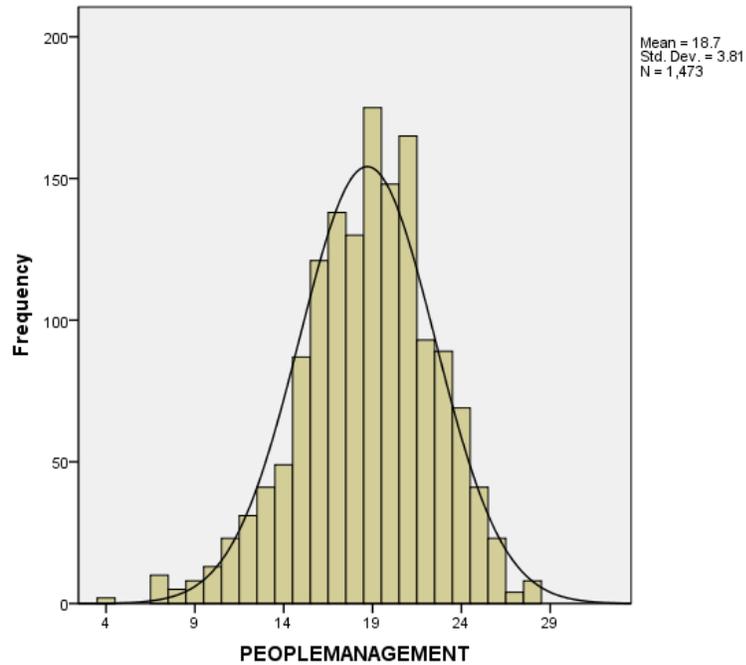
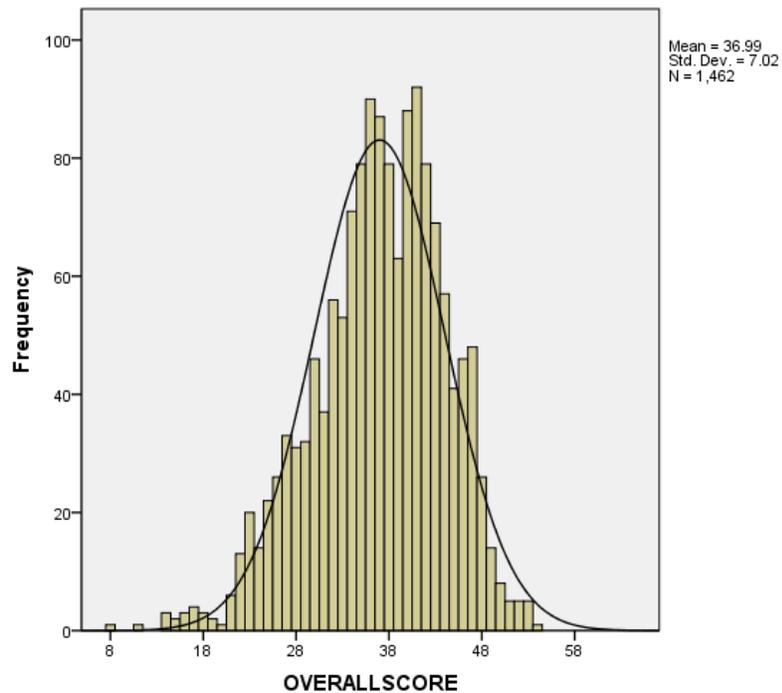


Figure 4.4.

Histogram depicting the normality of distribution for the OVERALLSCORE competency in the high structure data set.



Testing for linearity require regression-based analyses, plotting the dependent and independent variables against each other (Tabachnick & Fidell, 2007). In this study, the independent variable, race, was dichotomous in nature (0 = white; 1 = black) and therefore did not yield regression equations that could be used for the testing of linearity. The assumption of linearity was therefore assumed to be met, since it could not be determined.

Homoscedasticity refers to the assumption that the dependant variable shows similar variance when observed over different independent variables (Tabachnick & Fidell, 2007). In this study, there were only two independent variable options, i.e. white or black. The difference in the standard deviations of black applicant scores and white applicant scores in the low-structured condition indicated only a small difference in variance in the two conditions ($SD_{black} = 14.77$; $SD_{white} = 13.43$) of the independent variables. In the high-structure data set, the standard deviation of black applicant scores and white applicant scores also were small enough not to reject the assumption of homoscedasticity ($SD_{black} = 7.20$; $SD_{white} = 5.27$). This difference in variance was slightly bigger, but the robustness of multiple regression analysis could accommodate it (Kerlinger & Lee, 2000).

4.4 Descriptive Statistics and Inter-correlations of Study Variables

The descriptive statistics of both data sets are reported in this section. More detailed reports on item and construct level are also given for the high-structure data set. Interesting main effect comparisons that are discussed more thoroughly in Chapter 5 are possible from these tables.

4.4.1 Descriptive Statistics of the Low-structured Condition

Interesting observations can be made when comparing the means of ratings *given* by raters and *allocated to* applicants of the two observed race groups (Table 4.5). The averages of ratings by white raters and black raters, and scores allocated to white applicants and black applicants all fell in the range 59.35 to 61.36 (out of 100). Using a simplistic (but incorrect) interpretation of bias as mean differences in scores, this would indicate that the assessment was not biased against any group, since there was no systematic difference in the scores of different groups (Cleary, 1968; Theron, 2010). This, however, is not the conclusion that was drawn when the more detailed mean comparison, split by group membership, was considered. It is clear from Table 4.5 that each race group rated racially similar applicants higher than non-similar applicants, thereby advantaging the group to which they belonged – in each instance the similarity effects determined the group that was given the superior rating. The observation of the homogenous overall means only masked the very realness of the similarity effects, as the averaging brought inherently biased scores deceptively close to each other. The significance of these effects will be tested in later analyses.

Table 4.3.

Means, standard deviations, sample sizes, and interview ratings for each applicant-interviewer race in the low-structured condition

Applicant Race	Statistic	Interviewer Race		Overall
		Black	White	
Black	M	61.59	60.08	60.37
	SD	13.65	14.85	14.77
	Applicant N	4	4	8
	Interviewer N	152	629	781
White	M	57.10	62.40	61.36

SD	13.26	13.45	13.43
Applicant N	4	4	8
Interviewer N	152	627	779
Overall Mean	59.35	61.24	60.87

4.4.2 Descriptive Statistics of the Highly Structured Condition

The high-structure data portrayed quite the opposite to the low-structured condition discussed above. The same table was drawn up for the high-structure data (Table 4.6) and the means of the overall scores could similarly be compared. The means, though significantly different for the two groups, were consistently determined to be so by the raters of both groups. In other words, raters did not ascribe higher ratings to racially similar candidates, but both rater groups allocated the highest rating to the same race (whites) and the lower score to the other group (blacks).

A slight advantage however, was allocated to black applicants by black raters compared to the white raters' rating of black applicants ($M_{black\ raters} = 36.73$; $M_{white\ raters} = 33.61$). This effect was again reversed for white applicants, with black raters scoring the white applicants slightly higher than the white raters scored the white applicants ($M_{black\ raters} = 40.60$; $M_{white\ raters} = 39.94$). These results make the waters of clearly visible similarity effects very murky and it would seem that other random non-racial factors influenced group ratings in the highly structured condition. The statistical significance of these results is discussed after further analysis in SPSS and HLM, but it is clear that the results differed radically from the low-structured results. This was even more significant when considering that the same applicants and the same raters were used for both conditions. The moderating effect of interview

structure was clearly visible, but its true effectiveness had to be tested by subsequent analyses.

Table 4.4.

Means, standard deviations, sample sizes and interview ratings for each applicant-interviewer race in the highly structured condition

Applicant Race	Statistic	Interviewer Race		Overall
		Black	White	
Black	M	36.73	33.61	34.00
	SD	7.10	7.31	7.20
	Applicant N	4	4	8
	Interviewer N	92	644	736
White	M	40.60	39.94	40.02
	SD	5.47	5.08	5.27
	Applicant N	4	4	8
	Interviewer N	89	637	726
Overall Mean		38.67	36.78	36.99

4.4.3 Descriptive Statistics on Applicant Level

The following tables present the descriptive statistics for the ratings allocated to each of the eight applicants, first in the low-structured condition, against a total of 100, and then in the highly structured condition on both competencies and the aggregated overall score.

4.4.3.1 Low-structured Condition

The following tables provide basic statistical information on the rating of each of the eight applicants from the low-structure data set.

Table 4.5

Descriptive statistics on applicant level for applicant 1

APPL1			
(W*)	Mean	N	Std. Deviation
0	62.14	1365	13.851
1	51.94	195	12.633

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater

Table 4.6

Descriptive statistics on applicant level for applicant 2

APPL2			
(B*)	Mean	N	Std. Deviation
0	60.58	1365	14.338
1	62.90	195	12.244

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater

Table 4.7

Descriptive statistics on applicant level for applicant 3

APPL3			
(W*)	Mean	N	Std. Deviation
0	60.68	1365	14.419
1	62.19	195	11.674

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater.

Table 4.8

Descriptive statistics on applicant level for applicant 4

APPL4			
(B*)	Mean	N	Std. Deviation
0	62.24	1365	13.174
1	51.28	195	16.557

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater.

Table 4.9

Descriptive statistics on applicant level for applicant 5

APPL5			
(W*)	Mean	N	Std. Deviation
0	59.80	1365	14.037
1	68.36	195	12.251

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater

Table 4.10

Descriptive statistics on applicant level for applicant 6

APPL6			
(B*)	Mean	N	Std. Deviation
0	60.17	1365	14.148
1	65.76	195	12.858

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater

Table 4.11

Descriptive statistics on applicant level for applicant 7

APPL7			
(B*)	Mean	N	Std. Deviation
0	60.78	1365	14.341
1	61.45	195	12.398

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater

Table 4.12

Descriptive statistics on applicant level for applicant 8

APPL8			
(W*)	Mean	N	Std. Deviation
0	60.55	1365	14.350
1	63.06	195	12.118

*W = White applicant; B = Black applicant
0 = Black rater; 1 = White rater

In most cases observed above, the two rater groups scored the applicants differently, and quite severely so, with half of the applicant mean scores showing a difference in rater group rating of more than 5%. Standard deviations were reasonably constant regarding all applicants.

It is interesting to note that the opposite rater group allocated higher ratings to some of the individual applicants than the similar group (E.g. Applicants 1, 2, 6, 7). These ratings were, on average, 4.72% higher than ratings by the similar group, whereas the applicants that were rated higher by their own groups were rated 5.89% higher on average. These statistics prove inconclusive, but do highlight the fact that the two rater groups showed systematic differences in the way they rated the two applicant groups.

4.4.3.2 Highly Structured Condition

The following tables present basic statistical information on the rating of each of the eight applicants from the high-structure data set. It again can be observed that none of the applicants were allocated the same score consistently by both rater groups. A difference of between 4.42% (Applicant 3) and 12.48% (Applicant 4) could be seen in the mean ratings of individual applicants by different rater groups – indicating that the different groups consistently seemed to rate differently. This finding is the similar to that seen in the low-structured condition. However, the differences in scores are not always in favour of the racially similar applicant; five of the eight applicants quite contrarily were scored higher by the rater group of the opposite race (E.g. Applicant 2, 3, 4, 5 and 6).

Table 4.13

Descriptive statistics on applicant level for applicant 1

		OVERALL							PEOPLE				
APPL1 (W)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT	
0*	Mean	36.49	4.85	4.39	4.38	4.32	17.92	4.55	5.08	4.88	4.06	18.57	
	N	1279	1290	1291	1293	1290	1284	1294	1292	1292	1292	1288	
	Std. Deviation	7.144	1.721	1.293	1.299	1.714	4.321	1.240	1.312	1.507	1.585	3.874	
1*	Mean	40.47	5.67	5.37	4.66	5.12	20.81	4.93	5.86	3.34	5.51	19.64	
	N	183	184	185	184	185	183	185	185	185	185	185	
	Std. Deviation	4.826	1.220	1.076	1.049	1.288	2.962	1.022	.928	2.061	1.054	3.192	

*W = White applicant; B = Black applicant.. 0 = Black rater; 1= White rater

Table 4.14

Descriptive statistics on applicant level for applicant 2

		OVERALL							PEOPLE				
APPL2 (W)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT	
0*	Mean	37.70	5.05	4.46	4.59	4.53	18.61	4.70	5.25	4.83	4.29	19.08	
	N	1277	1289	1291	1292	1290	1282	1294	1292	1292	1292	1288	
	Std. Deviation	6.840	1.674	1.333	1.158	1.705	4.289	1.195	1.306	1.602	1.563	3.637	
1*	Mean	32.07	4.24	4.83	3.23	3.69	15.98	3.92	4.65	3.66	3.86	16.09	
	N	185	185	185	185	185	185	185	185	185	185	185	
	Std. Deviation	6.245	1.618	1.073	1.407	1.347	3.459	1.184	1.089	1.744	1.809	3.969	

*W = White applicant; B = Black applicant.. 0 = Black rater; 1= White rate.

Table 4.15

Descriptive statistics on applicant level for applicant 3

		OVERALL							PEOPLE				
APPL3 (B)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT	
0*	Mean	36.63	4.82	4.52	4.34	4.23	17.89	4.54	5.08	4.66	4.45	18.73	
	N	1277	1289	1291	1292	1290	1282	1294	1292	1292	1292	1288	
	Std. Deviation	7.252	1.714	1.369	1.297	1.650	4.332	1.237	1.290	1.713	1.485	3.926	
1*	Mean	39.46	5.90	4.43	4.92	5.74	20.99	5.01	5.84	4.88	2.75	18.48	
	N	185	185	185	185	185	185	185	185	185	185	185	
	Std. Deviation	4.420	1.101	.764	.955	1.318	2.658	1.019	1.138	1.276	1.593	2.874	

*W = White applicant; B = Black applicant. 0 = Black rater; 1 = White rater

Table 4.16

Descriptive statistics on applicant level for applicant 4

		OVERALL							PEOPLE				
APPL4 (W)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT	
0*	Mean	38.00	5.33	4.64	4.46	4.67	19.08	4.64	5.26	4.78	4.23	18.91	
	N	1277	1289	1291	1292	1290	1282	1294	1292	1292	1292	1288	
	Std. Deviation	6.425	1.380	1.237	1.273	1.586	3.780	1.203	1.226	1.687	1.656	3.626	
1*	Mean	30.01	2.32	3.60	4.11	2.72	12.75	4.31	4.61	4.03	4.30	17.25	
	N	185	185	185	185	185	185	185	185	185	185	185	
	Std. Deviation	6.997	1.234	1.430	1.242	1.365	3.388	1.306	1.602	1.341	1.159	4.667	

*W = White applicant; B = Black applicant. 0 = Black rater; 1 = White rater.

Table 4.17

Descriptive statistics on applicant level for applicant 5

		OVERALL						PEOPLE				
APPL5 (B)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT
0*	Mean	36.51	4.90	4.49	4.34	4.32	18.03	4.53	5.21	4.62	4.12	18.48
	N	1277	1289	1291	1292	1290	1282	1294	1292	1292	1292	1288
	Std. Deviation	7.014	1.747	1.279	1.239	1.705	4.311	1.218	1.329	1.709	1.614	3.832
1*	Mean	40.25	5.30	4.61	4.96	5.14	20.01	5.09	4.93	5.17	5.05	20.24
	N	185	185	185	185	185	185	185	185	185	185	185
	Std. Deviation	6.155	1.149	1.500	1.377	1.359	3.649	1.129	1.006	1.229	1.241	3.272

*W = White applicant; B = Black applicant, 0 = Black rater; 1= White rater

Table 4.18

Descriptive statistics on applicant level for applicant 6

		OVERALL						PEOPLE				
APPL6 (B)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT
0*	Mean	36.51	4.83	4.41	4.41	4.39	18.02	4.59	5.17	4.54	4.19	18.49
	N	1279	1290	1292	1293	1292	1284	1295	1293	1293	1293	1289
	Std. Deviation	7.018	1.709	1.299	1.311	1.746	4.330	1.215	1.311	1.673	1.622	3.804
1*	Mean	40.36	5.81	5.20	4.47	4.64	20.15	4.65	5.23	5.72	4.58	20.19
	N	183	184	184	184	183	183	184	184	184	184	184
	Std. Deviation	6.061	1.238	1.163	.975	1.176	3.401	1.263	1.190	1.180	1.408	3.516

*W = White applicant; B = Black applicant, 0 = Black rater; 1= White rater

Table 4.19

Descriptive statistics on applicant level for applicant 7

		OVERALL						PEOPLE				
APPL7 (B)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT
0*	Mean	37.42	5.00	4.59	4.45	4.63	18.65	4.64	5.26	4.61	4.26	18.77
	N	1279	1289	1292	1292	1290	1283	1294	1292	1293	1292	1289
	Std. Deviation	7.072	1.742	1.334	1.301	1.634	4.317	1.221	1.273	1.690	1.638	3.843
1*	Mean	33.95	4.62	3.93	4.17	2.98	15.71	4.35	4.60	5.20	4.10	18.24
	N	183	185	184	185	185	184	185	185	184	185	184
	Std. Deviation	5.811	1.202	.927	1.033	1.312	2.959	1.189	1.315	1.386	1.319	3.548

*W = White applicant; B = Black applicant, 0 = Black rater; 1 = White rater

Table 4.20

Descriptive statistics on applicant level for applicant 8

		OVERALL						PEOPLE				
APPL8 (W)*		SCORE	COM1	COM2	COM3	COM4	COMMUNICATION	PEOP1	PEOP2	PEOP3	PEOP4	MANAGEMENT
0*	Mean	36.64	4.84	4.57	4.36	4.29	18.05	4.61	5.10	4.57	4.31	18.59
	N	1289	1293	1293	1293	1293	1290	1294	1294	1293	1294	1293
	Std. Deviation	7.172	1.725	1.300	1.286	1.696	4.392	1.228	1.303	1.703	1.614	3.888
1*	Mean	39.54	5.78	4.09	4.80	5.36	19.97	4.56	5.68	5.51	3.77	19.51
	N	173	181	183	184	182	177	185	183	184	183	180
	Std. Deviation	5.102	1.083	1.295	1.114	1.275	2.867	1.174	1.123	1.071	1.434	3.092

*W = White applicant; B = Black applicant, 0 = Black rater; 1 = White rater

4.4.4 Descriptive Statistics on Item Level

Table 4.21 presents the descriptive statistics of the items, competencies and the overall score of the high-structure data set. The low-structured condition did not, according to its nature, make use of item ratings, but only of overall scores, therefore only the data from the highly structured condition will be reported through item-level statistics. The means and standard deviations were reasonably constant throughout the items measures, while the normality statistics indicated that the distributions were fairly normal, though all slightly skewed to the left, as with the low-structured condition.

Table 4.21
Descriptive statistics on item level for the high-structure data

	N	Minimum	Maximum	Mean	Std.	Skewness	Kurtosis		
					Deviation			Std. Error	Std. Error
	Statistic	Statistic							
OVERALL SCORE	1462	8	54	36.99	7.020	-.478	.064	.196	.128
COM1	1474	1	7	4.95	1.688	-.669	.064	-.429	.127
COM2	1476	1	7	4.51	1.308	-.137	.064	-.229	.127
COM3	1477	1	7	4.42	1.274	-.261	.064	.073	.127
COM4	1475	1	7	4.42	1.687	-.275	.064	-.740	.127
COMMUNICATION	1467	4	28	18.28	4.283	-.373	.064	-.407	.128
PEOP1	1479	1	7	4.60	1.221	-.025	.064	-.181	.127
PEOP2	1477	1	7	5.18	1.296	-.534	.064	-.043	.127
PEOP3	1477	1	7	4.69	1.666	-.546	.064	-.494	.127
PEOP4	1477	1	7	4.24	1.602	-.338	.064	-.592	.127
PEOPLEMANAGEMENT	1473	4	28	18.70	3.810	-.411	.064	.345	.127

4.5 Results of HLM and Multiple Regression Analyses

The analyses were undertaken with the student version of HLM7 (Raudenbush et al., 2011) as well as linear regression analyses in SPSS 19 and are reported on in this section by referring to the same steps outlined in Chapter 3 (3.8). The results of the low-structured and the highly structured conditions are reported separately and compared in a subsequent section. The hypotheses stated in Chapter 3 were tested and are reported on in this section.

4.5.1 Low-structured Condition Results

This section reports the results of the low-structured condition in respect of which the hypotheses that were tested stated that the presence of significant racial similarity effects would be observed under conditions where low-structured selection interviews were used

4.5.1.1 HLM Analysis Results

The analysis was done by, (1) running each set of regression equations as described in 3.8; (2) checking the results and testing hypotheses; and (3) continuing with subsequent steps as necessitated by previously accepted or rejected hypotheses.

The first set of regression equations, called the *null model* was tested and the ICC was calculated by using the formula $\rho = \tau_{00} / (\tau_{00} + \sigma^2)$. The data gave the result as $= 45.33759 / (45.33759 + 154.29890) = .23$. This ICC value explained that 23% of the total variance in the score could be explained by inter-interviewer variance and indicated significant idiosyncratic rater effects. This also proved that a significant nesting effect was present and that analysis

with HLM would provide rich information in the further steps of analysis (Sacco et al., 2003). The model also showed significant intercept variability ($p < .05$). For these reasons, Hypotheses 1 and 2 were not rejected and the following models were tested.

The next model, the *random coefficient regression model* was tested. This model tests whether applicant race is related to higher ratings (Sacco et al., 2003). This hypothesis was rejected ($p > .05$), which indicated that all applicants, irrespective of race, were scored similarly ($\gamma_{10} = -.98, p = .12$). There was no race main effect for the applicant race independent variable in the low-structured condition, as the descriptive results in Table 4.4 also indicated.

Table 4.22
Final estimation of fixed effects for results in HLM: Random coefficient regression model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	61.364706	0.654447	93.766	194	<0.001
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-0.979109	0.629043	-1.557	1364	0.120

The next model that was tested was the *intercepts as outcomes model* that determines whether interviewer race has significant main effects on ratings. This hypothesis was also rejected ($p > .05$) and it was concluded that the means of different interviewer race groups were similar. ($\gamma_{01} = -1.33, p = .349$). This also supported the descriptive results that indicated no main effect for rater race.

Table 4.23

Final estimation of fixed effects for results in HLM: Intercepts as outcomes model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	61.638286	0.716643	86.010	193	<0.001
RATERRAC, γ_{01}	-1.333003	1.421125	-0.938	193	0.349
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-0.979316	0.629042	-1.557	1364	0.120

The last model, the *slopes as outcomes model*, was tested to see whether significant race interaction effects had been found, even though no main effects were found. Very interestingly, and also derivable from the descriptive summary table, interaction effects were found, and significantly so, at $p < .05$ and even at $p < .001$, ($\gamma_{11} = 6.68$, $p < 0.001$). The γ_{11} value was the critical one, modelling the interaction effect, as explained in Chapter 3. This confirmed hypothesis one that hypothesised that significant racial similarity effects would be found in the low structure condition.

Table 4.24

Final estimation of fixed effects for results in HLM: Slopes as outcomes model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	62.324820	0.733186	85.005	193	<0.001
RATERRAC, γ_{01}	-4.674820	1.618228	-2.889	193	0.004
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-2.350070	0.701069	-3.352	1363	<0.001
RATERRAC, γ_{11}	6.681320	1.547743	4.317	1363	<0.001

4.5.1.2 Regression Analysis

As an alternative analysis, interaction effects were modelled by means of multiple regression analysis in SPSS. For the low-structure data, the variables *Applicant Race*, *Rater Race* and *Race x Race* (rater race x applicant race) were modelled as predictors with *Rating* (the overall rating of each applicant, by each rater) as the dependent variable. Correlation coefficients between the variables could not be obtained since the independent variables were dichotomous (0 or 1).

The model showed a deceptively small R-Square value ($R^2 = .013$). This implied that the independent variables did not explain much variance (1.3%) of the independent variable. This was to be expected as the model was not intended to be explanatory, but rater focused on one specific source of error variance – rater bias. When the fact that no significant main effects were found is taken into account, it makes sense that the R^2 value was so low. In this light, the value of .013 was not to be frowned upon, since it implied that applicant race, rater race and the interaction thereof undesirably explained 1.3% of the variance in the dependent variable, *Rating*.

The ANOVA model (Table 4.25) tests the null hypothesis stating that all the model coefficients are zero (Tabachnick & Fidell, 2007). The hypothesis was rejected. ($F(6.943)$, $p < 0.001$) This F-value was calculated by dividing the Mean Square (regression) by the Mean Square (Residual).

Table 4.26 indicates the coefficients, standard errors, standardised coefficients as well as the statistical significance of the models. The B values in the model show that black applicants,

on average, scored 2.3 points out of 100 lower than white candidates and also that black raters rated all applicants 5.3 points lower out of 100 than white raters ($B_{appl} = 2.322, p = .003; B_{rater} = 5.300, p = .000$). Both these main effects were significant at $p < .05$. The *Race x Race* interaction term indicates that racial similarity, i.e., the effect that similar-to-me ratings versus different-to-me ratings had in the low-structured condition were 6.81 out of 100, indicating significant racial similarity bias even at $p < .001 (B_{interaction} = 6.81, p = .000)$. This finding also confirmed the hypothesis that racial similarity effects would be found in a low-structured interview context.

Table 4.25
Model summary of the multiple regression analysis of the low-structured condition

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.115	.013	.011	14.031

Predictors: (Constant), Race x Race, Applicant Race, Rater Race

Dependent Variable: Rating

Table 4.26
ANOVA model of the multiple regression analysis of the low-structured condition

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	4100.413	3	1366.804	6.943	.000
	Residual	306317.237	1556	196.862		
	Total	310417.650	1559			

Predictors: (Constant), Race x Race, Applicant Race, Rater Race. Dependent Variable: Rating

Table 4.27
Coefficients model of the multiple regression analysis of the low-structured condition

Model		Unstandardised Coefficients		Standardised Coefficients		Sig.
		B	Std. Error	Beta	t	
1	(Constant)	62.399	.560		111.360	.000
	Applicant Race	-2.322	.792	-.082	-2.932	.003
	Rater Race	-5.300	1.269	-.149	-4.178	.000
	Race x Race	6.808	1.794	.143	3.796	.000

Model		Unstandardised Coefficients		Standardised	t	Sig.
		B	Std. Error	Coefficients		
1	(Constant)	62.399	.560		111.360	.000
	Applicant Race	-2.322	.792	-.082	-2.932	.003
	Rater Race	-5.300	1.269	-.149	-4.178	.000
	Race x Race	6.808	1.794	.143	3.796	.000

Dependent Variable: Rating

4.5.2 Highly Structured Condition Results

This section reports the results of the highly structured condition under which the hypotheses that assumed the absence of significant racial similarity effects in conditions where highly structured interviews are used were tested.

4.5.2.1 HLM Analysis

The analysis was done by; (1) running each set of regression equations as described in Chapter 3 (3.8); (2) checking the results and testing hypotheses; and (3) continuing with subsequent steps as necessitated by previously accepted or rejected hypotheses.

The first set of regression equations, called the *null model* was tested and the ICC was calculated by using the formula $\rho = \tau_{00} / (\tau_{00} + \sigma^2)$. The data gave the result to be $= 9.06286 / (40.31287 + 9.06286) = .18$. This ICC value explained that 18% of the total variance could be explained by inter-interviewer variance and indicated significant idiosyncratic rater effects. This also proves that a significant nesting effect was present and that analysis with HLM would provide rich information in the further steps of analysis (Sacco et al., 2003). The

model also showed significant intercept variability ($p < .05$), therefore the subsequent models were tested.

The next model, the *random coefficient regression model* was tested. This model tested whether applicant race was related to higher ratings. The hypothesis was confirmed at ($p < .05$), which indicated that applicants of different races were allotted significantly different scores in the interview ($\gamma_{10} = -6.022$, $p < 0.001$). This confirms with results observed in the descriptive statistics, where it was observed that white applicants were allocated significantly higher scores by raters of both groups.

Table 4.28

Final estimation of fixed effects for results in HLM: Random coefficient regression model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	40.017884	0.311969	128.275	184	<0.001
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-6.022226	0.286152	-21.046	1276	<0.001

The next model that was tested was the *intercepts as outcomes model* which determined whether interviewer race had significant main effects on ratings. This hypothesis was also confirmed at ($p < .05$), though not as convincingly as the previous hypothesis. It was concluded that the mean ratings given by the two interviewer race groups were significantly different ($\gamma_{01} = 1.856$, $p = .024$).

Table 4.29

Final estimation of fixed effects for results in HLM: Intercepts as outcomes model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	39.778037	0.326653	121.774	183	<0.001
RATERRAC, γ_{01}	1.855560	0.814495	2.278	183	0.024
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-6.023502	0.286153	-21.050	1276	<0.001

The last model, the *slopes as outcomes model* was tested to see whether racial interaction effects were found along with the significant main effects that were found. Significant similarity effects were found in the highly structured condition at ($p < .05$) and the null hypothesis that stated the γ_{11} variable to be insignificant ($\gamma_{11} = 2.737$, $p = .001$) was rejected.

Table 4.30

Final estimation of fixed effects for results in HLM: Slopes as outcomes model

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	39.955862	0.331118	120.670	183	<0.001
RATERRAC, γ_{01}	0.465351	0.922013	0.505	183	0.614
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-6.377038	0.305515	-20.873	1275	<0.001
RATERRAC, γ_{11}	2.737075	0.850072	3.220	1275	0.001

4.5.2.1.1 Additional HLM Analyses

Analyses were run to test the results, 1) without the missing value imputation, and also 2) on competency level, to see whether there were differences or points of interests in the way similarity effects were observed therein.

The HLM analysis was run without the imputation of the missing values at the PEO4 item. Only the last step of the HLM analysis was run to compare only the critical γ_{11} value that tests for a significant interaction effect. The results, as shown in the Table 4.31, indicate no significant racial similarity effect at $p < .05$ when the analysis is done without the imputation of the missing values ($\gamma_{11} = 1.599$, $p = .072$). The decision to impute the missing values, therefore, had a significant effect on the results of these analyses and introduced some uncertainty into the interpretation of the high-structure results. From the previous research discussed in Chapter 2 (McCarthy et al., 2010; Sacco et al., 2003) it is well documented that highly structured interviews are resistant to demographic similarity effects. Though the results of this study seem to be slightly vague at this point it could be considered that the imputation, which seemed the right thing to do logically, had an unrealistic effect on the data. This, however, could not be confirmed in this study.

Table 4.31

Final estimation of fixed effects for results in HLM (Overall rating – no MVI)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	39.960007	0.317121	126.009	183	<0.001
RATERRAC, γ_{01}	0.466843	0.883094	0.529	183	0.598

For APPLRACE slope, β_1

INTRCPT2, γ_{10}	-4.805111	0.317185	-15.149	1088	<0.001
RATERRAC, γ_{11}	1.598938	0.887875	1.801	1088	0.072

The HLM analyses were run for the aggregate score on the *communication* competency. Only the last step of the HLM analysis was run to compare only the critical γ_{11} value that tests for a significant interaction effect. The *communication* competency showed a significant effect at $p < .05$ ($\gamma_{11} = 1.152$, $p = .027$).

Table 4.32

Final estimation of fixed effects for results in HLM (Communication)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	20.427752	0.180040	113.462	183	<0.001
RATERRAC, γ_{01}	0.317175	0.501710	0.632	183	0.528
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-4.500694	0.187462	-24.009	1275	<0.001
RATERRAC, γ_{11}	1.151601	0.521574	2.208	1275	0.027

The HLM analyses were run for the aggregate score on the *people management* competency. It is within this competency that the missing values were imputed. Only the last step of the HLM analysis was run to compare only the critical γ_{11} value – which tests for a significant interaction effect. The *people management* competency also showed a significant effect at $p < .05$ ($\gamma_{11} = 1.583$, $p = .002$).

Table 4.33

Final estimation of fixed effects for results in HLM (People Management)

Fixed Effect	Coefficient	Standard error	t-ratio	Approx. d.f.	p-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	19.528219	0.186041	104.968	183	<0.001
RATERRAC, γ_{01}	0.150344	0.518273	0.290	183	0.772
For APPLRACE slope, β_1					
INTRCPT2, γ_{10}	-1.876325	0.185022	-10.141	1275	<0.001
RATERRAC, γ_{11}	1.583180	0.514796	3.075	1275	0.002

4.5.2 Regression Analysis

As an alternative analysis, interaction effects were modelled by multiple regression analyses in SPSS. In the high-structure data, the variables *Applicant Race*, *Rater Race* and *Race x Race* (rater race x applicant race) were modelled as predictors with *OVERALL SCORE* (the overall rating of each applicant, by each rater) as the dependent variable. Correlation coefficients between the variables could not be obtained since the independent variables were dichotomous (0 or 1).

The model showed a moderate R-Square value ($R^2 = .195$) that implied that the independent variables explained 19.5% of variance of the independent variable. This is not high in conventional terms, but was to be expected since the model was not intended to be explanatory, but rater focused on one specific source of error variance – rater bias. Most of this variance, though, was seen in the applicant race variable, where very big differences in the scores of two groups were observed.

The ANOVA model (Table 4.34) tested the null hypothesis that stated that all the model coefficients were zero (Tabachnick & Fidell, 2007). The hypothesis was rejected ($F(118.00)$, $p < 0.001$). This F-value was calculated by dividing the Mean Square (regression) by the Mean Square (Residual).

Table 4.35 indicates the coefficients, standard errors, standardised coefficients and the statistical significance of the models. The B-values in the model show that black applicants, on average, scored 6.33 out of 64 lower than white candidates ($B = -6.33$, $p = 0.000$). The model shows that all raters, irrespective of race, scored applicants similarly – black raters rated all applicants only 0.65 out of 64 lower than white raters ($B = .658$, $p = 0.356$). These two main effects are interesting in themselves as they confirm the descriptive statistics finding that, although applicants of different races were scored differently, it was not a direct function of interviewer race, since all interviewers consistently allocated similar scores to all applicants.

The *Race x Race* interaction term indicates that racial similarity, i.e., the effect that similar-to-me ratings versus different-to-me ratings had in the low-structured condition were 2.46 out of 64, indicating significant racial similarity effects ($B = 2.463$, $p = 0.014$). This finding contradicted the hypothesis that racial similarity effects would not be found in a highly structured interview context. In the light of the previously mentioned argument, this finding seemed somewhat out of place and could have been due to an aggregation of the two previous main effects without taking the nesting effect into account as the HLM analysis did.

Table 4.34
Coefficients model of the multiple regression analysis of the low-structured condition

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.442	.195	.194	6.304

Predictors: (Constant), Race x Race, ApplRace, RaterRace

Dependent Variable: OVERALL SCORE

Table 4.35
ANOVA model of the multiple regression analysis of the highly structured condition

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	14066.013	3	4688.671	118.000	.000
	Residual	57932.740	1458	39.734		
	Total	71998.753	1461			

Predictors: (Constant), Race x Race, ApplRace, RaterRace

Dependent Variable: OVERALL SCORE

Table 4.36
Coefficients model of the multiple regression analysis of the highly structured condition

Model		Unstandardised Coefficients		Standardised Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	39.937	.250		159.906	.000
	ApplRace	-6.330	.352	-.451	-17.971	.000
	RaterRace	.658	.713	.031	.923	.356
	Race x Race	2.463	1.001	.085	2.460	.014

Dependent Variable: OVERALL SCORE

4.6 Comparative Analysis

Since this was an experimental study that compared two conditions hypothesised to be significantly different, the purpose in this section is to highlight the differences and similarities of the two conditions. It is important to note, again, that the same applicants and raters were used under both conditions, thereby to provide control of variance and a clear view of the effect of interview structure as the only factor that was not held constant in the two conditions.

The descriptive statistics yielded interesting differences in the two conditions that indicated the standardising and convergent effect that interview structure had on interviewer ratings. From the descriptive statistics, it was derived that significant interaction effects were present in the low-structured condition – indicating racial similarity effects, although there seemed to be no major main effects. Each rater group allocated their highest rating to their matching applicant group, black-for-black 61.59 and white-for-white 62.40, whereas black-for-white gave 57.10 and white-for-black 60.08. In the high-structure setting, much more uniform ratings were given. Black applicants (34.00) consistently scored lower than white applicants (40.02). Interestingly, both rater groups rated the black applicant group lower than the white group, unlike the inverse rating effect found in the low-structured condition. This rater-converging effect could be explained by the rating guide that formed part of the highly structured interview, guiding the raters to rate by the answers applicants gave and not their subjective perception of applicant suitability.

The HLM analyses also revealed distinctly different results for the two conditions. The low-structured condition showed no significant main effects for either applicant or rater race, but interaction terms showed very significant effects with $p < .001$. This was in line with the

interpretation drawn from the descriptive statistics mentioned above. The highly structured condition revealed significant applicant main effects and less significant rater main effects while there was slight unease in reporting the interaction effect with certainty: With the missing values imputed, as explained in section 4.2, significant similarity effects were present at $p < .05$, but with the missing value-cases deleted, no significant similarity effects were found at $p < .05$.

The regression analyses gave slightly more inconclusive results, but the conclusion overall is that, although both conditions presented significant interaction effects, the difference in the size of these effects were significant as an indication of the moderating effect of interview structure. Significant rater and applicant main effects as well as a significant interaction effect at $p < .001$ were found in the low-structured condition. In the highly structured condition, only applicant main effects were found, which was similar to the indication given by the descriptive statistics in this regard. The rater main effect was not significant, but the interaction effect was proven significant at $p < .05$ with $p = .014$.

Overall, the hypotheses of this study were partially confirmed, proving one half of the coin to be true and the other to be more complicated than the hypothesis expected it to be. The hypothesis stating that significant racial similarity effects were to be found within a low-structured interview environment was powerfully confirmed. The hypothesis that stated that significant racial similarity effects would not be found in the highly structured condition was rejected. Although differences in the sizes of these effects were observed, both conditions yielded significant racial similarity effect sizes at $p < .05$.

When the descriptive statistics were compared with the results of the HLM and Regression analyses it was clear that the HLM provided a much more accurate picture of the reality that the data portrayed. This most likely was due to the fact that HLM takes into account the nesting effect that is present in the data, whereas normal regression analysis ignores such effects.

4.7 Summary

It was the aim of this chapter to give a clear account of the results of the data analysis of this study and to present the findings on the hypotheses. It can generally be concluded that highly significant similarity effects were present in the low-structured condition, while somewhat debatable results that tended to be significant in most analyses, were also found with the use of highly structured interviews. The difference in the results of the low- and highly structured conditions were noteworthy, however, and therefore indicated a moderating effect of interview structure on racial similarity effects. In the next chapter, further elaboration on the implications of the results is presented and recommendations for future research are made.

CHAPTER 5: DISCUSSION OF RESULTS AND RECOMMENDATIONS FOR FUTURE RESEARCH

5.1 Introduction

This chapter presents a discussion of the general conclusions that were reached after the results of the data analysis had been interpreted. Furthermore, (1) the results are placed within the context of the previous research and theoretical frameworks; (2) the gist of the findings are summarised in an abbreviated and coherent way; (3) the implications and conclusions drawn from the results are discussed; (4) a discussion of the limitations of the study is presented; (5) recommendations are made with regard to future research and (6) practical applications of the results are discussed.

5.2 Background

In the light of previous studies on racial similarity in selection interviews, the main concern of this study was to determine the moderating effect that interview structure has on possible racial similarity effects in selection interviews within a South African context. Before proceeding with this aim, an attempt was made to actually find racial similarity effects within interviews, since previous research into this has been inconclusive. Researchers like De Meijer et al. (2007) Goldberg (2005), McCarthy et al. (2010) and Sacco et al. (2003) all hypothesised their expectation of finding racial similarity effects in employment interview settings. While using highly structured interviews and highly trained, professional interviewers, none of these major studies could confirm the existence of racial similarity effects and therefore conclude that the use of the structured interview format was resistant to the influence of similarity effects – if they do in fact exist. The question therefore would

concern whether the phenomenon of racial similarity is a concern at all with the employment interview, since there seemed to be no evidence of such effects in research. The critique of such a statement would point out that there is no evidence that similarity effects would exist even in low-structured conditions.

The hypotheses of this study boldly endorsed the idea that racial similarity would indeed be found in low-structured conditions, while highly structured conditions were hypothesised to be resistant to such effects, in confirmation of the expectations and predictions of other researchers. The South African context within which this study was undertaken itself has made a unique contribution as this country, more than most unfortunately has a rich history and heritage of racial tension and discrimination. Another significant difference from parallel studies that have been conducted concerned the use of student raters instead of professional raters. To accomplish the objectives set for this study, an experimental design was chosen, and two samples were taken – a high and a low-interview structure sample from within a controlled and stable environment. The results obtained from this were analysed via SPSS and HLM 7 and were found to confirm most of the central hypotheses.

In contrast to previous research, this study found significant racial similarity effects under both structured and unstructured interview conditions. Various explanations for the discrepancy between these findings and other mainstream research can be offered with the unique South African context and lack of rater accountability and professionalism, being the most likely causes. However, there is a difference in the size of the similarity effects of the two conditions, which indicates that interview structure does play a moderating role. The context of the research seemed to have magnified the prevalence of similarity effects to the extent where structure could not completely contain it - as was the case in previous research.

This study proves the very real danger making biased decisions due to racial similarity effects in employment interviews and should cause the practitioner, especially in South Africa, to rethink and align interview practices to best practice guidelines. It also provides strong support for the underlying psychological theories that supported the hypotheses of this and previous studies advocating the existence of similarity effects in judgement contexts - like the employment interview.

5.3 Summary of Results

This section presents a brief summary and discussion of the results from the two types of structured conditions investigated before the more detailed discussion that follows in the next section.

5.3.1 Low-structured Condition

In the low-structured condition, hypotheses were confirmed and significant racial similarity effects were found. Rater effects were shown to predict 23% of the total variance of the ratings, providing reason for further analysis of the source of these effects. In HLM, the γ_{11} variable indicating interaction effects that provide evidence of similarity effects was shown to be significant even at $p < .001$, even though the power of the sample was low due to unequal distribution of white and black raters. Similar results were obtained by the regression analysis modelling the interaction as an applicant race x rater race regression on the dependent variable, *Rating*. The coefficient of this statistic was also shown to be statistically significant at $p < .001$.

The table that follows (Table 5.1) serves as a summary of the HLM results reported above and shows that significant racial similarity effects were found in the low-structured condition, as hypothesised. HLM was used as the primary method of analysis and is therefore emphasised more than the regression analyses. HLM as method of analysis is promoted as the most accurate way to model demographic similarity effects in interviews by researchers like de Meijer et al. (2007), McCarthy et al. (2010) and Sacco et al. (2003). The emphasis on HML also serves the purpose of facilitating comparison of this study with other studies conducted in this field with HLM as method of analysis.

Table 5.1

Sample sizes and summary of key HLM results for racial similarity in the low-structured condition

Coding	Groups	Sample Sizes			Level 1 Variance Components		Level 2 Parameter Estimates	
		Applicant N	Interviewer N	ICC	Intercept	Slope	Intercept (γ_{01}) (SE)	Slope (γ_{11}) (SE)
1	Blacks	779	38	.23*	.66*	.63	1.62*	1.55*
0	Whites	781	157					

Notes: L1 = Level 1 (applicant level); L2 = Level 2 (interviewer level); ICC = Intra-class Correlation; SE = Standard error; HLM = Hierarchical Linear Modelling. Race was coded as either zero or one. $x^* = (p < .05)$. Asterisks on the ICCs indicate that the intercepts had significant variability, rather than statistical significance tests computed on the ICCs themselves.

Interestingly, applicant and interviewer main effects were not found in the low-structured condition. This is noteworthy because the common, but incorrect, interpretation of bias that concludes that, if different groups obtain and give the same scores, on average, the assessment is unbiased (Theron, 2010). In this case it is clear that, even though white and black applicants' average scores were not significantly different, the race of the rater still moderated the scores allotted to applicants and were directly biased against members of the other race group. The first of the three broad hypotheses presented in Chapter 2 is therefore confirmed, in that significant racial similarity effects were found in the low-structured interview condition.

This analysis concurs with the theory as stated in Chapter 2 as well as the Sacco et al. (2003) research that indicated that racial similarity effects might be more prevalent in low-structured conditions, although some studies have even failed to find significant effects with the use of unstructured interviews (Elliot, 1981). Researchers like Sacco et al. (2003) and McCarthy et al. (2010) have indicated the possibility that unstructured interviews could be more susceptible to similarity effects, but this notion has not been put to test to date. The results of this study provide some clarity concerning very real danger of using unstructured interviews as the preferred type of interview.

5.3.2 Highly Structured Condition

In the hypotheses concerning the highly structured condition were not confirmed because significant racial similarity effects were found. Rater effects were shown to predict 18% of the total variance of the ratings, providing reason for further analysis of the source of these effects. In HLM, the γ_{11} variable indicating interaction effects that provide evidence of

similarity effects was shown to be significant at $p < .05$, even though the power of the sample was low due to the unequal distribution of white and black interviewers. Similar results were obtained in regression analysis, modelling the interaction as a race x race regression. The coefficient of this statistic was also shown to be statistically significant at $p < .05$.

The following table (Table 5.2) serves as a summary of the HLM results reported above and shows, contrary to the hypothesis, that significant racial similarity effects were found in the highly structured condition,. HLM was used as the primary method of analysis and was therefore emphasised more than the regression analyses. HLM as method of analysis is promoted as the most accurate way to model demographic similarity effects in interviews by researchers like De Meijer et al. (2007), McCarthy et al. (2010) and Sacco et al. (2003). The emphasis on HLM also serves the purpose of facilitating comparison of this study with others in this field that have used HLM as method of analysis. The tabled summary portrays the results that were obtained with the imputed missing values data set, since the researcher believes it serves the cause of the research better than an unbalanced applicant grouping would. It is important, however, to once again note that the results showed no significant similarity effects when the analysis was run without the missing value imputation,– therefore confirming the hypothesis. The researcher still maintains that the imputation provides a more realistic outlook for analyses and therefore has chosen to use the imputed data for reporting results.

Table 5.2
Sample sizes and summary of key HLM results for racial similarity in the highly structured condition

Coding	Groups	Sample Sizes			Level 1 Variance Components		Level 2 Parameter Estimates	
		Applicant N	Interviewer N	ICC	Intercept	Slope	Intercept (γ_{01}) (SE)	Slope (γ_{11}) (SE)
1	Blacks	734	24	.18*	.29*	.81*	0.92	0.85*
0	Whites	727	161					

Notes: L1 = Level 1 (applicant level); L2 = Level 2 (interviewer level); ICC = Intra-class Correlation; SE = Standard error; HLM = Hierarchical Linear Modelling. Race was coded as either zero or one. $x^* = (p < .05)$. Asterisks on the ICCs indicate that the intercepts had significant variability, rather than statistical significance tests computed on the ICCs themselves.

Together with the significant interaction effects, contrary to the low-structured condition, applicant main effects were found in the highly structured condition. This is noteworthy since the widely held, but incorrect, definition of bias concludes that, if different groups obtain the same scores on average, the assessment is unbiased (Theron, 2010). According to this definition, this specific assessment might be considered biased since one group consistently scored higher than the other group. More research is needed to determine whether these differences are *trait related* or not, in other words, biased in the psychometric sense. However, the aim of this study was to test for rater bias in assessment. When the ratings of different rater groups are compared, as they were allocated to applicants of a different race, there is not much difference in the way different rater groups rated the different applicant groups. Still, the HLM analysis confirms the presence of significant interaction effects at the γ_{11} variable ($p = .14$), indicating that, even though the effects are not as clearly evident as is the case in the low-structured condition, they do exist and that rater race, in this study, moderated the applicants' scores in the highly structured environment.

This analysis contradicts the McCarthy et al. (2010) research that declares that structured interviews are highly resistant to demographic similarity effects. The diverse and troubled social background of South Africa might be to blame for this finding that contradicts mainstream research findings, along with constructs like professionalism and accountability. The fact remains that similarity effects were greater in the low-structured condition and that interview structure did indeed play a positively moderating effect in the struggle against biased interview ratings.

5.3.3 Main Hypotheses Testing

The following hypotheses were formulated in Chapter 2 and conclusions are presented briefly on the basis of the reported results:

Hypothesis 1

Interviewers using low structured (unstructured) interviews will award higher ratings to same-race candidates compared to different-race candidates.

Hypothesis 1 is confirmed by the results of this study, since it is clear from all the results that, in the low-structured condition, white raters assigned higher ratings to white candidates than to blacks and black raters assigned higher ratings to blacks than to whites.

Hypothesis 2

No race-similarity or dissimilarity effect will be found in the ratings by interviewers of same-race candidates compared to different-race candidates when using highly structured, situational interviews.

Hypothesis 2 is rejected since, although less significant than in the low-structured condition, statistically significant racial similarity effects were still found in the highly structured condition.

5.4 Discussion of Results

This section presents a discussion of the results of the two conditions in a comparative way by highlighting interesting observations from the main effect results to the more imperative interaction- or similarity effect results.

From the descriptive statistics presented in section 4.4, noteworthy observations can be made in terms of the main effects present in the two structured conditions. The fact that the low-structured condition shows very little observable main effects may at first glance look deceptively positive, fulfilling the early (but incorrect) definition of fairness in that neither of the groups scored systematically higher or lower than the other (Cleary, 1968). The interesting fact is that respective rater groups did indeed score different applicant groups differentially, advantaging their own group by rating their in-group as the superior. The averaging of the ratings allocated by the two rater groups to each applicant group had the effect that the mean rating of both applicant groups was similar. This finding is consistent with social identity theory and the other theoretical frameworks presented in Chapter 2 that predict that individuals will show a preference for similar others. The subjectivity of ratings in the low-structured condition made it easy for raters to subconsciously allocate racially biased ratings. This should by no means be interpreted as overt racial discrimination, but rather as the manifestation of inherent psychological and sociological factors that would promote the perception of all human beings of their own group as a superior one.

The framework proposed by DiDonato et al., (2011) refers to intergroup differentiation, in-group favouritism and differential accuracy as the mechanisms whereby one group would favour their own over another. This, and most other theories, suggest that it could be expected for one group to favour their group above another, but not that a group would intentionally

down-rate another group. In other words, it is expected and confirmed by these results that groups do tend to show favouritism towards the in-group, esteeming them higher, but that they do not necessarily rate an out-group down. This portrays a much more convenient picture than one where the out-group is aggressively targeted and perceived as less competent. There seems to be a positively biased in-group effect and not a negatively biased out-group effect.

The subconsciously biased effect clearly visible in the results of this study should ring a few alarm bells in the employment selection practice, especially in South Africa. If the unstructured selection interview is so susceptible to these effects, the fact that 50% of South African practitioners and 65.3% of practitioners worldwide make use of unstructured or semi-structured interviews should be a serious concern to all human resource practitioners and industrial psychologists (Ryan et al., 2004).

In the highly structured condition the effects portray a completely different picture to the low-structured condition. It is clear that white applicants scored dramatically higher than black applicants, indicating assessment bias in the traditional definition of the term (Cleary, 1968). Although the researcher will not debate the fairness of the selection instrument, i.e., the highly structured interview competencies, questions and rating sheet, the focus will be kept on rater bias as a source of error variance. From this vantage point it is interesting to note that, in the highly structured condition, both rater groups scored the white applicants significantly higher than the black applicants. It was the aim of the structured interview to converge the rating inputs of the raters towards a uniform standard, whereby all raters, irrespective of race, would provide consistently similar ratings to applicants of both race

groups. This effect is clearly visible in the descriptive statistics and the effect of interview structure can be accredited for the much more uniform ratings.

The analyses with HLM and regression, however, indicate the presence of significant similarity effects, seemingly contrary to the results observed in the descriptive statistics. At first, this seems out of place, but when the individual ratings that each of the eight applicants were given are investigated, very interesting and rich information is obtained. Even though white applicants were allocated higher scores on average, it is clear that the racially similar raters scored similar applicants higher than the dissimilar rater group scored each applicant. In other words, white raters allocated consistently higher ratings than black raters to white applicants and vice versa. This was observed, in spite of the previously mentioned fact that there was overall consistency in the way applicants of a different race were rated. The HLM and regression analyses still cleverly identified the underlying racial interaction effects, however. In the highly structured condition, these effects were better masked than in the low-structured condition, requiring extensive analyses and thoughtful interpretation to conclude, that, although to a lesser extent, racial similarity effects were present even in the highly structured condition.

Several reasons can be put forward to explain this unexpected and unprecedented (in terms of parallel research in other parts of the world) results in the highly structured condition. First, the South African context might be more conducive to such effects due to an unfortunate history of racial separation and discrimination. The inherent psychological and sociological processes that facilitate racial similarity effects might be magnified in the social make-up of the average South African, explaining its tenacity in even filtering through the resistant effect of interview structure. Second, the sample of students who participated were by no means

trained interviewers who had any real incentive for making the right decision in terms of selecting the right individual for the right job in an organisation of which they are a part. This factor might put an even stronger hold on the manifestation of racially biased decisions. Third, it is possible that real differences existed. Appropriate measures of bias should disentangle themselves from these real differences in order to serve as meaningful indicators of bias (Guion & Highhouse, 2006).

In order to further make sense of the discrepancy between the results from highly structured interview conditions in this study and prior studies like that of De Meijer et al. (2007), McCarthy et al. (2010) and Sacco et al. (2003), some points of interest are raised. The main difference between this study and those mentioned above arises from the fact that those were field studies and this was an experimental laboratory study. The field studies could not detect the racial similarity effects that theory so boldly proclaims to exist, whereas this study found resounding support for the psychological and sociological theories underlying similarity effects. The studies mentioned above made use of 1) highly structured interviews; 2) highly skilled professional interviewers (in some cases, like McCarthy et al., 2010, industrial psychologists were used); and 3) a real-life job scenario where the raters were highly accountable for the decisions they make, because of, a) their professional reputation, and b) their obligation to provide their employer with their best service. It is clear that a lot more is at stake in these scenarios that could hinder the prevalence of subjectively inspired similarity effects than just interview structure. This study confirms that, when all these inhibitors are removed from the equation, people do make similarity judgements and do advantage their own group above others in rating, as theory suggests. This study identified interview structure as the moderator that would be able to, on its own, remove similarity bias from interviews. The results, however, indicate that structure alone might not be enough to eradicate this bias.

Accountability and professionalism, together with highly structured interviewing, are two constructs that might jointly be accountable for the removal of similarity bias. These factors were not present in this study due to the fact that unaffiliated students were used who were not to be held accountable for their ratings, nor were they professional raters in any sense. In prior studies reporting no significant similarity effects, these two criteria (accountability and professionalism) were present to a large extent, since the studies were field studies that made use of professional interviewers (De Meijer, 2007; McCarthy et al., 2010; Sacco et al., 2003).

It should be asked whether this study or the other studies that have been mentioned reflect the reality of interviewing in the general employment context best: Let's identify the three variables; structure, accountability and professionalism. In this study, only structure was utilised, whereas all three of these 'bias fighting' variables were present in the other studies. The researcher believes that the actual scenario in the general employment context might be plotted somewhere between what is found in comparing this study with prior research. It is known that only 34.7% of interviews worldwide in reality make use of structured interviewing, which immediately implies that there also is a lack of professionalism in interviewing practice (Ryan et al., 1999). Then it becomes very idealistic, even in structured interview settings, to propose that a highly skilled professional, like an industrial psychologist, could be expected to conduct most employment interviews. Second, in smaller businesses where selection decisions are made by executives and enterprise owners themselves, the degree of accountability becomes much smaller than in a large corporate setting. The conclusion can be drawn that, in practice, neither structure, nor professionalism or accountability is a given and also that the use of highly structured interviewing does not automatically include

professionalism and accountability. This study highlights the danger of neglecting either one of these criteria and shows the very real danger of making selection decisions that are influenced by racial similarity bias.

5.5 Limitations

This study did encounter limitations that needed to be addressed. The foremost limitation of the study involved the access to a sample of demographically representative and balanced students. The options with regard to students who could participate in the research were limited to classes for whom permission was granted; therefore it was difficult to obtain a sample of equal racial representation. There consistently were more white raters in the sample than black raters, which diminished the power of the sample, as well as the certainty with which conclusions could be drawn, specifically in terms of the black raters. The power of the sample, however, was adequate enough to form general conclusions with certainty, especially since the effects were so strongly evident in the data.

The sample of students was not affiliated with any organisation, and most of them had not been. They were not subject to the real employment world and lacked the credibility of trained professionals in this area. This created an environment in which all the raters are new to the interview scenario and also had no incentive to produce the best results for their company – as would more likely be the case in practice. This opened the door to subjective opinion in ratings wider, which is exactly what this study saw in the results, even in the proven bias-resistant highly structured interviews. Even though the generalisability of this study needs to be tested in a field study, the nature of the sample has highlighted some inherent psychological effects that were known in theory, but has not often been seen in research. This finding emphasises the importance of using trained professionals for job

interviews – people who have a reputation and a career that will benefit from accurate and unbiased decision making.

Only eight applicants were involved and the researcher is aware that one or two extreme cases of bad or excellent interview performance can shift results quite heftily. This seemed not to have been the case as the videos were carefully developed to check for such anomalies in the stimuli. Yet, it was a risk that had to be taken in order to make the laboratory study work with as much MAXMINCON as possible.

5.6 Recommendations for Future Research

The first and most obvious recommendation would be to take this study to the field, as the limitations of its laboratory characteristics will always be ill compared with the utility of data collected in the field of practice. The study could also be repeated in contexts other than the highly unique South African context, since no other experimental studies on racial similarity have been undertaken abroad. Furthermore, future research can replicate this study using other demographic variables like age and sex in the same way this study tested for racial bias.

In retrospect, the two constructs, professionalism and accountability, might play as big a part in moderating racial similarity effects as interview structure. Future studies could be used to attempt to operationalise these constructs in a manner similar to how interview structure has been operationalised in this study to test their moderating power.

Interesting differences were observed in the way similarity effects were present in the two competencies measured by the highly structured interview. This highlights another area for future research – the effect that different constructs or competencies have on the prevalence

of similarity effects. In other words, investigating which competencies should be avoided and which ones could be included in the selection interview without fear of evoking similarity bias. Findings on such constructs and competencies might also be culturally determined and could provide another niche for culture-specific investigation in this regard.

5.7 Practical Applications

Since highly structured interviews have been shown by this and other studies (McCarthy et al., 2010; Sacco et al., 2003; etc.) to be resistant to demographic similarity effects, the utilisation of structured interviews for employee selection purposes is highly recommended. In the South African context, this is all the more important since the Employment Equity Act No. 55 of 1998 also condemns the usage of invalid and unreliable assessments like the unstructured interview. It should be understood, as seen in the low-structured condition, that mean score comparisons might mask actual bias even though it might fit the Cleary (1968) definition of being unbiased. This study therefore highlights the danger of using unstructured interviews that might at first glance not reveal bias.

Campion et al. (1997) proposed 15 ways in which interviews can be shaped to fulfil the criteria of high structure. The 15 steps or criteria were derived from meta-analytic research on the structured employment interview by Campion et al. (1997), arguably the foremost author in the field of structured interviewing. This study would propose the incorporation of these 15 steps in the development of all selection interviews. The 15 criteria of a highly structured interview are:

1. Base questions on a thorough job analysis
2. Ask each applicant the same set of questions

3. Limit the probing and follow-up questions and elaboration on questions
4. Use good types of questions
5. Use longer interviews or ask more questions
6. Control ancillary information
7. Do not allow applicant to ask questions until after the interview
8. Rate each answer or use rating scales
9. Use anchored rating scales
10. Take detailed notes
11. Use multiple interviewers
12. Use same interviewers for all applicants
13. Do not discuss applicants or answers between interviews
14. Provide extensive interviewing training
15. Use statistical rather than clinical prediction

The practical applications of these results might not be very elaborative, but the essence of the fact that structured interviews should, from a legal and operational point of view, be promoted and implemented in all organisations, proves a very strong and almost imperative requirement.

5.8 Concluding Remarks

The study of racial similarity effects and the moderating effect of interview structure has clear relevance for personnel practice in South Africa. The results of the present laboratory study showed that interview structure does indeed moderate the influence of dyadic racial similarity bias in the employment interview and that the existence of these effects potentially poses a significant problem to practitioners using unstructured employment interviews with

regard to utility and potential legal risks. Interview structure is found to be a moderator that, together with other factors, like professionalism and accountability, can be applied to protect the employer from having racial similarity effects prevalent in selection interview ratings. Future field research should seek to replicate these findings and further, creatively explore the nuts and bolts of similarity effects in ratings, as well as to identify and test moderators that could be useful in limiting their unwanted influence that similarity effects have on judgement and decision making in the employment interview.

REFERENCES

- Alexander, P., & Halpern, R. (2004). Introduction: Comparing race and labour in South Africa and in the United States. *Journal of Southern African Studies*, 30(1), 5-18.
- Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281-322.
- Barth, J., Overby, L. M., & Huffmon, S. H. (2009). Community context, personal contact, and support for anti gay rights referendum. *Political Research Quarterly*, 62(2), 355-365.
- Bevan, S., & Fryatt, J. (1988). *Employee Selection in the UK*. Brighton: Institute of Manpower Studies.
- Blackman, M. C. (2002). Personality judgment and the utility of the unstructured employment interview. *Basic & Applied Social Psychology*, 24(3), 241-250.
- Bovey, W. H., & Hede, A. (2001). Resistance to organisational change: The role of defence mechanisms. *Journal of Managerial Psychology*, 534-548.
- Bretz Jr, R. D., Ash, R. A., & Dreher, G. F. (1989). Do the people make the place? An examination of the attraction-selection-attrition hypothesis. *Personnel Psychology*, 42(3), 561-581.
- Buckley, M. R., Jackson, K. A., Bolino, M. C., Veres III, J. G., & Field, H. S. (2007). The influence of relational demography on panel interview ratings: A field experiment. *Personnel Psychology*, 60, 627-646.
- Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A Review of Structure in the Selection Interview. *Personnel Psychology*, 50(3), 655-702.
- Campion, M. A., Pursell, D. E., & Brown, K. B. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology*, 41, 25-42.

- Capozza, D., & Brown, R. (2000). *Social identity processes*. London: SAGE.
- CIPD. (2008). *Competency and competency frameworks*. London: CIPD.
- Cleary, T. A. (1968) Test bias: prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement*, 5, 115-124.
- Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology*, 80, 565-579.
- Crook, T. R., Todd, S. Y., Combs, J. G., Woehr, D. J., & Ketchen, D. J. (2011). Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. *The Journal of applied psychology*, 96(3), 443-456.
- De Meijer, L. A., Born, M., Van Loon, H., & Van der Molen, H. T. (2007). Through the eyes of the assessor: Demographic and perceived similarity. In L. A. de Meijer, M. Born, H. Van Loon & H. T. Van der Molen, *Through the eyes of the assessor: Demographic and perceived similarity* (pp. 79-101). Amsterdam: Unknown.
- Dipboye, R. L., & Macan, T. M. (1988). A process view of the selection/recruitment interview. In R. S. Schuler, S. A. Youngblood & V. L. Huber,(Eds.), *Readings in personnel and human resource management* (pp. 217-232). St. Paul, MN: West.
- Dipboye, R. L. (2005). The selection/recruitment interview: Core processes and contexts. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell Handbook of Personnel Selection* (pp. 121–142). Malden, MA: Blackwell Publishing.
- DiDonato, T. E., Ullrich, J., & Krueger, J. I. (2011). Social perception as induction and inference: An integrative model of intergroup differentiation, ingroup favoritism, and differential accuracy. *Journal of Personality and Social Psychology*, 100(1), 66-83.
- Elliott, A. G. P. (1981). Sex and decision making in the selection interview: A real-life study. *Journal of Occupational Psychology*, 54, 265–273.

- Fiedler, A. M. (2001). Adverse impact on Hispanic job applicants during assessment center evaluations. *Hispanic Journal of Behavioral Sciences*, 23, 102–110.
- Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M.P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–108). San Diego, CA: Academic Press.
- Fiske, S. T., & Taylor, S. E. (2008). *Social Cognition*. (3rd edn.). New York: Addison-Wesley.
- Foxcroft, C., & Roodt, G. (2005). *An introduction to psychological assessment in the South African context* (2nd edn.) ed.). Cape Town: Oxford University Press.
- Gallois C, Callan V. J, Palmer J. M. (1992). The influence of applicant communication style and interviewer characteristics on hiring decisions. *Journal of Applied Social Psychology*, 22, 1041–1060.
- Gatewood, R. D. & Feild, H. S. (2001). *Human resource selection*. (5th edn.). Ohio: South-Western.
- Gehrlein, T. M., Dipboye, R. L., & Shahani, C. (1993). Nontraditional validity calculations and differential interview experience: Implications for selection interviews. *Educational and Psychological Measurement*, 52, 457-469.
- Glick, P., Zion, C., & Nelson, C. (1988). What mediates sex-discrimination in hiring decisions? *Journal of Personality and Social Psychology*, 55, 178-86.
- Goldberg, C. B. (2005). Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group Organizational Management*, 30, 597-624.

- Graves, L. M., & Powell, G. N. (1995). The effect of sex similarity on recruiters' evaluations of actual applicants: A test of the similarity-attraction paradigm. *Personnel Psychology*, 48, 85–98.
- Graves, L. M., & Powell, G. N. (1996). Sex similarity, quality of the employment interview and recruiters' evaluation of actual applicants. *Journal of Occupational & Organizational Psychology*, 69(3), 243-261.
- Guion, R. M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah: Lawrence Erlbaum.
- Guion, R. M., & Highhouse, S. (2006). *Essentials of personnel selection: Personnel assessment and selection*. Mahwah, NJ: Lawrence Erlbaum.
- Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology*, 42, 691-726.
- Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology*, 53, 575-604. doi:10.1146/annurev .psych.53.100901.135109
- Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1, 333-342.
- Hoffman, B., Lance, C. E., Bynam, B., & Gentry, W. A. (2010). Rater source effects are alive and well after all. *Personnel Psychology*, 63, 119-151.
- Hough, L. M., & Oswald, F. L. (2000). Personnel selection: looking toward the future--remembering the past. *Annual review of psychology*, 51, 631-64.
- Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it? *Psychological Methods*, 5(1), 64-86.

- Hoyt, W. T., & Kerns, M. D. (1999). Magnitude and moderators of bias in observer ratings: A meta-analysis. *Psychological Methods, 4*, 403-424.
- Huffcutt, A. I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment, 19*, 62-81.
- Jetten, J., Spears, R., & Manstead, A. S. R. (1998). Defining dimensions of distinctiveness: Group variability makes a difference to differentiation. *Journal of Personality and Social Psychology, 74*, 1481-1492.
- Judd, C. M., & Park, B. (1993). Definition and assessment of accuracy in social stereotypes. *Psychological Review, 100*, 109-128. doi:10.1037/0033-295X.100.1.109
- Kalin, R. & Hodgins, D. C. (1984). Sex bias in judgements of occupational suitability. *Canadian Journal of Behavioural Science - Revue Canadienne Des Sciences Du Comportement, 16*, 311-25.
- Kenny, D. A. (1991) A general model of consensus and accuracy in interpersonal perception. *Psychological Review, 98* (2), 155-163.
- Kerlinger, F. N., & Lee, H. B. (2000). *Foundations of behavioral research* (4th edn.). Orlando: Harcourt College Publishers.
- Kohn, L. S., & Dipboye, R. L. (1998). The effects of interview structure on recruiting outcomes. *Journal of Applied Social Psychology, 28*, 821-843.
- Kulik, C., & Bainbridge, H. (2006). Psychological perspectives on workplace diversity. In C. Kulik & H. Bainbridge, *Handbook of workplace diversity*. SAGE.
- Latham, G. P, Saari, L. M, Pursell, E. D, Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65*, 422-427.

- Lin, T., Dobbins, G. H., & Fahr, J. (1992). A field study of race and age similarity effects on interview ratings in conventional and situation interviews. *Journal of Applied Psychology*, 77(3), 363-371.
- Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resources Management Review*, 19, 203-218.
- Marchese, M. C., & Muchinsky, P. M. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment*, 1(1), 18-26.
- McCarthy, J. M., Van Iddekinge, C. H., & Campion, M. A. (2010). Are highly structured interviews resistant to demographic similarity effects. *Personnel Psychology*, 6 (2), 325-359.
- McCulloch, M.C., & Turban, D.B. (2007). Using person-organization fit to select employees for high-turnover jobs. *International Journal of Selection and Assessment*, 15(1), 63-71.
- McFarland, L. A., Ryan, A. M., Sacco, J. M., & Kriska, S. D. (2004). Examination of structured interview ratings across time: The effects of applicant race, rater race, and panel composition. *Journal of Management*, 30, 435-452.
- Mount, M. K., Judge, T. A., Scullen, S. E., Sytsma, M. R., & Hezlett, S. A. (1998). Trait rater and level effects in 360-degree performance ratings. *Personnel Psychology*, 51(3), 557-576.
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th edn.). New Jearsey: Pearson Education International.
- Ones, D. S., Viswesvaran, C., & Schmidt, F. L. (2003). Personality and absenteeism: a meta-analysis of integrity tests. *European Journal of Personality*, 17(S1).
- Posthuma, R. A., Morgenson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, 55.

- Prewett-Livingston, A. J., Veres III, J. G., Feild, H. S., & Lewis, P. M. (1996). Effects of race on interview ratings in a situational panel interview. *Journal of Applied Psychology, 81*(2), 178-186.
- Rand, T. M., & Wexley, K. N. (1975). Demonstration of the effect, “similar to me,” in simulated employment interviews. *Psychological Reports, 36*, 535–544.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd edn.). Thousand Oaks: Sage.
- Reid, P. T., Kleiman, L. S., & Travis, C. B. (2001). Attribution and sex differences in the employment interview. *The Journal of Social Psychology, 126*, 205–212.
- Republic of South Africa. (1998) *Employment equity act. Government Gazette, 400* (19370), Cape Town, 19 October.
- Riordan, C. M. (2000). Relational demography within groups: Past developments, contradictions, and new directions. In G. Ferris (Ed.), *Research in personnel and human resource management* (Vol. 19, pp. 131–173). Greenwich, CT: JAI Press.
- Riordan, C. M., & McFarlane Shore, L. (1997). Demographic diversity and employee attitudes: An empirical examination of relational demography within work units. *Journal of Applied Psychology, 82*(3), 342-358.
- Robertson, I., & Makin, P. J. (1986). Management selection in Britain: A survey and critique. *Journal of Occupational Psychology, 59*, 45–57.
- Ryan, A. M., McFarland, L., Baron, H., & Page, R. (1999). An international look at selection practices: Nation and culture as explanations of variability in practice. *Personnel Psychology, 52*, 359–392.
- Sacco, J. M., Scheu, C. R., Ryan, A. M., & Schmitt, N. (2003). An investigation of race and sex similarity effects in interviews: A multilevel approach to relational demography. *Journal of Applied Psychology, 88*(5), 852-865.

- Segrest-Purkiss, S. L., Perrewe, P. L., Gillespie, T. L., Mayes, B. T., & Ferris, G.R. (2006). Implicit sources of bias in employment interview judgments and decisions. *Organizational Behavior and Human Decision Processing*, 101, 152-167.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262–274.
- Schmitt, N. (1976). Social and situational determinants of interview decisions: Implications for the employment interview. *Personnel Psychology*, 29, 79-101.
- Schmitt, N., Pulakos, E. D., Nason, E., & Whitney, D. J. (1996). Likability and similarity as potential sources of predictor-related criterion bias in validation research. *Organisational Behaviour and Human Decision Processes*, 3(68), 272-286.
- Schneider, B. (1987). The people make the place. *Personnel Psychology*, 40 (3), 437-453.
- Scullen, S. E., Mount, M. K., & Goff, M. (2000) Understanding the latent structure of job performance ratings. *Journal of Applied Psychology*, 85 (6), 956-970.
- Simas, K., & McCarrey, M. W. (1979). Impact of recruiter authoritarianism and applicant sex on evaluation and selection decisions in a recruitment interview analogue study. *Journal of Applied Psychology*, 64, 483–491.
- Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures*. Bowling Green: Author.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. (5th edn.). New York: Allyn & Bacon.
- Tepstra, D. E., & Rozell, E. J. (1993). The relationship of staffing practices to organizational level measures of performance. *Personnel Psychology*, 46, 27-48.

- Theron, C. (2007). Confessions, scapegoats and flying pigs: Psychometrics and the law. *SA Journal of Industrial Psychology*, 33(1), 102-117.
- Theron, C. C. (2010a). Intermediate statistics and computer usage. Unpublished class notes. Industrial Psychology 815, University of Stellenbosch.
- Theron, C. C. (2010b) Research methodology & master's research proposal. Unpublished class notes. Industrial Psychology 712, University of Stellenbosch.
- Van Iddekinge, C. H., Perres, G. R., Verrewe, L. V. P., Perryman, A. A., Blass, F. R., & Heetderks, T. D. (2009). Effects of selection and training on unit-level performance over a latent growth modeling approach. *Journal of Applied Psychology*, 94(4), 829-843.
- Walsh, J. P., Weinberg, R. M., & Fairfield, M. L. (1987). The effects of gender on assessment centre evaluations. *Journal of Occupational Psychology*, 60, 305–309.
- Wiesner, W. H., & Cronshaw S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, 61, 275–290.
- Winter, P. A., & Kjørlien, C. L. (2000). Community college faculty recruitment: Effects of job mobility, recruiter similarity, dissimilarity and applicant gender. *Community College Journal of Research and Practice*, 24, 547-566.
- Wilk, S. L., & Cappelli, P. (2003). Understanding the determinants of employer use of selection methods. *Personnel Psychology*, 56(1) 103–124.
- Wiley, M. G., & Eskilson, A. (1985). Speech style, gender stereotypes, and corporate success: What if women talk more like men? *Sex Roles*, 12, 993–1007.

Addendum A

Low structure interview rating sheet and consent form



UNIVERSITEIT • STELLENBOSCH • UNIVERSITY
jou kennisvenoot • your knowledge partner

CONSENT TO PARTICIPATE IN RESEARCH

Research title: The moderating effect of interview design on simulated interview dimension ratings

The proposed study is part of the Master's research by Mr DB Hauptfleisch (Stellenbosch University, Department of Industrial Psychology) and intends to explore how interviewers go about making judgments of others in interview settings. There are a number of factors that we are aware of, from research findings, which may influence the ratings that interviewers assign to interviewees. In this research project, we will ask you to look at a number of videotaped interviews and pretend that you are an interviewer conducting the interview. Then, on forms provided by the researcher, we ask that you rate each interviewee according to the evaluative criteria that are indicated on the ratings sheets.

1. PURPOSE OF THE STUDY

The purpose of the research is to investigate the prevalence of demographic similarity effects in interviews in a South African context and to test for the moderating influence that interview structure might have on these effects.

2. PROCEDURES

You will be required sit in a classroom or auditorium for the duration of your participation in this study. After these consent forms have been filled in and collected, rating sheets will be distributed to you. Instructions will follow that will guide you in the rating process you will be required to take part in. Next, videos of interviews will be shown to all participants on a big screen in front of the classroom or auditorium. You will be asked to rate the interviewees in the video as they would be rated in a real job

interview situation by using the rating sheets provided. The rating sheets will be collected and a post experimental briefing will be given to all the participants before the researcher will thank and dismiss everyone who participated. The rating sheets will be securely stored in a cabinet and will be destroyed two years after the completion of the study.

2.1. RECEIPT OF RATING SHEET

The rating sheets will be handed out to the participants, in person, by the researcher.

2.2. COMPLETION OF RATING SHEET

The participants are requested to read thoroughly and complete the rating sheets individually as they would for employee selection purposes. There are no right or wrong ratings.

2.3. RATING SHEET COLLECTION

After completion, the researcher will personally collect all rating sheets from participants immediately.

3. POTENTIAL RISKS AND DISCOMFORTS

There are no potential risks or discomforts envisaged in this study

4. POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY

Participants will gain experience and skill in the job interview process. A short training (briefing) will be provided on appropriate rating procedure as well as overall briefing on the use of interviews in the selection process.

4.1. BENEFIT TO SCIENCE/SOCIETY

There exists a gap in research on the influence that interview structure has demographic similarity effects in interview ratings. This research will provide clarity by experimentally varying structure. This has implications in terms of fairness, validity, utility and legal defensibility. In pursuit of affirmative development, this study provides insights into fair selection procedures, across racial groups.

5. PAYMENT FOR PARTICIPATION

No payment will be made to participants for participating in this study.

6. CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with the participant will remain confidential and will be disclosed only with the participant's permission or as required by law. Confidentiality will be maintained by means of using coding procedures. The participants are not required to write their names or particulars on the rating sheets. The rating sheets will be issued to the participants by the researcher, in person. On completion of the rating sheets, the researcher will personally collect the rating sheets from all participants. The results will be published in the form of a completed dissertation as well as in an accredited journal, but confidentiality will be maintained. No real names will be published.

7. PARTICIPATION AND WITHDRAWAL

You can choose whether or not to participate in this study. If you volunteer to take part in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer and still remain in the study. The investigator may withdraw you from this study if circumstances arise which warrant doing so, such as ill-health or resignation.

8. IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact DB Hauptfleisch (14550687@sun.ac.za / 084 846 3440) Or Mr FS de Kock (fsdk@sun.ac.za / 021-808 3016 / 082-780 4652)

9. RIGHTS OF RESEARCH SUBJECTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research subject, contact Ms Marlene Fouche (mfouche@sun.ac.za / 021 8084622) at the Division for Research Development, Stellenbosch University.

SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

The information above was described to _____ (your name) by DB
Hauptfleisch in English. I am in command of this language or it was satisfactorily translated to me. I was
given the opportunity to ask questions and these questions were answered to my satisfaction.

I hereby consent voluntarily to participate in this study.

Signature of Participant

Date

SIGNATURE OF INVESTIGATOR

I declare that I explained the information given in this document to _____ [*name
of the subject/participant*]. [He/she] was encouraged and given ample time to ask me any questions. This
conversation was conducted in English and no translator was used

Signature of Investigator

Date



**PLEASE WAIT FOR INSTRUCTION FROM THE
RESEARCHER BEFORE TURNING THE PAGE**

Thank you so much for participating in my research!

**INTERVIEW RATING SHEET**

--	--

Instructions

Ctrl no.

Assume you are a manager who has to choose a successful candidate for a general supervisory position at an imaginary company. You have to assess a few applicants on the grounds of their job interview and judge their suitability for a supervisory position. Please provide a rating of the candidate, from 1 (lowest score) to 100 (highest score), as indicated on the rating sheet below.

You will be shown a series of 3-5-min long interviews on video. Please rate the candidates while you watch, at any point during the video session. There are only 10 sec between interviews and 30 sec at the end in which to finish your rating.

Candidate 1 (Mr Basson):	/100
Candidate 2 (Ms Mfasa):	/100
Candidate 3 (Ms Botha):	/100
Candidate 4 (Mr Hlophe):	/100
Candidate 5 (Ms Johnson):	/100
Candidate 6 (Ms Zuma):	/100
Candidate 7 (Mr Molotsi):	/100
Candidate 8 (Mr Coetzee):	/100

PLEASE MAKE SURE YOU HAVE ASSIGNED A SCORE TO EACH APPLICANT



You will see that applicants respond to each of these questions:

- 1) *Tell me about yourself and your background. Who are you and where do you come from?*
- 2) *What are your other hobbies and interests?*
- 3) *Where do you see yourself in 5 years' time?*
- 4) *How do you usually handle interpersonal conflict situations?*
- 5) *What are some of your personal strengths and weaknesses?*

PARTICIPANT INFORMATION

Please tick the appropriate box

Race

- Black
- Coloured
- Asian
- White

Gender

- Male
- Female

Age

- Age in Years
- Year of study



Addendum B

High structure rating sheet and consent form



UNIVERSITEIT • STELLENBOSCH • UNIVERSITY
jou kennisvenoot • your knowledge partner

CONSENT TO PARTICIPATE IN RESEARCH

Research title: The moderating effect of interview design on simulated interview dimension ratings.

The proposed study is part of the Masters research of Mr. DB Hauptfleisch (Stellenbosch University, Department of Industrial Psychology) and intends to explore how interviewers go about making judgments of others in interview settings. There are a number of factors that we are aware of, from research findings, which may influence the ratings that interviewers assign to interviewees. In this research project, we will ask you to look at a number of videotaped interviews and to pretend that you are an interviewer conducting the interview. Then, on forms provided by the researcher, we ask that you rate each interviewee according to the evaluative criteria that are indicated on the rating sheets.

1. PURPOSE OF THE STUDY

The purpose of the research is to investigate the prevalence of demographic similarity effects in interviews in a South African context and to test for the moderating influence that interview structure might have on these effects.

2. PROCEDURES

You will be required to sit in a classroom or auditorium for the duration of your participation in this study. After these consent forms have been filled in and collected, rating sheets will be distributed to you. Instructions will follow that will guide you in the rating process you will be required to take part in. Next, videos of interviews will be shown to all participants on a big screen in front of the class room or auditorium. You will be asked to rate the interviewees in the video as they would be rated in a real job interview situation by using the rating sheets provided. The rating sheets will be collected and a post



Masters research data collection

experiment briefing will be given to all the participants before the researcher will thank and dismiss everyone who participated. The rating sheets will be securely stored in a cabinet and will be destroyed two years after the completion of the study.

2.1. RECEIPT OF RATING SHEET

The rating sheets will be handed out to the participants, in person, by the researcher.

2.2. COMPLETION OF RATING SHEET

The participants are requested to read thoroughly and complete the rating sheets individually as they would for employee selection purposes. There are no right or wrong ratings.

2.3. RATING SHEET COLLECTION

After completion, the researcher will personally collect all rating sheets from participants immediately.

3. POTENTIAL RISKS AND DISCOMFORTS

There are no potential risks or discomforts envisaged in this study.

4. POTENTIAL BENEFITS TO SUBJECTS AND/OR TO SOCIETY

Participants will gain experience and skill in the job interview process. A short training (briefing) will be provided on appropriate rating procedure as well as an overall briefing on the use of interviews in the selection process.

4.1. BENEFIT TO SCIENCE/SOCIETY

There exists a gap in research on the influence that interview structure has on demographic similarity effects in interview ratings. This research will provide clarity by experimentally varying structure. This has implications in terms of fairness, validity, utility and legal defensibility. In pursuit of affirmative development, this study provides insights into fair selection procedures across racial groups.

5. PAYMENT FOR PARTICIPATION

No payment will be made to participants for participating in this study.



6. CONFIDENTIALITY

Any information that is obtained in connection with this study and that can be identified with the participant will remain confidential and will be disclosed only with the participant's permission or as required by law. Confidentiality will be maintained by means of using coding procedures. The participants are not required to write their names or particulars on the rating sheets. The rating sheets will be issued to the participants by the researcher, in person. On completion of the rating sheets, the researcher will personally collect the rating sheets from all participants. The results will be published in the form of a completed dissertation, as well as in an accredited journal, but confidentiality will be maintained. No real names will be published.

7. PARTICIPATION AND WITHDRAWAL

You can choose whether or not to participate in this study. If you volunteer to take part in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you do not want to answer and still remain in the study. The investigator may withdraw you from this study if circumstances arise which warrant doing so, such as ill-health or resignation.

8. IDENTIFICATION OF INVESTIGATORS

If you have any questions or concerns about the research, please feel free to contact DB Hauptfleisch (14550687@sun.ac.za / 084 846 3440) Or Mr F. S. De Kock (fsdk@sun.ac.za / 021-808 3016 / 082-780 4652)

9. RIGHTS OF RESEARCH SUBJECTS

You may withdraw your consent at any time and discontinue participation without penalty. You are not waiving any legal claims, rights or remedies because of your participation in this research study. If you have questions regarding your rights as a research subject, contact Ms Marlene Fouche (mfouche@sun.ac.za / 021 8084622) at the Division for Research Development, Stellenbosch University.



SIGNATURE OF RESEARCH SUBJECT OR LEGAL REPRESENTATIVE

The information above was described to _____ (your name) by DB Hauptfleisch in English. I am in command of this language or it was satisfactorily translated to me. I was given the opportunity to ask questions and these questions were answered to my satisfaction.

I hereby consent voluntarily to participate in this study.

Signature of Participant

Date

SIGNATURE OF INVESTIGATOR

I declare that I explained the information given in this document to _____ [*name of the subject/participant*]. [He/she] was encouraged and given ample time to ask me any questions. This conversation was conducted in English and no translator was used

Signature of Investigator

Date



**PLEASE WAIT FOR INSTRUCTION FROM THE
RESEARCHER BEFORE TURNING THE PAGE**

Thank you so much for participating in my research!



INTERVIEW RATING SHEET

Instructions

Ctrl no.

Assume you are a manager who has to choose a successful candidate for a general supervisory position at your company. You are part of a diverse panel that are jointly responsible for and accountable to each other for the interviewer rating. You need someone who has strong communicational and people management competence. You have to assess a few applicants on the grounds of their job interview to help make an assessment of the applicant on the above-mentioned grounds.

You will be shown a recorded interview that will comprise of the same questions in the same order as they appear on the rating sheet below. Please read through the rating sheet and then circle the rating you feel is appropriate according to the answer the applicant gives. Please give the rating while the applicant is answering, towards the end of the answer. Study the 'Rating guide' section for guidance on how to rate the individual. A rating of 1 would indicate a very unfavourable answer, in terms of the question and competency at hand, whereas a 7 would indicate an excellent answer, showing brilliant savvy in the specific competency.

COMPETENCY	DESCRIPTION, QUESTIONS AND RATING
<ul style="list-style-type: none"> Communication 	<p>Description: The successful applicant should able to communicate clearly and effectively in a variety of circumstances.</p>
	<p>Q1: What would you do if your manager shouts at you for something you did not do?</p> <p style="text-align: center;">Rating guide</p> <p><i>1: Probably lose my temper, because it's unfair.</i></p> <p><i>4: I would tell him there and then that it is not my fault and tell him who the culprit was.</i></p> <p><i>7: Stay calm and handle the conflict constructively by organizing a meeting to discuss the problem.</i></p> <p style="text-align: center;">Circle your rating</p> <p style="text-align: center;"> Poor Excellent </p> <p style="text-align: center;">C1: 1 – 2 – 3 – 4 – 5 – 6 – 7</p>



	<p>Poor</p> <p>C2: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C3: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C4: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C5: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C6: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C7: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C8: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
<p>Q2: How would you handle a situation where your work colleagues ignore your ideas and input?</p>	
<p style="text-align: center;">Rating guide</p> <p><i>1: I would stop to give my ideas and input.</i></p> <p><i>4: I will go to the manager...</i></p> <p><i>7: I would identify the root of the problem and organize a meeting with individuals I feel would be most effective in resolving the situation.</i></p> <p style="text-align: center;">Circle your rating</p>	
	<p>Poor</p> <p>C1: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C2: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C3: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C4: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C5: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C6: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C7: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
	<p>Poor</p> <p>C8: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>
<p>Q3: How would you go about establishing rapport with your new work colleagues?</p>	
<p style="text-align: center;">Rating guide</p> <p><i>1: I would focus on doing my job well, not the people around me.</i></p> <p><i>4: I would go with the flow and get to know everyone as I work with them.</i></p> <p><i>7: I would focus on serving those around me and find out how I can make the lives of those around me better, while being excellent in my job.</i></p> <p style="text-align: center;">Circle your rating</p>	
	<p>Poor</p> <p>C1: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Excellent</p>



	<p>Poor Excellent</p> <p>C2: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C3: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C4: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C5: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C6: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C7: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C8: 1 - 2 - 3 - 4 - 5 - 6 - 7</p>
	<p>Q4: How would you explain a complex technical issue to someone who has less technical knowledge than you?</p> <p style="text-align: center;">Rating guide</p> <p><i>1: I would try to explain it as well as I can.</i></p> <p><i>4: I would use examples they might understand.</i></p> <p><i>7: I would find out what level of competence they have and explain it in their terms, making sure they follow me, allowing any question.</i></p> <p style="text-align: center;">Circle your rating</p> <p>Poor Excellent</p> <p>C1: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C2: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C3: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C4: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C5: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C6: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C7: 1 - 2 - 3 - 4 - 5 - 6 - 7</p> <p>Poor Excellent</p> <p>C8: 1 - 2 - 3 - 4 - 5 - 6 - 7</p>
<ul style="list-style-type: none"> • People Management 	<p>Description: <i>The applicant should be able to manage employees in achieving organisational goals as well as managing adversity.</i></p>
	<p>Q5: How would you go about motivating your subordinates to work overtime on a big project?</p> <p style="text-align: center;">Rating guide</p> <p><i>1: Threaten them with some kind of punishment.</i></p> <p><i>4: I would find out what incentive I can offer them to make it worthwhile.</i></p> <p><i>7: I would set the example, whilst providing incentives to work over time.</i></p> <p style="text-align: center;">Circle your rating</p> <p>Poor Excellent</p> <p>C1: 1 - 2 - 3 - 4 - 5 - 6 - 7</p>



	<p>Poor</p> <p>C2: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C3: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C4: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C5: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C6: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C7: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C8: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p>
	<p>Q6: How would you go about resolving racial tension between staff members working in your team/department?</p> <p style="text-align: center;">Rating guide</p> <p><i>1: I would not interfere and would let them sort out their own issues.</i></p> <p><i>4: Rebuke them and tell them to get along with one another at work.</i></p> <p><i>7: I would find the root of the problem and work on changing perceptions. If tension amounts to action, disciplinary steps will be taken.</i></p> <p style="text-align: center;">Circle your rating</p> <p>Poor</p> <p>C1: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C2: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C3: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C4: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C5: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C6: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C7: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p> <p>Poor</p> <p>C8: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p>
	<p>Q7: What would you do if your subordinates want to strike in order to get higher salaries?</p> <p style="text-align: center;">Rating guide</p> <p><i>1: Tell them you will take action against them if they strike.</i></p> <p><i>4: Say that we can't offer them the salaries they want and convince them not to strike.</i></p> <p><i>7: Assess the need of the employees against what you can offer. Organize negotiating meetings to resolve the problem before it leads to a strike.</i></p> <p style="text-align: center;">Circle your rating</p> <p>Poor</p> <p>C1: 1 – 2 – 3 – 4 – 5 – 6 – 7</p> <p>Excellent</p>



	Poor									Excellent			
C2:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor											Excellent	
C3:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor											Excellent	
C4:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor											Excellent	
C5:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor											Excellent	
C6:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor											Excellent	
C7:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor											Excellent	
C8:	1	-	2	-	3	-	4	-	5	-	6	-	7
Q8: How would you address the problem that most of your subordinates come late for work on a regular basis?													
Rating guide													
<i>1: Threaten them with some kind of punishment.</i>													
<i>4: Have no tolerance and discipline them for the late coming.</i>													
<i>7: Organize a staff meeting where the severity of coming late, in terms of organisational profit and personal advance will be explained – as well as understanding any legitimate reasons for coming late.</i>													
Circle your rating													
	Poor												Excellent
C1:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C2:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C3:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C4:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C5:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C6:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C7:	1	-	2	-	3	-	4	-	5	-	6	-	7
	Poor												Excellent
C8:	1	-	2	-	3	-	4	-	5	-	6	-	7

PLEASE MAKE SURE YOU HAVE ASSIGNED A SCORE TO EVERY QUESTION OF EACH APPLICANT



PARTICIPANT INFORMATION

Please tick the appropriate box

Race

African

Coloured

Asian

White

Gender

Male

Female

Age

Age in Years

Year of study



Addendum C

Ethical clearance form

ETHICS COMMITTEE APPLICATION FORM
UNIVERSITY OF STELLENBOSCH
SUBCOMMITTEE A

March 2011

Application to the University of Stellenbosch SUBCOMMITTEE A
for clearance of new/revised research projects

This application must be typed or written in capitals

Name: Prof/Dr/Mr/Ms:	MR DB HAUPTFLEISCH
Position/Professional Status:	STUDENT
Affiliation: Research Programme /Institution:	UNIVERSITY OF STELLENBOSCH
Telephone and extension number	084 846 3440

Title of research project: (*Do not use abbreviations*)

A multilevel investigation of demographic similarity effects in interviews and the moderating effect of interview structure

Where will the research be carried out?

At the University of Stellenbosch

All the following sections must be completed (Please tick all relevant boxes where applicable)

1. FUNDING OF THE RESEARCH: How will the research be funded?

Self-funded/Private

2. PURPOSE OF THE RESEARCH:

The purpose of the research is to investigate the prevalence of demographical similarity effects in interviews in a South African context and to test for the moderating influence that interview structure might have on these effects.



3. AIMS AND OBJECTIVES OF THE RESEARCH: (Please list objectives)

In an attempt to address the research needs and the purpose of this study, the proposed study will focus on providing further insight into the following research objectives:

- 1) To establish the extent to which racial similarity effects influence the judgement process in employment interviews in South Africa;
- 2) To determine to what degree interview structure moderates these effects;
- 3) To contribute to determining the generalisability of results done on demographic influence in South Africa to the rest of the western world, by comparing results with similar studies from the United States and Europe.

4. SUMMARY OF THE RESEARCH (give a brief outline of the research plan – not more than 200 words)

Extensive research has been conducted in attempts to uncover the influence that demographic similarity effects have on interviewer judgement and subsequent ratings (Buckley, Jackson, Bolino, Veres III, & Field, 2007; Graves & Powell, 1996; Goldberg, 2005; Harris, 1989; Lin, Dobbins, & Farh, 1992; McCarthy, Iddekinge, & Campion, 2010; Prewett-Livingston, Veres III, Feild, & Lewis, 1996; Sacco, Scheu, Ryan, & Schmitt, 2003; Schmitt, 1976). In summary, though providing a platform for future research, the aggregate of the findings of these studies tend to be inconclusive (Huffcutt, 2011).

Data will be collected from students at the University of Stellenbosch. Practical classes of industrial psychology lecturers will be used, with permission and cooperation of the relevant lecturers. During these classes students will be shown video stimuli of recorded simulated interviews and will be asked to rate the 'applicants' in the videos as they would for a real job. This exercise will form part of a module on conducting research as well as personnel selection.

The data will be analyzed by using HLM (Raudenbush & Bryk, 2002), which is intended to accommodate nested or multilevel data structures. This technique of analysis is appropriate for this study since a number of interviewees will be nested within each interviewer. A series of models will be evaluated in order to test hypothesis when HLM is used as proposed (Sacco et al., 2003).

5. NATURE AND REQUIREMENTS OF THE RESEARCH

5.1 How should the research be characterized (Please tick ALL appropriate boxes)

5.1.1 Personal and social information collected directly from participants/subjects	X
5.1.2 Participants/subjects to undergo physical examination	
5.1.3 Participants/subjects to undergo psychometric testing	
5.1.4 Identifiable information to be collected about people from available records	
5.1.5 Anonymous information to be collected from available records	
5.1.6 Literature, documents or archival material to be collected on individuals/groups	



5.2 Participant/Subject Information Sheet attached? (for written and verbal consent)

YES	
NO	X

5.3 Informed Consent form attached? (for written consent)

YES	X
NO	

5.3.1 If informed consent is not necessary, please state why:

NB: If a questionnaire, interview schedule or observation schedule/framework for ethnographic study will be used in the research, it must be attached. The application cannot be considered if these documents are not included.

5.4 Will you be using any of the above mentioned measurement instruments in the research?

YES	X
NO	

6 PARTICIPANTS/SUBJECTS IN THE STUDY

6.1 If humans are being studied, state where they are selected:

Industrial psychology students from the University of Stellenbosch will be used as permission is granted by lectures and students. Participation will be voluntary unless the lecturer uses the opportunity as part of a practical assessment.

6.2 Please mark the appropriate boxes:

Participants/subjects will:	YES	NO
be asked to volunteer		
be selected	x	



6.3 Are the participants/subjects subordinate to the person doing the recruiting?

<i>YES</i>	
NO	x

6.3.1 If yes, justify the selection of subordinate subjects:

6.4 Will control participants/subjects be used?

<i>YES</i>	
NO	X

6.4.1 If yes, explain how they will be selected:

6.5 What records, if any, will be used, and how will they be selected?

N/A

6.6 What is the age range of the participants/subjects in the study?

18 – 30 years of age

6.6.1 Was assent for guardians/consent for participants/subjects obtained?

<i>YES</i>	
------------	--



6.8 Will the research benefit the participants/subjects in any direct way?

<i>YES</i>	X
NO	

6.8.1 If yes, please explain in what way:

Students will gain experience and skill in the job interview process.

7. PROCEDURES

7.1 Mark research procedure(s) that will be used:

Literature	
Documentary	
Personal records	
Interviews	X
Survey	
Participant observation	X
Other (please specify) _____ _____	

7.2 How will the data be stored?

Data will electronically be captured from the electronic rating sheets. The original rating sheets will be destroyed within two years after capture. The captured data will be stored in a specially created restricted and secure repository on my computer's personal drive. The password will be



8. RESEARCH PERIOD

(a) When will the research commence:

August 2011

(b) Over what approximate time period will the research be conducted:

1-2 month period

9. GENERAL

9.1 Has permission of relevant authority/ies been obtained?

<i>YES</i>	X
NO	

9.1.1 If yes, state name/s of authority/ies:

Department of Industrial Psychology, Stellenbosch University

9.2 Confidentiality: How will confidentiality be maintained to ensure that participants/subjects/patients/controls are not identifiable to persons not involved in the research:

Participants are not required to divulge their names or any information which will personally identify them in any way. The participants who complete the rating sheets will not be prompted for any personal information except their racial information. Personal information of videotaped participants will not be disclosed in the results, but will only possibly be recognizable by participants who know them personally and are part of the rating sample group. Consent



9.6 Any other information which may be of value to the Committee should be provided here:

Date: July 2011

Applicant`s signature: DB Hauptfleisch

Who will supervise the project?

Name: MR FRANCOIS DE KOCK

Institution/Department: MASTERS PROGRAM, DEPARTMENT OF INDUSTRIAL PSYCHOLOGY, UNIVERSITY OF STELLENBOSCH.