

CONSTRUCT VALIDITY OF A MANAGERIAL ASSESSMENT CENTRE

by

ZOVUYO CHULEKAZI NAKO

Submitted in accordance with the requirements for the degree

MASTER IN ARTS

in the subject

INDUSTRIAL AND ORGANISATIONAL PSYCHOLOGY

at the

UNIVERSITY OF SOUTH AFRICA

SUPERVISOR: PROF H A BARNARD

DECEMBER 2011

DECLARATION

I, (Zovuyo Chulekazi Nako, 7605190693 declare that *the Construct Validity of a Managerial Assessment Centre* is my own work, and that all the sources that I have used or have quoted from have been indicated and acknowledged by means of complete references.

SIGNATURE

DATE

ACKNOWLEDGEMENTS

I would like to thank:

- My supervisor, Antoni Barnard, for her advice and guidance on this study;
- My partner, Joerg Taeuber, for his support and the sacrifices he made;
- Anne Buckett for her advice.

SUMMARY

COMPETENCY DIMENSIONS OF A MANAGERIAL ASSESSMENT CENTRE

This was a correlation study exploring the relationships between scores on various dimensions within and across different exercises in the leadership assessment and development centre (LADC) of an auditing firm in Johannesburg. The study specifically aimed at investigating the discriminant and convergent validity of the LADC. LADC ratings collected from a sample of 138 were analysed using a Pearson Product Moment Correlation (r) and principal component analysis (PCA) was conducted to discover the main dimensions or constructs.

Twenty one dimensions were measured using six different exercises in the LADC. The large correlations found in the study showed lack of discriminant validity amongst the majority of different dimensions measured in same exercises whilst, the PCA showed some convergent validity among various dimensions measured across exercises for the LADC. Lastly, the findings of the principal component analysis (PCA) supported a two-factor structure, indicating that assessors are able to differentiate between interpersonal and performance-related dimensions.

KEY TERMS

Assessment centre, competencies, competency-based assessment, dimensions, construct validity, discriminant validity, the principal component analysis, Pearson Product Moment Correlation

TABLE OF CONTENTS

	PAGE
CHAPTER 1 SCIENTIFIC ORIENTATION TO RESEARCH	
1.1 INTRODUCTION	12
1.2 BACKGROUND AND MOTIVATION FOR RESEARCH	12
1.3 PROBLEM STATEMENT	19
1.4 AIMS OF THE STUDY	19
1.4.1 General aim	19
1.4.2 Specific aim	20
1.5 PARADIGM PERSPECTIVE	20
1.5.1 Disciplinary framework	21
1.5.2 The underlying theoretical orientation (trait-activation theory)	21
1.6 RESEARCH DESIGN	22
1.6.1 Type of study	23
1.6.2 Unit of analysis	23
1.6.3 Strategies to ensure reliable and ethical research	23
1.6.3.1 Ensuring reliability	24
1.6.3.2 Complying with ethical research principles	24
1.7 RESEARCH METHODOLOGY	25
1.7.1 Literature review	25
1.7.2 Empirical study	26

1.7.2.1	<i>Sampling procedure</i>	26
1.7.2.2	<i>Measuring instruments</i>	26
1.7.2.3	<i>Data gathering</i>	29
1.7.2.4	<i>Data analysis</i>	29
1.7.2.5	<i>Formulation of hypothesis</i>	30
1.8	CHAPTER LAYOUT	30
1.9	CONCLUSION	31

**CHAPTER 2 COMPETENCY ASSESSMENT AND ASSESSMENT
CENTRE METHODOLOGY**

2.1	INTRODUCTION	32
2.2	HISTORY OF PSYCHOLOGICAL TESTING IN SOUTH AFRICA	33
2.3	COMPETENCY BASED ASSESMENTS	36
2.3.1	Conceptualisation of competencies	38
2.3.1.1	<i>Definition of competency</i>	38
2.3.1.2	<i>The importance of competency-based assessment in the workplace</i>	42
2.3.2	Competency Development	45
2.3.2.1	<i>The design of competencies</i>	45
2.3.2.2	<i>Development of competency frameworks</i>	47
2.3.2.3	<i>Development of the competency - based - assessment battery</i>	51
2.4	ASSESSMENT CENTRE METHODOLOGY	54

2.4.1	Conceptualisation of assessment centre methodology	54
2.4.1.1	<i>Definition of assessment centre</i>	55
2.4.1.2	<i>Characteristics of assessment centres</i>	55
2.4.1.3	<i>The uses of assessment centres</i>	57
2.4.1.4	<i>Designing an assessment centre</i>	58
2.4.2	The utility of assessment centres	62
2.4.2.1	<i>The predictive validity in assessment centres</i>	62
2.4.2.2	<i>Construct validity issues in assessment centres</i>	64
2.4.3	Research: AC construct validity	65
2.4.3.1	<i>Correlation studies</i>	65
2.4.3.2	<i>Factor analysis</i>	67
2.5	CONCLUSION	79

CHAPTER 3 ARTICLE

ABSTRACT	81
Assessment Centres	82
Validity studies on assessment centres	83
METHOD	88
Participants	88
The assessment centre methodology	89
<i>Dimensions</i>	89
<i>Assessment centre exercises</i>	89
<i>Assessors and assessor training</i>	91

	Procedure	91
	RESULTS	92
	DISCUSSION	105
	REFERENCES	110
CHAPTER 4	CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS	
4.1	INTRODUCTION	117
4.2	CONCLUSIONS	117
4.3	LIMITATIONS	122
4.4	RECOMMENDATIONS	123
4.4.1	Recommendations for practice	123
<i>4.4.1.1</i>	<i>Assessment centre dimensions</i>	<i>124</i>
<i>4.4.1.2</i>	<i>Assessors</i>	<i>124</i>
<i>4.4.1.3</i>	<i>Rating process</i>	<i>125</i>
4.4.2	Recommendations for future research	125
4.5	CONCLUSION	126
	REFERENCES	127

LIST OF TABLES

		PAGE
Table 1.1	Dimensions and LADC exercises	29
Table 2.1	Competency models containing key managerial competencies	50
Table 2.2	Previous research of discriminant validity WER vs AER	67
Table 3.1	Race and gender breakdown of LADC participants (N=138)	88
Table 3.2	Dimensions and LADC exercises	91
Table 3.3	Descriptive statistics for dimensions including composite scores	93
Table 3.4	Correlations between dimension in the SPS	94
Table 3.5	Correlations between dimensions in the executive interview	94
Table 3.6	Correlations between dimensions in the ALS	95
Table 3.7	Correlations between dimensions in the CBI	95
Table 3.8	Correlations between dimensions measured in the PSYCH	96
Table 3.9	Correlations between dimensions measured in the PON	97
Table 3.10	Correlations between dimensions across exercise (convergent validity)	99
Table 3.11	Rotated Factor Matrix ratings on the assessment centre exercises	101
Table 3.12	Factor matrix of composite scores on 21 dimensions	104

LIST OF FIGURES

		PAGE
Figure 2.1	Conditions leading to direct and indirect discrimination and adverse impact	36
Figure 2.2	The iceberg model and surface and core personality competencies	41
Figure 2.3	The basic design model	60

APPENDICES

Appendix A **Definitions of dimensions measured in**
the LADC

142

CHAPTER 1
SCIENTIFIC ORIENTATION TO RESEARCH

1.1 INTRODUCTION

Owing to their ability to predict future job performance, assessment centres (ACs) have gained much popularity over the years both internationally and in South Africa. In South Africa the popularity of ACs has further increased due to its positive impact on fairness in a multicultural assessment context, as it encompasses fairness principles based in its job-relatedness, it being competency driven and in its multi-method approach. Despite the potential fairness in the approach and despite evidence of predictive validity, controversy in the use of ACs has been sparked by limited evidence available to support their construct validity. Given this dilemma, this study aimed to investigate the construct validity of ACs. In particular this study focused on the competency dimensions in a leadership assessment and development centre (LADC). The LADC was intended to identify leadership potential among business unit managers for admission as business partners in a South African auditing firm.

In chapter 1, the context within which the study was conducted, is explained. The chapter commences with a background and motivation for the research. Thereafter, the problem statement, the research aims, the theoretical orientation of the research, the research design and research methodology are discussed. The chapter concludes with a chapter summary highlighting the remaining chapters in this dissertation.

1.2 BACKGROUND AND MOTIVATION FOR THE RESEARCH

In ACs several dimensions, exercises and assessors are used to arrive at judgments about the potential shown by candidates, such as the potential to

perform well in a particular position. Firstly, ACs are constructed around various dimensions or competencies¹ related to a particular job or work environment. Performance-related dimensions, such as decision-making, problem-solving, analytical thinking and adaptability, are examples of dimensions assessed and the various exercises included in the AC to measure these dimensions are referred to as methods (Greyling, Visser & Fourie, 2003; Lievens, 1998). Usually more than one assessor administers the AC and evaluates AC participants on the various dimensions with the different methods included in the AC. The AC is therefore termed a multi-method, multi-trait, multi-rater approach (various exercises are used and multiple dimensions are observed and measured by more than one assessor) and the assessors are normally trained in observing the dimensions exhibited (Thornton & Rupp, 2006).

AC exercises may consist of psychometric testing, interviews and simulation exercises. Joiner (2000) suggested an inclusion of at least one simulation exercise, which must be job relevant enough to elicit performance related behaviors from the candidates. ACs are also unique in that candidates' behaviours are classified and categorised meaningfully, into relevant dimensions or attributes (Joiner, 2000). Whilst other forms of assessment look at the candidate's current performance, ACs' advantage is that they have an ability to identify potential of candidates particularly that of managers (Charoux, 1991). Another distinguishing characteristic of ACs is the feedback participants receive. Lievens (2002) pointed out that ACs are advantageous in that participants receive detailed feedback concerning their strengths and weaknesses on the dimensions measured. In compliance with AC guidelines, the intention of an AC is to measure behavioural dimensions in a number of ways and to recognise that these measures provide repetitive rather than unique information about distinct work-related attributes (Haaland & Christiansen, 2002; Joiner, 2000). The AC exercises are designed to have different behavioural requirements to ensure that a wide range of relevant job behaviours is demonstrated (Joiner, 2000). According to Joiner (2000), a proper job analysis of job-relevant behaviours should be conducted to determine the dimensions, attributes and job performance indices that are

¹ Competencies or dimensions are often used interchangeably in literature; however, in this dissertation dimensions will be used to refer to the competencies used in ACs.

important to job success in order to identify what dimensions should be evaluated in a particular AC.

As a result of the popularity that ACs have gained, many organisations are currently using them for selection, development and promotional purposes (Spychalski, Quinones, Gaugler, & Pohley 1997). Because of the increase in the use of ACs, the need for standardisation of AC procedures has gained momentum. Consequently, much research has been conducted since the late eighties to examine AC validity, especially in the United States and in the United Kingdom (Kriek & Charoux, 1994). Investigating the utility of AC in the South African context, Kriek and Charoux (1994) reported on the fairness of ACs based on race, gender and age, and found no evidence of bias in the predictive validity of ACs. Schlebusch and Roodt (2008) referred to ACs as an ideal method for selection decisions, owing to the methods' cultural appropriateness, which minimises the chances of unfair discrimination. Moreover, Schlebusch and Roodt (2008) highlighted the adherence of the AC procedure to South African legislation, such as the Employment Equity Act (EEA), Act no 55 of 1998.

Due to the fact that AC exercises are loaded with more interpersonal skills than cognitive ability, Harel, Arditi-Vogel and Janz (2003) suggested that the likelihood of adverse impact is a lot less, especially for Black participants. According to Greyling et al. (2003) the research focus in studies on ACs has mostly been on the predictive nature of ACs. According to Schmidt and Hunter (1998) the predictive validity of a measure is the most important property in personnel selection and the cumulative research knowledge makes it possible for employers to increase the productivity, output and learning ability of their workforce substantially by using procedures that work well and by avoiding those that do not. Within the South African context, EEA requires the use of assessment procedures that provides sufficient support for inferences of job relatedness, thus validity amongst others (Muchinsky, Kriek & Schreuder, 2002).

Validity as a fundamental characteristic of a measure refers to the appropriateness, accuracy and the extent to which a test measures what it claims to measure (Gregory, 2004; Muchinsky et al., 2002). There are three concepts which contribute to a total picture of a test's validity namely, predictive, content and construct validity concepts (Schmidt, 2006). These according to the author fall under the ambit of a trinitarian view to validity (Schmidt, 2006). Within this view, predictive validity is concerned with the extent to which a measure allows for accurate concurrent and future predictions of criteria, while the content validity is concerned with the degree to which a measure provides a good sample of the domain of behaviors that it is intended to measure (Gregory, 2004). The construct validity which is considered as the most important form of validity is concerned with the extent to which a test measures the psychological constructs it is supposed to measure (Gregory, 2004; Messick, 1980; Muchinsky et al., 2002). The trinitarian view according to Gregory (2004) and Schmidt (2006) is a more technically focused concept of validity, thus, viewing validity within the scientific domain which excludes the concept of values. Value concepts refer to issues such as biasness or fairness of a measure (Messick, 1980; Muchinsky et al., 2002). Although the three validity concepts are not mutually exclusive, the trinitarian approach claims that depending on the purpose in which the test is used for, one does not necessarily need evidence of all three (Schmidt, 2006). Moreover, according to Muchinsky et al. (2002), construct validity is the most important of the three and therefore encompasses both the content and predictive validities. Thus, the above applied in an AC would mean that if content and predictive validities are evident in an AC as widely believed, then construct validity evidence should also be expected.

Although the scientific and technical domains of validity are important, one wonders to what extent this view is applicable in the South African context, where the EEA requires tests used to be fair, and free of bias while providing validity evidence. Against this background, this is a gap that lies within the trinitarian approach, in which only the technical and scientific aspects are considered in test use, while ignoring the value judgements. This view was also expressed in Messick (1980), in which the author argued that tests cannot be evaluated only in scientific terms, but rather evaluated also in their potential social consequences. Similarly, SIOPSA (2005) advocates for continued

research in fairness and bias issues in testing, particularly those used locally. In fact, according to Gregory (2004) psychological testing is an applied science that occurs within a political and social context, not necessarily a neutral undertaking. Therefore, in South Africa where companies are under moral economic and political pressures to diversify their workforce, psychological tests that are fair and free from bias and adverse impact are required (Theron, 2009).

Predictive validity studies on AC clearly show the utility of AC in selection and development practices. In a meta-analytical study Gaugler, Rosenthal, Thornton and Bentson (1987), for example, found validity outcomes of AC dimensions to be as high as .40. Similarly, Arthur, Day, McNelly and Edens (2003) and Jansen and Stoop (2001) reported on the predictive evidence of ACs, particularly for advancement criteria such as career progress, salary advancement, long-term promotion and potential development. Harel et al. (2003) reported 17 reviews of meta-analytical studies, through which the AC method is ranked as the most valid assessment method, when job performance ratings are used as criteria. Although the above discussion indicates the predictive support of ACs among researchers, lack of convergent and discriminant validity and thus construct validity has been reported in literature (Arthur, Woehr & Maldegen, 2000; Kleinmann, 1993; Lievens, 1998; Shore, Thornton & Shore, 1990).

The absence of construct validity in studies on ACs has been explained differently by various researchers. Arthur et al. (2000) explained poor construct validity of AC in terms of AC procedures and implementation, which may add measurement errors that prevent appropriate convergent and discriminant validities from being obtained. Furthermore, the authors also claimed that ACs measure constructs other than those originally intended by the centre design (Arthur et al., 2000). In addition, Jansen and Stoop (2001) suggested in their longitudinal study of AC validity that the variance and changes in validity of ACs is due to changes in critical work elements as a result of organisational and societal developments, as well as changes during an individual's career progression. Lievens (2008) proposed that the lack of construct validity in ACs could be due to the fact that AC performance is exercise-specific. Thus,

inconsistency in performance may result from some participants performing better in one exercise and worse in other exercises, consequently obtaining higher ratings in all the dimensions for those exercises in which they performed better and lower ratings in the dimensions of exercises where their performance was worse. According to Arthur, Day, and Woehr (2008) this is due to the fact that AC theory has not been adequately and satisfactorily tested. In fact, the authors recommend a focus on the manner in which AC dimensions are conceptualised, scored and operationalised, in order to mitigate the exercise specificity in ACs.

Another factor according to Lievens (2008) is that some performance dimensions are more observable in some exercises than in others. Thus, a candidate may obtain a low score in one exercise, owing to the inability of the exercise to reveal the candidate's behaviour. In this regard, Howard (2008) claimed that more research on how well the dimensions are differentiated across exercises is required. Furthermore, dimensions should be designed to ensure that they elicit relevant behaviors and guide assessors in terms of their relevance (Howard, 2008).

Lievens (2002) reported that assessors' bias and inaccuracy, which are the result of poor assessment design, adds another dimension to the lack of construct validity in ACs. The halo effect may also be a contributory factor because assessors may not distinguish between dimensions in the same exercise. Typical factors that may affect assessors' ratings, according to Greyling et al. (2003), are social desirability, personality factors, the actor-observer effect and physical attractiveness. The within-exercise attribute rating method is indicated in Lance (2008) as contributing to the lack of construct validity in ACs. When using this approach, assessors tend to give within-exercise attribute ratings after observing one candidate in one exercise. However, moderate levels of convergent validity have been indicated of the same attribute evaluated in more than one exercise. High method variance has also been indicated when this approach is used, which is demonstrated by high correlations between different dimensions in the same measure (Greyling et al., 2003).

Several approaches have been used to study the construct validity of dimension ratings of ACs. Shore et al. (1990) have suggested Campbell and Fiske's (1959) multi-trait multi-method (MTMM) matrix to determine construct validity. This occurs when a set of traits is measured by a number of methods and the results are presented in a correlation matrix called the MTMM matrix (Greyling et al., 2003; Lance, 2008). According to this approach, when strong correlations occur between the same dimensions measured in different AC exercises, convergent validity is displayed, and when the correlation is weak between different dimensions as measured within one method, discriminant validity is demonstrated. To confirm the construct validity of an AC, one would expect that there should be convergent validity between the same dimensions across different methods and discriminant validity across different dimensions within one method.

Lance's (2008) argument is that the lack of construct validity in ACs is actually due to the use of the MTMM matrix design. He further reported that the exercises fail to assess behaviour with regard to the dimensions that are presented to assessors as they are presumed to reflect cross-situationally consistent behavioural categories. Another method that has been used to study construct validity in ACs has been confirmatory factor analysis (CFA) to determine whether dimensional ratings in exercises can be presented by a small subset of dimensions, grouped together, for example, into problem-solving or interpersonal sensitivity (Lievens & Van Keer, 2001). Although CFA has been widely used to generalise construct validity evidence, Kudish, Ladd and Dobbins (1997) suggested that caution is advised and more research needs to be conducted. The fact that AC construct validity has until recently been much debated by researchers from various perspectives (e.g Arthur et al., 2008; Howard, 2008; Lance, 2008; Lievens & Conway, 2001; Meriac, Hoffman, Woehr & Fleisher, 2008), further emphasised the need for continued research in this regard.

From the discussion above it seems evident that most studies investigating construct validity of ACs use the MTMM approach as well as CFA, but that

construct validity of AC remains a contentious issue that should continually be explored and confirmed.

1.3 PROBLEM STATEMENT

Extensive knowledge exists about the predictive nature of ACs and AC ratings have been found to demonstrate high ability to predict candidates' subsequent job success (Gaugler et al., 1987; Schmidt & Hunter 1998). Despite all the evidence of predictive validity, less support has been shown in research for the construct validity of ACs, leading to the conclusion that AC methods are probably not as effective in measuring the dimensions they claim to measure (Arthur et al., 2000; Greyling et al., 2003; Haaland & Christiansen, 2002; Lievens, 2008; Lievens & Conway, 2001). Furthermore, it is important to understand why selection instruments work and what constructs they measure. In order for test score inferences to be useful for personnel selection, they must be based on underlying constructs common to both test performance and job behaviour and these underlying constructs should be involved in all measurement. It is also important to know whether assessors' judgments accurately reflect the underlying constructs in dimensions, since these ratings constitute the basis for judgments of overall managerial potential as well as for planned development (Haaland & Christiansen, 2002). Hence construct validity of AC dimensions requires continued investigation.

1.4 AIMS OF THE STUDY

Having highlighted the problems associated with the construct validity of ACs, the following general and specific aims of the research were formulated.

1.4.1 General aim

The general aim of this study was to investigate the construct validity of an LADC applied in a South African auditing firm over the period of five years and in

doing so, to extend research and build on existing knowledge about the construct validity of ACs.

1.4.2 Specific aims

In investigating the construct validity of ACs, the specific aims of the study were to:

- Conduct a literature review on AC as a competency-based assessment in the multicultural assessment context of South Africa
- Determine whether different dimensions measured within exercises are significantly discriminated from one another (discriminant validity) in the LADC
- Determine whether there are significant correlations between similar dimensions across exercises in the LADC (convergent validity)
- Whether the number of dimensions measured in the LADC can be collapsed by meaningfully grouping dimensions into a smaller number of dimensions.

1.5 PARADIGM PERSPECTIVE

Paradigms refer to interrelated systems of thinking and practice that govern the nature and the manner in which to enquire and on which to base assumptions (Terre Blanche & Durrheim, 2004). Through the use of a particular paradigm, the researcher is provided with the rationale for the research to commit to certain methods of data collection, observation and interpretation (Terre Blanche & Durrheim, 2004). Implicit in explicating one's research paradigm, is an understanding of the disciplinary framework in which the research is conducted as well as the underlying theoretical orientation driving the research.

1.5.1 Disciplinary framework

The research falls within the discipline of industrial and organisational psychology. Particularly, the research focuses on industrial psychological assessment, which is a sub-discipline of industrial and organisational psychology. Muchinsky et al. (2002) described psychology as studying both the thinking and behaviour of humans and animals scientifically. Evans (1997), cited in Muchinsky et al. (2002), explained industrial and organisational psychology as a scientific study of people in their work environments, which includes scientific observations, evaluation, optimal utilisation and influencing. Foxcroft and Roodt (2001) described psychological assessment as a process of gathering a wide range of information through the use of assessment measures and other sources such as interviews, which is then evaluated and integrated to reach a conclusion. The underlying theoretical orientation to the study is based on the interactionist approach (trait-activation theory) which is congruent to many practices and principles in the field of psychological assessment.

1.5.2 The Underlying Theoretical Orientation: Interactionist (Trait-activation Theory)

The interactionist approach which is a combination of the trait and situational approaches is based on the assumption that neither traits, nor situation specificity are determinants of behaviour; it is rather the interaction of these two factors that determines behaviour (Heindl, 1997). Endler & Magnusson (in Heindl, 1997), suggested that there is a continuous interaction between individuals and the situations they find themselves in. It is therefore important that both the person variable and environmental variable be taken into account in order to understand the person-situation interaction. For the purpose of this study, behaviour exhibited in the AC exercises and its evaluation was conceptualised within the recent interactionist approach, trait-activation theory.

The trait-activation theory explains how the behaviour of job-relevant behaviors is elicited and rated in ACs. The theory focuses on the person-situation interaction to explain behaviour based on responses on trait-relevant

cues found in situations (Lievens & Thornton, 2005). These behavioural responses serve as the basis for behavioural ratings (Tett & Guterman, 2000). According to the trait-activation theory, a person's trait level is expressed as trait-relevant behaviour at work (Haaland & Christiansen, 2002) and only when trait-relevant cues are present are such traits expressed. This theory is relevant to AC research in that a person's trait level is measured as a score on a trait-related competency, which is based on behaviour in certain AC exercises. However, to obtain an objective rating, the trait-action potential of an AC exercise must be available, which is determined by trait-relevant exercise cues and rewards provided (Tett & Guterman, 2000). Haaland and Christiansen (2002) described trait-activation potential as the capacity to observe differences in trait-related behaviour in a given situation. Exercise cues are job demands, social demands and the organisational structure, which are simulated to elicit trait behaviour, while rewards are presented in specific instructions that provide information to participants about what to do (Lievens & Thornton, 2005).

A study of the construct validity of ACs according to the trait activation theory provides a deeper understanding of both the convergent and discriminant validity of ACs. Lievens (2002) asserted that the theory is advantageous in AC validity in that convergence can be expected among dimension ratings, but only in exercises that provide an opportunity to observe behaviour related to the same trait. The theory is also used as it provides a greater theory-driven strategy and it entails mapping the trait-activation potential of ACs.

1.6 RESEARCH DESIGN

Research design is referred to as a blueprint of the research project that is synonymous with decision-making during the research process (Monk, 2001). Terre Blanche and Durrheim (2004) defined research design as a strategic framework for action that provides a bridge between the research question and the implementation of the research. In explicating the research design driving this study, the type of study is discussed next, the unit of analysis is

highlighted and strategies followed to ensure reliable and ethical research are noted.

1.6.1 Type of study

The study is quantitative, non-experimental and explorative in nature. Variables were relationally analysed with no sample randomisation, manipulation of variables or use of control groups.

1.6.2 Unit of analysis

For the purpose of this study the unit of analysis was the individual. There are four different units of analysis, namely, groups, organisations, social artifacts and individuals (Terre Blanche & Durrheim, 2004). It was mentioned earlier that in ACs individual behavior is measured against a number of dimensions using various exercises. Against this background, individual assessment ratings on identified dimensions were collected for analysis in this study, hence the unit of analysis is at an individual level.

1.6.3 Strategies to ensure reliable and ethical research

The EEA requires the use of reliable and valid assessment tests. Thus, any psychological test must meet such requirements. Reliability refers to a measure's consistency in measuring what it measures (Foxcroft, 2004). Reliability in the LADC under study was ensured through a standardised assessor training process, to guarantee consistency in rating of the candidate's behaviour. According to Joiner (2000) assessors are supposed to receive an intensive training prior to participating in an AC. The rights and safety of the participating firm and that of LADC participants were protected by complying with the widely used research ethical guidelines and principles. According to Terre Blanche and Durrheim (2004), any research design should reflect careful considerations to ethical issues attached to research projects. In this instance, the researchers in this study followed the ethical guidelines and principles documented in Terre Blanche and Durrheim (2004). The manner in

which reliability and compliance to ethical guidelines and principles was ensured under this study is discussed below.

1.6.3.1 Ensuring reliability

Assessors included three partners/directors and three business unit managers. One business unit manager was an industrial psychologist. Assessors attended a four-hour training workshop, designed for fundamental assessor skills, focusing on observation, recording, classification and the rating of candidate's behaviour. Moreover, the assessor training model of Saunders (2000) informed the training framework. The workshop included sharing with the assessors information about how the competencies were identified and how the LADC was designed as well as providing them with knowledge of the context in which the AC was being conducted. The relevance and definition of each dimension was explained to the assessors. Furthermore, assessors were instructed to write behaviour descriptions of candidates' performance classify these behaviours according to dimensions and thereafter, rate each candidate's performance. This was followed by instructions on how to minimise bias, subjectivity and poor judgements in assessments. The manner in which reports were to be compiled and feedback was to be provided was also discussed during the training workshop.

Assessor training workshops were held four times a year. The information was sent to the relevant partners/directors in the firm and they were required to sign up for one workshop.

1.6.3.2 Complying with ethical research principles

According to Terre Blanche and Durrheim (2004) there are three ethical principles to research (e.g autonomy, nonmaleficence and beneficence) and the ethical guidelines relate to consent, confidentiality and researcher competence. It is important to note that the current study used secondary data, in that data was initially gathered for identification of leadership potential than for this research. Therefore, no consent was sought from the assessment

participants, but rather obtained from the participating firm. The researcher forwarded a formal letter to the participating firm to obtain the LADC ratings for the purposes of this research. In this regard, the anonymity and confidentiality of both the company and participants were protected in all aspects of the research process. Thus, nothing was revealed about the company that would allow anyone to identify the company or the candidates, nor would anyone know which information had been provided by the firm. The names and the identities of the candidates were not used, nor were any personal details given. Furthermore, the firm was made aware that the research was part of studies for a degree at a university and that the information obtained would be read by the researcher's supervisor. The information obtained would be used for the dissertation, further research and publications (for example, in journal articles and for presentations at conferences or public talks), but with strict adherence to the above undertaking about anonymity and confidentiality.

Firstly, the autonomy principle was addressed through obtaining informed consent from the participating firm. Secondly, the nonmaleficence was addressed through protecting the identity of the firm and that of participants in the LADC. Lastly, the beneficence principles will be addressed through recommendations for future practice and research in construct validity of ACs.

1.7 RESEARCH METHODOLOGY

The research method consisted of a literature review, the empirical study and the conclusions, limitations and recommendations. These are discussed below.

1.7.1 Literature review

A literature review of competency-based assessment and AC methodology was conducted in the multi-cultural assessment context of the South African workplace.

1.7.2 Empirical study

Methodologically this study is similar to the construct validity study of Greyling et al. (2003). This study however differs in that it was conducted in a different research and business context. For the purpose of this study, ratings from an LADC of South African auditing firm was collected, while Greyling et al. (2003) studied AC dimensions assessed in a team leader AC. Furthermore, in the current study data were gathered from LADCs spanning the period from 2006 to 2010, whilst in the Greyling et al. (2003) study, data were gathered over 11 months. The Greyling et al. (2003) AC also included different dimensions from those used in the current study. The empirical study consisted of the following steps:

1.7.2.1 Sampling procedure

The sample consisted of the results of 138 candidates who participated in the LADC in the research organisation for admission as a partner/director in the company. The LADC was designed to assess leadership potential rather than managerial performance and candidates were assessed against the benchmark of third-year partners/directors in their respective service lines. This enabled the assessment panel to determine how close or far off candidates were in terms of readiness for admission to partnership. Furthermore, the LADC was designed to provide the firm with an objective assessment of an individual's ability to function at a leadership level, and as a means of providing guidance to the business unit and strengths and development areas to be addressed towards admission to the partnership.

1.7.2.2 Measuring instruments

In order to assess the identified competencies, assessment techniques must be designed to evaluate dimensions identified in the job analysis (Joiner, 2000). Twenty one dimensions were measured using seven exercises. Out of the seven exercises, there were four simulations exercises, an executive interview and two psychometric tests, (Occupational Personality Questionnaire (OPQ) of SHL and the Cognitive Process Profile (CPP) of Cognadev. In this

section, exercises and the competencies measured in the LADC are discussed.

i. Competency dimensions

A job description or job profile of a partner or director was developed prior to assessment, which was used as a basis for competency identification. Therefore, the candidates were assessed against behaviors required for the job rather than a set of generic or fixed competencies. The job profiles informed the 21 competencies measured in the LADC. Appendix 1 provides definitions of all the competencies measured in the LADC. Behavioural anchors were then assigned to these dimensions. These behavioural anchors on particular dimensions were rated on a 5-point Likert scale (1 - unacceptable, 2 - less than acceptable, 3 - acceptable, 4 - more than acceptable, and 5 - very high level of ability). Thus, 1 represents a very negative display according to the criterion and 5 a very positive display of the dimensions. The 21 dimensions were client orientation, applied judgment & insight, business acumen, knowledge-up-to-datedness, practice leadership, winning business, product and service knowledge, creativity and innovation, written communication, oral communication, personal impact personal integrity, interpersonal skills, achievement drive, team work, respect for diversity, negotiation, resilience, flexibility, leadership and general judgement.

ii. Assessment centre exercises

The following exercises were used to measure the dimensions stated above. Among these were four simulation exercises and an interview.

- *Self-presentation skills (SPS)* - the candidates are required to prepare a ten-minute presentation of the practice office nomination exercise described below, using any presentation aids, to the panel.
- *The executive interview (EXEC)* - This is a one-on-one interview conducted by an Exco member as an interviewer. The interview guidelines are provided to the interviewer prior to the interview.

- *Applied Leader Skills Assessment* - The ALS comprises a number of emails that a company's partner/director might typically find in his or her inbox. Once candidates have had an opportunity to respond to the emails, they meet with the ALS panel to discuss their responses. The purpose of the ALS is to assess the candidates' strengths and development needs on a number of competencies considered important for leaders. Furthermore, the ALS exercise is tailored to be relevant for each specific business unit.
- *Criteria-based Interview (CBI)* - This assessment comprises a panel interview focused on defined core competencies.
- *Psychometric assessment (PSYCH)* - as stated earlier psychometric tests comprised of the OPQ and the CPP. The CPP measures various constructs such as the cognitive styles (namely; explorative, analytical, structured, holistic, intuitive, memory, integrative, logical reasoning, reflective, a balanced profile, learning, random, impulsive, metaphoric and quick insight) and work- related processing aspects (Van Heerden, 2005). The OPQ is a personality questionnaire that describes 32 personality dimensions of people preferred styles of behavior at work (Nobre, 2005). The 32 dimensions measured by the OPQ are persuasive, controlling, outspoken, independent minded, outgoing, affiliative, social confident, modest, democratic, caring, data rational, evaluative, behavioral, conventional, conceptual, innovative, variety seeking, adaptable, forward thinking, detail conscious, conscientious, rule following, relaxed, worrying, tough minded, optimistic, trusting, emotionally controlled, vigorous competitive, achieving and decisive (Nobre, 2005). These two psychometric exercises are not directly linked to the LADC, but are rather used as aiding tools and measured independently from the LADC. Consequently; assessors identified dimensions on each of the PSYCH contributing to or inhibiting the competencies assessed in the LADC (see in table 1.1 below). Norm scores obtained on the PSYCH on the related dimensions were then adjusted to the 5-point Likert scale used to assess the other measures in the LADC.
- *Practice office nomination (PON)* - Candidates are required to prepare and write a 1 500-word paper/report on any topic of relevance to their career: the company's services, clients, the company's people, the

firm, technical issues or their profession as a whole. The paper is required to be the candidate's own original work. Subsequent to this, candidates are required to prepare a ten-minute presentation in any form using any aids on the same topic chosen for the written report for presentation to demonstrate self-presentation skills.

Table 1.1 below depicts the assessment matrix on which the LADC was based, noting the various competencies assessed in the different exercises.

Table 1.1: Dimensions and LADC exercises

Dimension	Exercise Method					
	SPS	EXC	ALS	CBI	PSYCH	PON
Client orientation			X			X
Applied judgment and insight			X			X
Business acumen			X			X
Knowledge-up-to-datedness			X			X
Practice leadership			X			X
Winning business			X			X
Product and service knowledge			X			X
Creativity and innovation						X
Written communication	X	X				X
Oral communication	X	X				X
Personal impact	X	X				X
Personal integrity						X
Interpersonal skills				X	X	X
Achievement drive				X	X	X
Team Work				X	X	X
Respect for diversity				X	X	X
Negotiation				X	X	X
Resilience				X	X	X
Flexibility				X	X	X
Leadership				X	X	X
General judgement				X	X	X

1.7.2.3 Data gathering

Assessment ratings of all the candidates who participated in LADC for partnership admission purposes were collected from the project coordinator of the firm. The ratings collected spanned from 2006 to 2010. This implies that the data used for this study was secondary data, due to the fact that the LADC was designed for a different purpose than for research.

1.7.2.4 Data analysis

Statistical Programmes for the Social Sciences (SPSS) was used for management and statistical analysis of the research data. Descriptive statistics such as percentages, means and standard deviations were used to describe the AC results. To determine the magnitude of relationships between variables (that is, of dimensions across methods and of dimensions within methods) the Pearson Product Moment Correlation (r) was used. To explore and discover the main dimensions or constructs, the Principal Component Analysis (PCA) was conducted. Principal components that represent interrelations among the variables under study were extracted, and once the total number had been determined, the factors were rotated for interpretation.

1.7.2.5 Formulation of hypothesis

The following hypotheses are formulated for the study:

- H1: There is discriminant validity amongst the dimensions measured in each exercise.
- H2: There are significant correlations to be found in ratings of individual dimensions across exercises in the LADC procedure (convergent validity).
- H3: LADC dimensions can be grouped into a smaller number of dimensions.

1.8 CHAPTER LAYOUT

This chapter has demonstrated a need for continued research of construct validity in ACs. In this study therefore; the following chapters will consist of the literature review on competency assessments and the AC methodology, the scientific study and lastly, the conclusions, limitations and recommendations. The following highlights the remaining chapters of this study.

Chapter 2: Competency –based assessment and AC methodology

Chapter 3: Article: Empirical study in which statistical techniques of analysis of variance, correlation and principal factors will be used to test the hypothesis. Results will be presented in the form of tables and graphs

Chapter 4: Conclusions, limitations and recommendations will be formulated regarding the construct validity of competency dimensions in a managerial assessment centre.

1.9 CONCLUSION

From the literature review it is evident that a great deal of research has been undertaken on the predictive nature of ACs, yet little evidence of the construct validity of this method exists. This chapter therefore, discussed and highlighted the background and motivation to the research, research problem statement, research aims, the theoretical orientation of the research, research design and method, and concluded with the listing of remaining chapters in the dissertation. The next chapter focuses on a literature review, particularly on competency-based assessment and AC methodology.

CHAPTER 2

COMPETENCY ASSESSMENT AND ASSESSMENT CENTRE METHODOLOGY

2.1 INTRODUCTION

The burgeoning demand for the implementation of affirmative action by South African organisations has seen a rise in the identification and development of potential among the Black, Asian and Coloured populations of the country (Charoux, 1991; Magwaza, 1995). This has further given rise to the need for selection procedures, including psychological assessment tools that are culturally fair, appropriate and transparent (Paterson & Uys, 2005; Potgieter & Van der Merwe, 2002). As a result, competency-based assessment (CBA) in South Africa has developed into a strategy aimed at not only integrating various human relations (HR) functions, but also at curbing unfair assessment practices in the workplace (Potgieter & Van der Merwe, 2002). Despite the evolving variety of conceptualisations of competence and competency most practitioners today, prefer the competency-based approach to assessment. CBA is preferred in comparison to the traditional paper and pencil method, because people who are assessed regard it as fair, objective and transparent (Boam & Sparrow, 1992; Saunders, 2000).

ACs are based on CBA principles. Competencies are the basic building blocks of ACs, which measure a number of competencies using various exercises (Potgieter & Van der Merwe, 2002). As such, the AC procedure discussed in this chapter has evolved as a means to promote fair and non-discriminatory assessment in response to the South African legislative framework. In fact, South African labour law requires job-relevant psychological assessment in the workplace, which includes the measurement of job-related competencies (Erasmus & Arumugam, 1997; Foxcroft & Roodt, 2001; Muchinsky et al., 2002).

This chapter aims to conceptualise CBA and AC methodology in the context of psychological assessment in the South African workplace. Firstly, a brief history of assessment in South Africa is provided. The development of the competency approach to assessment as an attempt to curb unfair assessment practices in South Africa and how the approach differs from other methods of assessments will be discussed. The conceptualisation of competencies is then elaborated on, with reference to the definition and the importance of CBA in the workplace. Competency development methodologies and the standards involved in designing the related competencies will also be discussed. Additionally, the manner in which competency frameworks are developed is explained together with the existing managerial competency frameworks. Lastly, the section will conclude with the discussion of the steps involved in developing a CBA battery. AC methodology will be discussed in terms of the definition, characteristics and process followed in the implementation thereof. The chapter is concluded with a chapter summary.

2.2 HISTORY OF PSYCHOLOGICAL TESTING IN SOUTH AFRICA

Prior to the implementation of legislative requirements and strict control of the use of assessments, South Africa was characterised by apartheid policies and practices, as well as the unequal distribution of resources based on racial categorisation into Blacks, Indians, Coloureds and Whites (Foxcroft & Roodt, 2001). During this time psychological assessment practice reflected the racially segregated South African society with tests that were developed for different population groups (Claassen, 1997). For example, tests used for Blacks were designed to investigate their potential for education and occupational fitness, while test developed for Whites were to determine leadership and skilled trade abilities and general classification tests for lower-grade occupations (Biesheuvel, 1959).

Political developments in South Africa led to the economic integration of Black and White employees. These saw a rise in the use of psychometric tests to examine the basic abilities of Black people and compare to that of Whites (Claassen, 1997; Fick, 1929; Taylor, 1990). Taylor and Radford (1986) noted however that even with similar educational levels, different ethnic groups obtain

significantly different scores on the same tests. Moreover, the authors tested a sample of Black and White students enrolled for engineering studies and found that Black students scored significantly lower than their White counterparts (Taylor & Radford, 1986). Researchers (e.g. Fick 1929; Goldstein, Yusko, Braverman, Smith & Chung, 1998; Taylor, 1990; and Taylor & Radford, 1986) raised concerns regarding the fairness or appropriateness of comparing exercise scores of different groups (e.g. Blacks and Whites) using tests standardised for one group only. Taylor and Boeyens (1991) also questioned the integrity of psychological constructs measured regarding their definition and meaning as they are usually understood and interpreted differently in different cultures. Comparability of scores which is related to the fairness of a measure is an investigation of the extent to which a test measures the constructs it claims to measure in the different groups (Taylor & Boeyens, 1991). The above discussion points to the ineffectiveness of testing when there is no evidence that constructs measured really exist in the population groups for which the test has not been standardised.

From the 1960s however, test developers began to realise that it was impossible to ignore cultural influences in test performance, and that culture is a moderator variable, influencing behaviour and constructs being measured (Bedell, 2001). Subsequently, the emphasis shifted towards the consideration of the impact of home environment, schooling and nutrition on test use and development. Additionally, scientific investigation of potential bias and problems associated with testing in a multicultural context also ensued (Van Der Vijver & Rothmann, 2004). Towards the end of apartheid (late eighties) Blacks were generally thought to be under-represented in higher-level jobs and post-secondary jobs (Claassen 1997).

Currently matters pertaining to the use, control and possession of assessment tools are legally controlled through the Employment Equity Act (EEA), Act 55 of 1998, the Labour Relations Act (LRA), Act 66 of 1995, and the Health Professions Act, Act 56 of 1974 (Foxcroft & Roodt, 2001; Mauer, 2000). Both the EEA and LRA deal with matters of individuals' rights and with specific substantive issues, while the Health Professions Act, Act 56 of 1974, regulates training, registration and practice standards for professional psychologists and

the protection of the public against unethical behaviour of professionals. The Health Professions Act attempts to protect the worth and dignity of individuals, irrespective of race, gender, marital status and language, by establishing standards and rules of conduct for practising psychologists. Moreover, the Health Professions Act and Health Professions Council of South Africa address the scope of the profession and practices of psychology, responsibilities and duties. The LRA provides guidelines for fairness in personnel selection decisions which award protection for individuals against unfair discrimination on any uninformed grounds, including sex, race, gender, ethnicity and disability in the workplace (Muchinsky et al., 2002). LRA also, encourages discrimination based only on job competence (Saunders, 2000). The EEA delineates parameters for psychological assessment practice and proposes the implementation of affirmative action measures to redress the past imbalances in employment experienced by previously disadvantaged groups (Muchinsky et al., 2002). Additionally, the EEA promotes the transparent use of psychological assessments and creates awareness regarding the implementation of fair employment practices and procedures. Provision for psychological assessments is made in Chapter 3 of the EEA, which states the following (see Muchinsky et al., 2002):

Psychological testing and other similar assessments of an employee are prohibited unless the test or assessment used:

- 1. Has been scientifically shown to be valid and reliable,*
- 2. Can be applied fairly to all employees, and*
- 3. Is not biased against any employee or group.*

Given the above, from 1994 to date, the focus shifted towards firstly, the development and administration of culturally appropriate assessments, assessment of job-related competence and the need for transparent, objective and practical assessment (Potgieter & Van der Merwe, 2002; Saunders, 2000). Adverse impact which refers to the degree of impact of results of assessment measures on selection decisions that involve different designated groups, has drawn attention amongst assessment practitioners in South Africa (Muchinsky et al., 2002). Adverse impact is shown by a significant difference in averages found between the performance of different ethnic and gender groups on an assessment test (SHL, 2001). When there is lack of evidence of

the validity of such a test, there may be an assumption that the group with lower average performance was being discriminated against directly or indirectly (Muchinsky et al., 2002). Figure 2.1 is a representation of conditions that lead to direct or indirect discrimination and adverse impact.

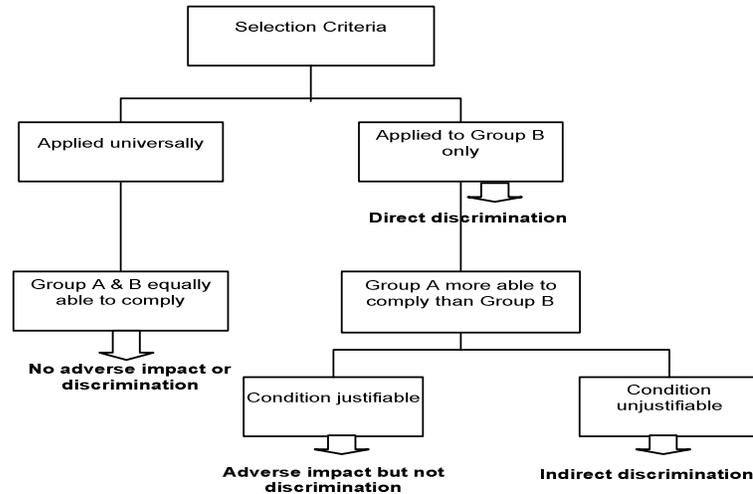


Figure 2.1: Conditions leading to direct and indirect discrimination and adverse impact (adapted from SHL, 2001).

The above discussion highlights the unfairness in testing prior to the implementation of legislative regulations such as the LRA and EEA in South Africa. Such legislation has placed the practice of assessments into the spotlight with regard to the extent of fairness, bias and cultural appropriateness of psychological tests used. Consequently, researchers and practitioners have sourced ways to evade and address the adverse impact that was evident previously in assessment practice, through the use of CBA approaches. The CBA approach thus constitutes the focus of discussion in the following section.

2.3 COMPETENCY-BASED ASSESSMENTS

There has been an increase in competency-based programmes in South Africa in recent years. Some researchers ascribe this increase to the challenges posed by the traditional HR tools such as paper and pencil tests

and other aptitude tests. The argument in literature is centred on paper and pencil tests' inability to predict individual's behaviour as opposed to the predictive value of simulation and work sample tests (Hagan, Konopaske, Bernardin & Tyler, 2006). Although there are similarities between CBA and the traditional paper and pencil tests, CBA differs from the latter with regard to the nature and the process followed, as well as the judgements made which are based on 'competent' or 'not yet competent' (Erasmus & Arumugam, 1997; Fletcher, 1997; Potgieter & Van der Merwe, 2002; Taylor, 1990). Firstly, CBA is aligned to South African legislation (e.g. LRA) in that candidates are assessed against job-relevant competencies, in which predetermined standards are established, as opposed to assessing performance against a selected sample (Fletcher, 1997; Potgieter & Van der Merwe, 2002; Taylor, 1990). Secondly, the approach cuts across all race groups and cultural differences in that people are measured in terms of what they actually do in the job, instead of what they say they do (Erasmus & Arumugam, 1997). Thirdly, CBA focuses on behavior rather than personality or cultural factors, therefore, may be seen as fair by participants. (Potgieter & Van der Merwe, 2002). Lastly, there is a perception that CBA minimises discrimination against candidates with actual job skills rather than education (Potgieter & Van der Merwe, 2002). This is closely linked to the fact that in the competency-based approach, the focus is on outcomes and meeting all set standards and candidates are not compared with other individuals' results or a sample group (Boam & Sparrow, 1992; Fletcher, 1997).

The concept of competency testing instead of testing for intelligence according to Mirabile (1997) was proposed in the early 1970s by McClelland, in an attempt to eliminate bias inherent in intelligence and aptitude tests. In addition, Lawler (1994) specified the need for organisations to move away from the traditional focus on jobs. The author recommended a focus on individual competencies as well as on individuals occupying a job that is characterised by a specific set of activities that can be confined in fixed job descriptions. Spencer and Spencer (1993) found in their earlier studies that aptitude and knowledge tests did not predict job performance and were most often biased against minorities, women and people from lower socioeconomic levels. Similarly, Goldstein, Yusko, and Nicolopoulos (2001) investigated the degree to which job-relevant managerial competencies differ when measured in AC

exercises and found significant mean differences between the subgroups, when using cognitive ability measures rather than those with less cognitive content. These findings imply an indication of a lower job- selection rate, particularly for Black applicants. CBA, as a basic building block of ACs, is an assessment methodology in which job-relevant competencies are identified, defined and measured using a combination of exercises and various assessors to evaluate the performance of the candidates for the job (Potgieter & Van der Merwe, 2002). Given the above, this section conceptualises competency, discusses the manner in which competencies are developed, discusses various competency frameworks that exists and concludes with a process to follow when developing a CBA battery.

2.3.1 Conceptualisation of competencies

The use of competencies and CBA is increasing at a fast pace and various definitions with different meanings exist in literature. That said; the legislative requirements in South Africa have resulted in changes in the development and practice of psychological assessments, which has seen the rise in the use of CBA. The aim of this section is therefore to provide literature on the competency-based approach with specific focus on the definition of competency. Additionally, the importance of CBA not only as means to comply with South African legislation, but it is discussed as a strategy to remain competitive and improve individual performance.

2.3.1.1 Definition of competency

Various definitions of competency exist in literature, denoting different meanings. Kim and Hong (2005) reported that these definitions have developed over the years in addition to those provided by McClelland (1973) who was the pioneer of competency modelling. Before competency is defined it is important to note that there are differences between competence and competency. The difference between the two concepts lies in the job aspects at which a person is competent, as well as the aspects that allow the person to be competent (Heinsman, Hoogh, Koopman, & Van Muijen, 2007; Potgieter & Van der Merwe, 2002; Whiddett & Hollyforde 2003). The focus of competency is on job output, while competence refers to a more abstract concept that can

be measured through behaviour or performance (Heinsman et al., 2007; Potgieter & Van der Merwe, 2002).

In defining competency, Woodruffe (1990) defined it in relation to motives, traits, self-concepts, content knowledge and cognitive behavioural skills and further distinguishes between functional, task and relevant competencies. Furthermore, competence is the underlying characteristics of a person which he/she utilises (Woodruffe, 1990). In comparison to the above definition, Kim and Hong (2005) described competencies as composed of performance, which is associated with clusters of life outcomes. In Burnett and Dutsch (2006) competency is defined as the individual's work-related characteristics such as skills, knowledge, attitudes, beliefs, motives and traits. Similarly, Heinsman et al. (2007) defined competency as a cluster of skills, knowledge, abilities or personality characteristics. According to Mirabile (1997), competency refers to knowledge, skill, ability or characteristics that are closely related to high performance in a specific job. Spencer and Spencer (1993) identified competency as the individual's underlying internal characteristics that result in effective and superior performance. Potgieter and Van der Merwe (2002) explained competency in terms of behaviour patterns that an incumbent brings into a particular position in order to perform the tasks which make the incumbent valuable to both the organisation and its customers. From these definitions, authors therefore seem to agree that competency consists of skills, knowledge, motives, traits, ability and cognitive behavioural skill and that competencies form the underlying characteristics required to perform well in a job context.

Spencer and Spencer (1993) defined competency by using the analogy of an iceberg highlighted in Figure 2.2 below. The authors suggest that competency is composed of motive, traits, self-concept and knowledge and skills. At the top of the iceberg are knowledge and skills, which are visible elements of competence. Both of these elements are surface competencies, which are relatively easy to develop (Juceviciene & Lepaite, 1995). According to Monk (2001) both knowledge and skills are crucial for cost-effective training and excellent job performance in securing such competencies. Motive, traits and self-concept are the most difficult to develop and are referred to as hidden

elements (Juceviciene & Lepaite, 1995). Such elements are central to personality and are more likely to predict a behavioural skill, which in turn predicts job performance (Monk 2001). According to Weiss and Hartle (1997), it is more cost-effective to recruit and select in line with such characteristics. Vathanophas and Thai-Ngam (2007) concluded that visible competencies in the iceberg are the technical and basic competencies required on the job, while hidden competencies are the behavioural competencies that drive the individual's job performance. Training is the most cost-effective way to secure employees' abilities (Spencer & Spencer, 1993). Core competency, according to Spencer and Spencer (1993), refers to those underlying elements of competency that are not easily identifiable or are more difficult to assess and develop, such as personality attributes, behaviour, motives, traits and values.

The iceberg model consists of five types of competency characteristics, i.e. motive, traits, self-concept, knowledge and skills. These are defined as follows (Vathanophas & Thai-Ngam 2007; Monk, 2001):

- Motive: individual's consistent thoughts and wants, which stimulate action
- Traits: responses and physical characteristics that are consistent and assist individuals to respond to situations and information
- Self-concept: individual's self-image, values or attitude
- Knowledge: the contextual or procedural information a person has in a specific area.
- Skills: the physical and mental ability to perform certain tasks.

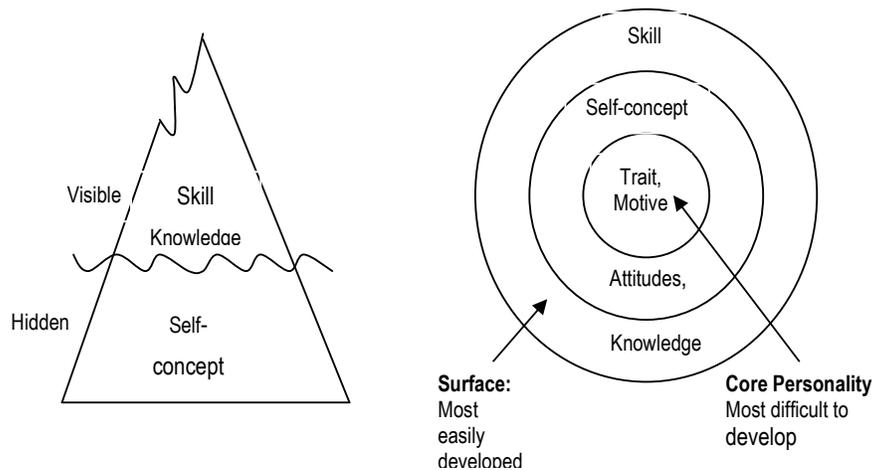


Figure 2.2: The iceberg model and surface and core personality competencies (Spencer & Spencer, 1993, p.11).

The various definitions of competency provided above point to the difficulty in reaching a consensus definition of competency among researchers. Some researchers define competency in relation to the individual's underlying characteristics that are associated with superior performance (Burnett & Dutsch, 2006; Heinsman et al., 2007; Mirabile, 1997), while others define competency in the realm of observable behaviour (Potgieter & Van der Merwe, 2002; Spencer & Spencer, 1993; Woodruffe, 1990). Furthermore, it can be deduced from the above definitions that competency refers to those unique individual characteristics that can be measured and that differentiate between superior and average performers.

The definition of competency from Chen and Naquin (2006) is useful in that it integrates common aspects from the previously stated conceptualisations into one definition. The authors defined competency as a set of work-related characteristics including knowledge, skills, abilities, attitudes, behaviour, traits and organisational capabilities that have a direct link to superior performance, which in turn results in the organisation's sustainable competitive advantage. This definition is further valuable as it views competencies as part of the organisation's strategic function, while taking into account the organisation's

vision and mission as well as its future survival. In the next section the importance of competency-based approach is highlighted.

2.3.1.2 The importance of competency-based- assessment in the workplace

Organisations are going through various changes prompted by among others competition, globalisation, technology, organisational change, structures etc. (Boam & Sparrow 1992; Lawler 1994; Meyer 1995; Potgieter & Van der Merwe, 2002; Theron & Roodt, 2000). According to Lawler (1994), these changes have an impact on the level of performance required to allow organisations to remain competitive and successful. This implies a demand for high performance on the side of the employee and therefore, a requirement for different assessment methods (Theron & Roodt, 2000). In fact, for an organisation to be successful in this changing environment, according to Lawler (1994), the focus should be on individual employees and their competencies rather than the job, specifically for South African organisations, where challenges and changes have been imposed by the labour regulations highlighted in section 2.2 above, in an attempt to eliminate unfair labour practices and ensure equality in the workplace, including the use of psychological assessments (Foxcroft & Roodt, 2001; Potgieter & Van der Merwe, 2002).

In particular, the National Qualifications Framework defines workplace assessment as a process of collecting and interpreting evidence that job performance standards have been met (Meyer, 1996). Therefore, the practice of psychometric assessments in the workplace can be read as unfair labour practice in terms of the LRA. That said, Monk (2001) argued that there is little correlation between aptitude tests and individual performance at work. The author also believed that these tests were biased. She further disputed the accuracy of administering such a test to a real life problem for which the test was not really designed. Lastly, Monk (2001) is held the view that assessing for competency is more efficient in identifying superior performance in specific lines of work, instead of attempting to predict job performance based on intelligence tests.

Moreover, the need to implement affirmative action measures as a result of legislation in South Africa has resulted in increased focus on training management with regard to core competencies and on the upgrading of people as a requisite for global competitiveness, organisational excellence and labour productivity (Mbokazi, Visser & Fourie, 2004; Potgieter & Van der Merwe, 2002). According to Potgieter and Van der Merwe (2002), in order for an organisation to be successful in any subsequent training strategy or intervention, competencies that enable employees to cope with external and internal changes should be identified and developed. Furthermore, it is suggested in Hagan et al. (2006) that a focus on individuals and competencies is more valuable than the evident focus on only jobs.

Industrial psychologists in South Africa are therefore faced with the challenge to provide their clients with job-related and appropriate assessment that is transparent to organisations, unions and employees, in compliance with the above legislation (Potgieter & Van der Merwe, 2002). This has led to a rise in the use of CBA in the workplace (Boam & Sparrow 1992; Potgieter & Van der Merwe, 2002; Theron & Roodt, 2000). Many practitioners prefer CBA because these are job-specific, open and transparent to all those with interest in the process, in addition to being objective and having increased validity (Lawler, 1994; Meyer, 1995; Potgieter & Van der Merwe, 2002; Theron & Roodt, 2000). Potgieter and Van der Merwe (2002) define CBA as an assessment process that evaluates an individual's skill and behaviour relevant to job performance, against the job for which the assessment is conducted. Moreover, by using a competency-based approach, one accumulates sufficient evidence that the individual can perform according to specified standards in a specific role (Fletcher, 1997). Similar to ACs, CBA measures psychological constructs (Potgieter & Van der Merwe, 2002). In addition, assessment information is gathered through systematic observation using assessment methods such as role play, work sample tests and case studies, rather than only paper and pencil tests (Lawler, 1994; Meyer, 1995; Potgieter & Van der Merwe).

To develop and define skills that reflect the organisation's current and future HR needs, CBA has consequently been used in many organisations to address such needs (Lange, Fourie & Vuuren, 2003). According to Burnett and

Dutsch (2006) a competency-based approach to management and leadership training and development enables organisations to become more competitive and results-oriented, providing increased productivity and efficiency. Furthermore, the competency-based approach can be used for selection, career planning, promotion, strategic planning, organisational development, succession planning, compensation, HR planning and performance appraisals (Burnett & Dutsch, 2006).

Despite the obvious value of using CBA, particularly for HR practitioners, some researchers question the feasibility of the approach. Lawler (1994) for example, refers to the costs and time involved, especially on the individual level. Consequently, some organisations would not be able to afford to conduct an exercise where a model could be cross-validated using a criterion sample. In fact, when the performance criteria are not effectively or clearly defined, a competency framework will identify inaccurate competencies (Spencer & Spencer 1993). Obstacles due to resistance to change posed by the implementation of the method are highlighted as one of the problems in Lawler (1994). Hagger, Gonczi and Athanasou (1994) argued that a CBA approach is limited, as it reduces jobs that are complex in nature to a series of discrete tasks where an observer mainly ticks off a checklist of desirable performance. In addition to this, constructs are context-specific, therefore cannot be measured outside a particular context (Hagger et al., 1994).

In summary, the competency-based approach seems to have been introduced as a way to mitigate the risk of unfair labour practice relating to the use of assessments, particularly in South Africa. Furthermore, the approach appears to be closely linked to organisations' strategic attempts to respond to environmental changes by focusing on individual performance rather than job tasks. As beneficial as the approach is to the HR practitioner, there are various disadvantages attached to it. In the following section competency development is discussed.

2.3.2. Competency development

Ulrich (in Potgieter & Van der Merwe, 2002) reports on the increased competitive challenges that have recently been faced by many organisations, including globalisation, profitability through growth, technology, intellectual capital and change. In South Africa and Africa, to deal with such challenges there is a greater need for strategies and interventions such as skills training and upgrading of people in order to remain competitive and productive (Khumalo, 1999). Consequently, the development of competencies that will enable people to cope with the above challenges is crucial in the implementation of such strategies and interventions (Potgieter & Van der Merwe, 2002). This section therefore provides a discussion on how to design competencies and the related competency frameworks. The section concludes with the process followed in development a CBA battery.

2.3.2.1. The design of competencies

Competency development (sometimes referred to as competency modelling/ profiling) is defined as a process in which competencies that represent job proficiency are identified, resulting in a competency framework for a particular job or role in an organisation (Saunders, 2000). The effectiveness of such a competency framework depends on the quality standards and principles applied during competency development. According to Whiddett and Hollyforde (2003), applying such standards ensures that the framework is effective and note that the standards should:

- be clear and easy to understand;
- be relevant to all people who will be affected by the framework;
- take account of expected changes;
- have discrete elements, e.g. behavioural indicators do not overlap;
- contain elements of the same type;
- ensure behaviour that is necessary and appropriate; and
- be fair to all affected by their use.

In addition to the above standards, Van Rooyen (2008) recommended that the competency model must be related to the job in question as well as the

organisation. Moreover, the author suggested the inclusion of only eight competencies that need to be measured. In developing a competency framework, job analysis plays an important role in terms of gathering relevant job information (Monk, 2001).

Job analysis differs from competency development in that it represents a bottom-up approach, while competency development reflects a top-down approach (Thornton & Rupp, 2006). Thus, instead of starting with the job itself as a point of analysis, it is the overall organisation that is considered first, including its vision, mission, strategy, culture and values (Schippmann et al., 2000). Competency modelling, according to Thornton and Rupp (2006), starts with the analyst interviewing the top management team to understand the broader context of the organisation, while the job analyst seeks to establish individual level competencies.

Monk (2001) similarly pointed out that job analysis and competency modelling differ in terms of purpose, focus, level of detail, face validity, orientation to time and focus on personality and values. While the purpose of job analysis is based more on selection and performance appraisals, competency modelling is used as the basis for training and development. Job analysis focuses on the work, including specific jobs and the job characteristics that differentiate between jobs, while competency development is worker-focused and includes individual-level competencies that cut across occupational categories. The former is highly detailed, while the latter is broader. With regard to time span, job analysis is more short-term based while competency has a long-term focus on organisational fit. Competency development methods produce results that take into account the culture of the organisation, whereas job analysis shows little consideration for the organisation's culture.

Various methods are used to identify competencies and to design an appropriate competency model. Monk (2001) listed various competency development methods used by organisations such as the job competence assessment model, modified job competence assessment method, customised generic model method, flexible job competency model method and the generic

model overlay method. The first two methods are reasonably straightforward: average and excellent behaviours are identified, resulting in a competency framework. The last three methods seem to be an involved process, as organisations review their strategies in order to align competency requirements. Other approaches to the development of competencies, according to Spencer and Spencer (1993), include the use of criterion samples, expert panels and studies of single incumbent and future jobs where there are not enough jobholders to offer samples of superior and average performance.

It is clear from the above discussion that CBA is a scientific and standardised method of assessment in which guidelines are documented to guide the design of competencies. Additionally, various tools to gather information about such competencies exist, to ensure relevance to specific organisations. Once competencies have been identified through methods discussed above a competency framework can be developed. Therefore, the focus of the next section is to discuss the development of such competency frameworks and existing managerial competency frameworks.

2.3.2.2 Development of competency frameworks

Before the 1960s, the success of an organisation was primarily reliant on specific managerial techniques, but because of the internal and external challenges that have emerged over the years, a need for new managerial and leadership competencies arose (Naquin & Holton, 2006). Furthermore, traditional measures that sustained organisational viability have changed and organisations are currently faced with pressures related to efficiency, being results-driven and increased competition (Melancon & Williams, 2006). Consequently, an interest in developing and redesigning management development strategies and systems that are aligned to competencies has become an essential focus in many organisations. According to Jubb and Robotham (1997), such a focus is driven by a need for effective management performance that will result in potential benefit for overall organisational performance. This trend has emanated from the need to entrust managers with the responsibility to transform their organisations to deal with changing circumstances, including increased competition (Mbokazi et al., 2004).

Therefore, it is important to identify the development needs and potential of managers and evaluate them against a clearly defined criterion of performance. According to Cockerill, Hunt and Schroder (1995), a competency-based approach to management development is ideal for such applications.

Various efforts have been made to identify core managerial competencies and their implications across management education and development, in order to accelerate the development of competent managers (Mbokazi et al., 2004). Generic management competency models have been researched and implemented (e.g. Bartram, 2005; Cockerill et al., 1995; Schroder, 1989; SHL, 2007; Spencer & Spencer, 1993). Once competencies have been identified for a managerial job by using any of the methods mentioned in section 2.3.2.1 above, they must be categorised into a logical structure (Monk, 2001). Furthermore, these competencies must be presented in clusters (Spencer & Spencer, 1993). According to Mirabile (1997) a cluster refers to a grouping of competencies, which is organised for the purpose of simplification. For example, problem solving and problem resolution may form a competency cluster with relevant behaviours that describe the cluster for a particular job group. Under this cluster, for example, analytical thinking, judgment and decision-making can be grouped. That said, Spencer and Spencer (1993) categorised competencies according to the job performance criterion that they claim to predict; these categories are threshold and differentiating competencies, which are explained below.

Threshold competencies: These refer to crucial attributes that a person must have in order to perform effectively, within minimum standards in a particular job (Spencer & Spencer, 1993).

Differentiating competencies (organisation-specific competencies): In contrast to threshold competencies, differentiating competencies provide a distinction between superior performers and average performers. These competencies are further categorised into generic and specific competencies (Monk, 2001). On the one hand, generic competencies are universally applied across organisations and on the other hand, organisation-specific competencies refer to what an organisation is best at and are not behaviour-specific

competencies, but rather technical in nature (Monk, 2001; Whiddett & Hollyforde, 2003).

Increased consensus among researchers has emerged, suggesting a universal framework of managerial competencies (Bartram, 2005; Cockerill et al., 1995; SHL, 1993; Spencer & Spencer, 1993). The authors seem to share consensus particularly on competencies such as leadership and managing others, interpersonal relationships and building relationships, teambuilding, analytical skills and the development of self and others. Another competency that seems to be commonly studied by researchers is achievement orientation (Cockerill et al., 1995; Spencer & Spencer, 1993). The various managerial competency models are highlighted in Table 2. 1 below.

Spencer and Spencer's (1993) model of managerial generic competencies specifically distinguishes high performers from average performers based on the competencies in Table 2.1 below. Their model was developed according to 36 managerial competencies, covering a wide range of levels of functioning (Spencer & Spencer, 1993). Similarly, Cockerill et al. (1995) designed a high performance managerial competencies model (HPMC), consisting of 11 competencies (see Table 2.1). The HPMC is also used to differentiate between high and average performers. Furthermore, the HPMC is beneficial to a more dynamic environment rather than a stable organisational environment. This is due to the fact that in a stable environment, HPMCs do not have predictive value (Cockerill et al., 1995). The authors suggest that firstly, in a dynamic environment the HPMC constitutes managerial behaviours which are critical to organisational performance. Secondly, the model can be used as a prerequisite for the transformation of bureaucracies to fit into dynamic environments.

Competencies in the SHL's IMC model are grouped according to three clusters, as shown in Table 2.1. Competencies one to four in the table are referred to as managerial qualities, while competencies 5 to 8 fall under professional qualities. Competencies 9 to 12 are grouped under

entrepreneurial qualities, while the last four competencies, 12 to 16, fall under personal qualities. Bartram’s (2005) competency model, referred to as the great eight competencies (see Table 2.1) is an articulation of work performance that is consistent with other models used by practitioners in competency research. In addition, the model provides understanding of job performance that can be applied consistently across jobs, measurement tools and cultural contexts.

Table 2.1: Competency models containing key managerial competencies

Spencer & Spencer (1993)	SHL, 1993	Cockerill, Hunt & Schroder (1995)	Bartram(2005)
1. Impact and influence	1. Leadership	1. Managing interaction	1. Leading and deciding
2. Achievement orientation	2. Planning & Organising	2. Interpersonal search	2. Supporting and cooperating
3. Teamwork and cooperation	3. Quality Orientation	3. Development orientation	3. Interacting and presenting
4. analytical thinking	4. Persuasiveness	5. Presentation	4. Analysing and interpreting
5. Initiative	5. Specialist Knowledge	6. Conceptual flexibility	5. Creating and conceptualising
6. Developing others	6. Problem Solving & Analysis	7. Self-confidence	6. Organising and executing
7. Self confidence	7. Oral Communication	8. Information search	7. Adapting and coping
8. Directiveness/assesstiveness	8. Written Communication	9. Achievement orientation	8. Enterprising and performing
9. Information seeking	9. Commercial Awareness		
10. Team leadership	10. Creativity & Innovation		
11. Conceptual thinking	11. Action Orientation		
	12. Strategic		
	13. Interpersonal Sensitivity		
	14. Flexibility		
	15. Resilience		

It is evident from the above discussion that through the use of the competency approach to leadership and management development, organisations are able to remain competitive and results-driven, leading to increased productivity and efficiency. Although various universal managerial competency frameworks exist, it might be more desirable for organisations to develop frameworks that take into account their unique culture, vision, mission and strategic direction, instead of customising an already existing framework.

2.3.2.3 Development of the competency- based- assessment battery

Once competencies have been identified from using any of the methods stated above and a competency framework developed, a battery to measure such competencies can be defined. Typically, in the CBA approach various competencies are measured using various simulation and psychometric exercises. Once the battery has been developed a method of administering such a battery (e.g assessment centre) can be designed. The following section therefore, presents steps involved in the development of a CBA battery. Following this process ensures an objective and fair competency assessment process. Before competency identification begins, the organisation needs to identify jobs to be studied, which have high value in relation to the organisation's strategic intent and structure (Spencer & Spencer, 1993). Potgieter and Van der Merwe (2002) identified the following steps in the development of CBA:

Step 1: Planning

This phase includes information-gathering, including the definition of a job family in which the assessment is conducted, in order to ensure that job performance factors are included in the competency model. During this phase the assessment battery is also defined, as well as the purpose of assessments (e.g. development, performance appraisal or selection). Because the CBA method is time-consuming, its relevance, advantages or disadvantages should be considered in this phase. In addition, the organisation's stakeholders' needs and the parties affected by the process are identified. These may include unions, management, selectors, customers and shareholders (Saunders, 2000). This ensures that the assessment battery is linked to the needs of the client and to transparency and openness in the model (Potgieter & Van der Merwe, 2002; Saunders, 2000). Whiddett and Hollyforde (2003) suggested that even when external consultants are used, it is important that they are seen as supporting the project and not leading the project themselves. It is crucial to consider getting buy-in from the key people, which implies clarification of purpose and planning.

Step 2: Establish competencies

This step involves the identification of competencies required for successful performance in the job, using the job analysis method. This ensures accurate reflection of the components of successful job performance. Job analysis can be defined as a systematic procedure used to investigate the tasks, duties and responsibility of a job in an organisation (Carrel et al., 1996). Furthermore, the authors claim that the investigation includes the level of decision-making by the employees in a particular job category, skills required by employees to perform their jobs adequately and the mental efforts required to perform the job.

According to Cascio (1998) job analysis information gives organisations insight into their needs related to selection problems. Job analysis information can be gathered using questionnaires, individual or group interviews, diaries, observations and subject matter experts (Van Rooyen, 2008). According to Monk (2001) the most commonly used methods to analyse jobs, particularly in South African organisations, are the Functional Job Analysis Questionnaire (FJA), Position Analysis Questionnaire (PAQ), Common-metric questionnaire (CMQ), Critical Incident Technique (CIT), Repertory Grid and the Work Profiling System (WPS).

Once the competencies have been identified, a qualitative validation and adjustment of the competency profile may be required, through the use of focus groups or questionnaires (Potgieter & Van der Merwe, 2002). Moreover, the identified competencies can then be grouped into competency clusters, summarising the essence of competencies in a particular cluster (Potgieter & Van der Merwe, 2002).

Step 3: Develop assessment battery

This step includes deciding on the relevant exercises to measure the identified competencies (Potgieter & Van der Merwe, 2002). Current or newly designed measurement methods (e.g. role play, case study, psychometric tests, etc.) are decided upon (Saunders 2000). Potgieter and Van der Merwe (2002) noted that assessment tools to be used must elicit behaviours that are related

to the core competencies identified to ensure that the battery is relevant and job-related. Moreover, during this phase the rating criteria must be established. Potgieter and Van der Merwe (2002) suggested that a pilot study should be conducted so that adjustments or revision of competencies occur when necessary, which must be integrated into the competency-based tool.

Step 4: Implement and evaluate the battery

This is a roll-out phase in CBA, which involves the application of assessment of competence throughout the organisation, with the resultant database of current skills levels in the organisation (Potgieter & Van der Merwe, 2002; Saunders, 2000; Whiddett & Hollyforde, 2003). An important element in this phase is the training of assessors and ensuring compliance with the guidelines regulating the use of tests in the workplace, as stipulated in the Health Professions Act of 1974 (Saunders, 2000; Potgieter & Van der Merwe, 2002).

Validation of the assessment battery is important in this phase in terms of construct and criterion validity, reliability and cultural bias towards designated groups (Potgieter & Van der Merwe, 2002; Saunders, 2000; Whiddett & Hollyforde, 2003). According to Potgieter and Van der Merwe (2002), battery validation is important, as it ensures that the battery meets its original intended purpose within the organisation.

In summary, literature reviewed in the practice of competencies and CBA, with the objective of ensuring transparent and culturally fair assessment methods, highlighted developments in the practice of psychological assessments prior to 1994. Literature demonstrates that psychological testing in South Africa was characterised by unfair practices, with tests being used in different populations (e.g. Blacks) for which they were not originally standardised. As such, means to mitigate unfair labour practices and unfairness in psychological testing practices were implemented through legislative frameworks such as the EEA and LRA. Specifically, the EEA advocates the use of tests that are fair, unbiased against any group and proved to be valid and reliable. Literature reviewed also indicated the utility of a competency based approach to assessment in aligning assessment practices to LRA and EEA fairness

requirements. AC methodology is based on CBA principles and is discussed in the following section.

2.4 ASSESSMENT CENTRE METHODOLOGY

The legislative framework discussed in section 2.2 above advocates for the use of reliable, objective, valid, fair and assessment of job related competencies. This has resulted in many organisations using CBA as a strategic intent. As a result the inherent rise in the use of assessment centre methodology has ensued as part of implementing the CBA approach. The use of ACs is particularly relevant in South Africa in that changes in legislation have resulted in the need to identify potential amongst the previously disadvantaged employees. Therefore, this type of assessment does not look at the candidate's current performance but rather the potential demonstrated by such candidates (Charoux, 1991). The aim of this section is therefore to discuss literature on AC methodology. Firstly, ACs will be conceptualised in terms of the definition, their unique characteristics, the uses of ACs in industry and the process followed when designing and AC. Secondly, the utility of ACs will be discussed in relation to the predictive nature of ACs. This will be followed by an examination of construct validity of, including issues that have an impact on such validity. This includes a review of available literature using various methods, such as factor analysis, MTMM and correlation studies.

2.4.1 Conceptualisation of assessment centre methodology

The current emphasis of psychological assessment practice as a means of curbing unfairness in assessments has been on using culturally appropriate tests, objective discrimination based on job competence and transparent testing amongst others. Given the legislation governing the practice of psychological assessments in South Africa (discussed in section 2.2) above, the AC methodology has grown in use, partly as a response to the requirements implied by such legislation. Against this background, ACs seem to have an ability to compare individuals fairly and accurately, even in a diverse society, such as South Africa (Charoux, 1991). Moreover, because

they simulate real life situations using a variety of exercises, candidates have a better opportunity to demonstrate their performance. ACs have been in use for more than 40 years and the use of the procedure is growing in industry for purposes such as selection, training and development and promotion (Lance, Foster, Nemeth, Gentry & Drollinger, 2007). The reasons for the continued use of ACs can be ascribed to their ability pertaining to content, face and predictive validities and perceived fairness (Lance et al., 2007). In fact, a lot of research has been conducted on the predictive nature of ACs, with findings supporting ACs' validity to predict future job performance (Arthur et al., 2003, Gaugler et al., 1987; Jansen & Stoop, 2001).

This section therefore, discusses characteristics that differentiate ACs from other forms of assessments, such as the traditional paper-and pencil tests or the psychometric tests. To achieve the above, it is important to provide a definition and the characteristics that make ACs unique. Furthermore, the uses of ACs are discussed. The section will conclude on how an AC is designed.

2.4.1.1 Definition of assessment centre

Before an AC is defined, it is important to note that an AC does not necessarily refer to a place or venue; instead it refers to a concept related to an approach or a method (Woodruffe, 1990). An AC is defined as a standardised method for evaluating behaviours comprising inputs that are observed by a number of trained assessors (Joiner, 2000; Thornton & Rupp, 2006).

2.4.1.2 Characteristics of assessment centres

ACs consists of a variety of assessment tools that are combined to assess job-related performance of assessees in a variety of exercises that are based on the targeted job (Dayan, Fox & Kasten, 2002; Woodruffe, 1990). Exercises include simulations, interviews and ability tests; etc. (Joiner, 2000). Simulation exercises must be designed in a manner that resembles a real-life situation (Heindl, 1997). Thus, ACs are designed to measure specific competency dimensions related to job performance across situations or exercises.

According to Thornton and Rupp (2006) this method allows for the measurement of multiple competencies that are otherwise difficult to measure.

Joiner (2000) indicated that in order to establish the job-related behaviours stated above, a thorough job analysis must be conducted. Job analysis is common practice in ACs to identify and determine the dimensions, characteristics and attributes required for job performance success. Thus, job analysis methodology forms the basis for establishing dimensions that are often conceptualised as equivalent to competencies that are to be measured in an AC (Chen, 2006; Joiner, 2002; Lievens, 1999).

To observe and evaluate each participant, several assessors must be used, who participate in the AC simultaneously (Woodruffe, 1990). Assessor training is a crucial part of the AC technique (Joiner, 2000). In order to avoid errors, Cadwell, Thornton and Gruys (2003) suggested that assessors should receive training with regard to the process of observing and classifying dimensions and behaviours, as well as demonstrate skill in order to evaluate assessee's performance properly. The assessment information obtained from different assessors and different assessment methods is integrated and an overall judgment of each participant is formed, usually during a wash-up session (Heindl, 1997; Thornton & Rupp, 2006). In summary, the International Task Force on Assessment Centre Guidelines has identified the following ten features of ACs (Joiner, 2000). They:

- conduct a job analysis of relevant behaviours;
- classify behaviours into meaningful and relevant dimensions or competencies;
- establish a link between the classified dimensions or competencies and assessment techniques;
- conduct multiple assessments;
- develop and implement job-related simulations to elicit behaviours related to classified dimensions or competencies;
- use multiple assessors to observe and evaluate each participant;
- train assessors on performance standards;

- record specific behaviour observations;
- report observations made during each exercise before the integration discussion; and
- integrate information through a statistical integration process validated in accordance with professionally accepted standards.

Given the distinct features of ACs discussed above, it is clear what ACs really are and what activities distinguishes them from other standardised methods of assessments. The next section highlights a variety of uses of ACs.

2.4.1.3 The uses of assessment centres

ACs emerged during World War 1, when the German government used them to select capable military leaders in the 1930s (Chen, 2006; Thornton & Rupp, 2006; Woodruffe, 1990). Over the past years, however, the use of ACs has escalated and currently the technique is being used in both the public and private sectors. Lievens (1999) confirmed that the technique has become popular, particularly in managerial selection.

Moreover, various changes in the world of work, such as the growing job market, changes in the composition of labour markets (e.g. more women entering the workplace) and changes in the nature of jobs (e.g. high performance work systems, knowledge workers and the new employment contract) have contributed to the growing use of ACs (Thornton & Rupp 2006). Consequently, these changes call for innovation in AC technology development practices. Despite the costs of setting up an AC, the use of this method is growing tremendously, especially for selection, placement and the identification of management potential and training (Gaugler et al., 1987). In a study conducted by Spsychalski et al. (1997) it was found that the most popular reasons for the use of ACs were selection (50%) and promotion (45.8%). Development planning (39.2%) and training (34.4%) also constituted some of the main reasons cited in the study. According to Arthur and Woehr (2003), another reason for the escalation of the use of ACs, is their ability to predict future job performance. Furthermore, Bartram (2004) claimed that

organisations use ACs for tactical and strategic reasons in an attempt to increase their effectiveness.

Notwithstanding, the need for Black empowerment and Leadership development in South Africa has resulted in an inherent rise in the use of ACs in the country (Charoux, 1991). According to Saunders (2000) this is because of their ability to simulate a real-life situation as well as the practical appeal to line and staff managers. Moreover, in ACs several assessors who have been trained are used, they appear to reduce the risk of bias (Charoux, 1991) and apart from the fact that the procedure is regarded as fair by the participants, reduced adverse impact has also been noted by some studies (Thornton & Rupp, 2006). Cascio and Aguinis (2005), cited in Dean, Roth & Bobko (2008), suggested that AC research has demonstrated less adverse impact compared to cognitive ability tests and interviews.

In summary, ACs as a unique form of assessment was initially used in the military context; however, with the latest developments and legislative challenges, more innovative ways of assessments became a requirement, such as the ACs. As such ACs are used in many sectors for reasons such selection and development. In the next section the steps involved in designing an AC are highlighted.

2.4.1.4 Designing an assessment centre

For the purpose of this study the integrated AC design model of Schlebusch and Roodt (2008) will be applied. According to the authors, the model is similar to that applied in a training context and consists of four stages depicted in Figure 2.3 below. The stages are the following (Schlebusch & Roodt, 2008):

Stage 1 Analysis: This stage is conducted in order to analyse the task or job and the organisational context. This ensures that the design team and the organisation are clear about what is required in terms of competencies. The purpose of this stage is to develop a systematic understanding of the organisational context and of the organisation, its dynamics and its specifics, as well as to provide clarity about the job/role and its demands. Once the

analysis has been conducted, competency profiles and elements of the validation criteria can be delivered.

Stage 2 Design: During this stage a programme is designed to ensure that the objects of the programme are achieved. This stage results in an AC manual and a technical AC manual. This is done in order to deliver a functional AC. During this stage the following is achieved:

- Translation of information from the analysis stage into an operational AC blueprint that guides simulation and AC design.
- Simulations and support documentation that allows observers to make a valid and reliable assessment of the behaviour (see section 2.3.2.3 for the design of a simulation battery).
- An AC process that effectively combines the various simulations.
- An AC that is ready for implementation.

Stage 3 Implementation: During this stage the programme that was designed is implemented and put into practice. This includes handing over the process to the process owners. This stage results in a pool of readily available, competent process owners, an AC practice that adheres to the AC manual, and a post-AC process that ensures positive action and a sustained AC process. The intended purpose here is to operationalise the AC, including achieving part of the target organisation's culture as well as an AC system, process and practice that ensure a reliable and valid AC.

Stage 4 Evaluation: The focus during this last stage is on evaluation and validation, which includes two steps, namely content evaluation and the reliability and validity evaluation steps. During this stage a content-valid, reliable and valid programme process is produced.

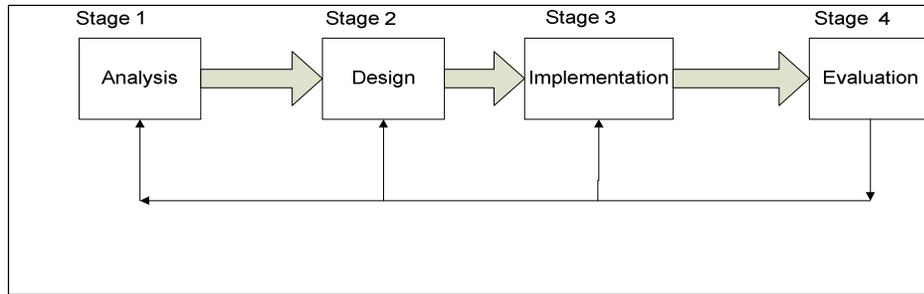


Figure 2.3: The basic design model (adapted from Schlebusch & Roodt, 2008).

The sections above conceptualised factors that may have an impact on the construct validity of ACs. Thus, conforming in the manner discussed in the above section when using the AC approach may facilitate the validity of ACs. Concerns have been mentioned in literature regarding the increase in the misuse of the AC process, which lies in the validity area. Some researchers ascribe this to misunderstanding and lack of clarity about the guidelines and ethical considerations on the side of AC users (Cadwell et al., 2003). The composition and the structures discussed above indicate what AC activities are and what does not form part of these activities.

Despite this information, a number of studies have revealed poor construct validity evidence of ACs. ACs are used widely because of their ability to predict some future job performance (Schmidt & Hunter 1998). However, there is a dilemma among researchers regarding AC construct validity, which is discussed in the section that follows. This dilemma seems to concern the methodological factors discussed above, including AC dimensions, exercises, assessor characteristics and training, observation and evaluation rating, the integration approach, the design, and the purpose of ACs as outlined in the AC guidelines. The aforementioned methodological factors lead to inaccuracies and bias on the side of the assessor, resulting in ACs' poor convergent and discriminant validity.

The purpose for which an AC is designed may have an impact on the construct validity of the AC. Arthur and Woehr (2003) hypothesised that the

purpose of the AC has a relationship with construct validity in that dimension-convergent validity will be higher for ACs conducted for training and developmental purposes than those designed for selection or promotion purposes, and discriminant validity evidence will be lower for ACs conducted for training and developmental purposes compared to those conducted for selection and promotion.

ACs, according to Joiner (2000) are meant to measure a number of dimensions using various exercises. Moreover; according to Howard (1997) ACs with a larger variety of exercises produce more validity evidence. In addition, in designing ACs, failure to identify stable dimensions may lead to poor construct validity, which has implications when the AC is used for developmental purposes (Lievens, 2001). When designing the centre, for example, one has to ensure that all characteristics match the purpose for which it was designed. The number of dimensions measured in an AC may act as a moderating factor in AC construct validity. One may therefore conclude that the number of dimensions measured in an AC may have an impact on the construct validity of the AC. In fact, the recommended number of dimensions to measure is eight, as indicated in section 2.3.2.1 above.

The quality of AC depends on the quality of the assessors. Joiner (2000) emphasises basic essential elements for ACs to be regarded as ACs, among others assessor training in observations, classification of behaviours into competencies and rating errors that may occur during ratings. Thus, assessors observe and assign ratings that are supposed to reflect performance on certain job-related dimensions. This indicates that assessor training is a crucial component in providing a shared understanding of the dimensions measured among assessors to improve the reliability and construct validity of ACs. The guidelines suggest intensive training prior to participating in ACs and this should focus on the quality rather than the length of training. A number of studies have attempted to examine the impact of using psychologists as assessors instead of managers (Jackson, Stillman & Atkins, 2005; Lievens, 1998; Lievens & Klimoski, 2001). In most cases, construct validity improved when using psychologists. Moreover, the rating approach used may influence convergent and discriminant validity. For example, Arthur and Woehr (2003)

reported that the across-exercise approach results in better evidence of construct validity than the within-exercise approach.

In summary, although the use of ACs is popular, the above points to some gaps between actual practice and the stipulated guidelines, resulting in poor construct validity of ACs. The following section discusses the utility of ACs with specific focus on the predictive validity of ACs as well as the findings and implications of construct validity research studies.

2.4.2 The utility of assessment centers

It was stated earlier in this study that ACs have been in use for a number of years and are particularly used for selection, promotion and development amongst others. Some researchers ascribe this increase to the methods ability to predict future job performance, content and face validities as well as perceived fairness. Although there a number of studies that support the predictive validity of ACs, many studies question the construct validity of ACs. The focus of this section is therefore to discuss the utility of ACs in terms of firstly, the predictive validity of the method. Secondly, to discuss literature on construct validity issues of ACs, including methods used to study the construct validity. The section will conclude with the discussion of factors that impact on construct validity of ACs.

2.4.2.1 Predictive validity in assessment centres

A large proportion of research reflected in literature supports the predictive validity of ACs. Predictive validity is the extent to which a measure is able to predict job performance as well as potential job performance (Jansen & Stoop, 2001). Predictive validity of ACs is determined by correlating predictor scores (e.g overall assessment ratings) gathered before the hiring of a person into the job, with criterion measures obtained sometime after hiring for a representative sample (Jansen & Stoop, 2001). Additionally, Schmidt and Hunter (1998) claimed that the most important value of any measure in the selection process lies within its ability to predict some future job performance. The vast amount of literature that exist regarding predictive validity evidence of ACs has been

shown using advancement criteria (e.g. career progress, salary advancement, long-term promotion, as well as potential development) and job performance criteria (Arthur et al., 2003; Gaugler et al., 1987; Hermelin, Lievens & Robertson, 2007; Hunter & Hunter, 1984; Jansen & Stoop, 2001; Thornton & Rupp, 2006). In fact, Jansen and Stoop (2001) are of the opinion that ACs predictive validity still outperforms other assessment instruments, with the exception of mental ability tests.

The studies of Gaugler et al. (1987), Hermelin et al. (2007) and Hunter and Hunter, (1984) focused on the overall assessment rating (OAR) of ACs to predict future job performance, and the results showed true validity estimates ranging from .28 (Hermelin et al., 2007), .37 (Gaugler et al., 1987), and .43. (Hunter & Hunter, 1984). Although Hermelin et al. (2007) found lower validity estimates than his counterparts, the above provides some evidence of ACs to predict future job performance. With regards to managerial performance, ACs have been reported to predict such criteria. For example Gaugler et al., (1987) in their meta analysis study, found validity estimates of .53, whilst .63 was found in Hunter and Hunter (1984). Similarly, Thornton and Rupp (2006) reported on a management progression by The American Telephone and Telegraph (AT&T), in which 42% of the sample predicted to advance to middle management level, had actually done so, indicating that the OAR were able to predict management potential. In terms of promotion and salary advancement criterion, Gaugler et al. (1987) reported .30 validity results for promotion and Jansen and Stoop (2001) reported .39 validity estimates for salary progression. Moreover, in Arthur et al. (2003) meta analysis, validity estimates between .25 and .39 were found for dimensions such as communication, problem solving, planning and organising, drive, influencing others and consideration for others dimension.

There have been some studies in South Africa that also focused on the predictive validity of ACs (e.g Heindl 1997; Olivier 1994; Spangenberg, Esterhuysen, Briedenhann & Calitz, 1989). Olivier (1994) found no evidence of a Middle Management Assessment Centre (MMCA) to predict entry into Senior Management Training Course. Heindl (1997) found modest mean predictive validity for both the trait ratings and the task ratings to predict future job

performance. More positive predictive validity results of ACs were found however in Spangenberg et al. (1989). In this study BARS were correlated with the AC ratings and a multiple correlation of .37 was obtained. Additionally, Britz (in Heindl, 1997) reported correlations of between .556 and .625 when validating an AC against five Management-By-Objective criteria.

In conclusion, the majority of available research suggest conclusive evidence that ACs are valid predictors of some form of criteria of interest such as managerial potential, career advancement, future job performance, salary progression and promotion amongst others. The next section focuses on issues that impact on the construct validity of ACs.

2.4.2.2 Construct validity issues in assessment centres

The predictive nature of ACs is reported in literature to be the major reason for the popularity of ACs in industry today. However, some research have contended ACs' ability to measure the constructs that they purport to measure, thus, construct validity. Furthermore, research in this regard argues the importance of understanding whether ACs work and whether they measure what they truly claim to measure (Arthur et al., 2008; Meriac et al., 2008).

A method widely used to establish construct validity of ACs has been the MTMM matrix, as recommended by Campbell and Fiske (1959). The focus of this method is on measuring several traits using multiple methods. Applied in the context of the AC process, dimensions are seen as traits, while methods are viewed as exercises (Arthur et al., 2008; Greyling et al., 2003). In terms of construct validity, the MTMM places emphasis on the need for strong correlations that are consistent between ratings of the same dimensions across exercises and low correlations between two different dimensions measured in the same exercise (Lievens & Conway, 2001).

When correlations between the same dimensions measured in two different exercises are strong, convergent validity is shown, while weak correlations

between two dimensions measured in the same exercise demonstrates strong discriminant validity (Greyling et al., 2003; Meriac et al., 2008). Thus, a weaker correlation (i.e. low r) between different dimensions within the same exercise demonstrates high discriminant validity and strong correlation (i.e. high r) between the same dimensions from across exercises demonstrates high convergent validity. Although findings demonstrating convergent validity have been shown in literature, little support has been shown for discriminant validity (Thornton & Rupp, 2006). In a study conducted by Kolk, Born and Van der Flier (2004a) convergent validity was found to have improved from .35 to .67. In addition, Lievens (2008) reported on studies that used the MTMM approach, that found consistently that ratings of the same dimensions from across exercises correlated highly (i.e. good convergent validity), while ratings of different dimensions within the same exercise had strong correlations (i.e. weak discriminant validity). The fact that construct validity of ACs is the main focus of this dissertation permits the following further explication of research explaining AC construct validity.

2.4.3 Research: AC construct validity

The aim of this section is to discuss issues that contribute to the lack of construct validity in ACs. The discussion will begin by discussing various methods that have been used to study the construct validity and conclude with factors that contribute to the lack of construct validity including the evidential and conceptual factors.

2.4.3.1 Correlation studies

Various methodological factors that may moderate construct validity findings of ACs have been reported in literature, such as the rating approach used by assessors to evaluate candidates' performance (see Arthur et al., 2008; Lievens & Conway, 2001; Meriac et al., 2008). There are two primary rating approaches namely; the within- exercise approach (WER) and the across-exercise approach (AER). Whilst the WER, assessees are rated on each dimension after the completion of each exercise, AER evaluation of similar dimensions only occurs after the completion of all exercises, when assessors

take a consensus approach based on ratings of performance from all exercises (Arthur et al., 2008; Lievens & Conway, 2001; Meriac et al., 2008). It is reported that lack of discriminant validity is enhanced when ratings are concluded after each exercise (WER) compared to the integration process that occurs after all the exercises have been completed (AER) (Robie et al., 2000). This implies that, using the AER approach results in a better distinction of factors being measured. In fact, according to Howard (1997), the WER method creates problems with regard to discriminant validity results. Howard (1997) held the view that assessors are most likely to think about the similarities in behavior across exercises when they wait until the end of the assessment centre to rate dimensions across different exercises. Regardless of the theoretical assumptions that AC dimensions within exercises should be separate and converge across different exercises; research on construct validity of ACs has shown continuous evidence that correlations between different dimensions within same exercise rated WER are much higher (weak discriminant validity) than correlations between different dimensions within the same exercises rated AER (thus higher discriminant validity (Heindl, 1997)). Thus, it is evident that there are more similarities among dimensions rated in WER and a smaller relationship between those obtained in the AER. Sackett and Tuzinski (in Lievens & Klimoski, 2001), for example, found that in an AC, the average correlation between dimensions AER is .25, while that of dimensions in WER is .58.

Another study by Donahue, Donald, Truxillo and Gerrity (1997) investigated the effect of using behavioural checklists versus using graphic rating scales on the construct validity of ACs using MTMM and confirmatory factor analysis, finding lack of convergent validity for both scales. The findings, however, showed an improvement in discriminant validity when using a behavioural checklist over graphic rating scales. Kudisch et al. (1997), in their findings on the study of construct validity of a diagnostic AC, reported a mean correlation AER was .29, while that of dimension ratings WER was .41. This study showed support for lack of construct validity in the form of discriminant validity. Kleinmann & Koller (1997) also found similar results in which the correlation AER (.31) was lower than WER (.91).

In another study by Fleenor (1996) a mean correlation of .42 for WER was reported and a mean correlation of .22 for AER. These findings indicate poor discriminant validity in ACs when using both rating approaches, although it is slightly better in AER. In addition, Greyling et al. (2003), when investigating the construct validity of ACs, found no evidence of discriminant validity and very little convergent validity based on only two dimensions (i.e. motivating others and analytical thinking). Table 2.2 below shows results of previous research on correlation studies regarding the construct validity of ACs, specifically discriminant validity.

Table 2.2: Previous research of discriminant validity WER vs AER

Authors	Correlations	
	Within Exercise Ratings	Across Exercise Ratings
Donahue et al. (1997)	.24/61	.25/26
Fleenor (1996)	.42	.22
Jansen & Stoop (2001)	.62	.28
Kleinmann & Koller (1997)	.91	.31
Kudisch, Ladd & Dobbins (1997)	.41	.29
Sackett & Tunsiski (2001)	.58	.25
Arthur & Woehr (2003)	.55	.34

2.4.3.2 Factor analysis

Factor analysis is referred to as a systematic approach used to identify a group of variables which have a tendency to correlate highly with one another that represents some commonalities among variables (Thornton & Rupp, 2006). The use of factor analysis in the study of AC construct validity is twofold. Firstly, it has been used by researchers to group performance dimensions into general constructs and secondly, it has been used in a confirmatory manner to establish construct validity of performance dimensions (Kaufmann, Jex, Love & Libkuman, 1993). Thus, factor analysis is used to determine whether the underlying factor structure of ratings strongly represent the dimensions or the exercises in ACs.

In a factor analysis conducted by Fleenor (1996), for example, where it was hypothesised that AC ratings would present dimensions rather than exercises, it was found that the underlying factors represented exercises rather than dimensions. In another study by Greyling et al. (2003) it was confirmed that ratings were clustered according to exercises rather than performance dimensions and when nine dimensions were grouped to establish whether they supported a two-factor structure, this yielded only two factors, interpersonal and problem-solving.

Literature reports on factor analytical studies that show support for the construct validity evidence of an AC, for example Shore et al. (1990), as well as Arthur and Woehr (2003). Shore et al. (1990) hypothesised that performance style dimensions would cluster more to the cognitive ability measures rather than interpersonal style. Consequently, convergent validity of all interpersonal style and performance style dimensions was found, while discriminant validity was established for two interpersonal style dimensions and one performance style dimension. In support of these findings, another factor analytical study conducted by Damitz, Manzey, Kleinmann and Severin (2003), using the PCA to investigate the construct validity of AC ratings, showed assessors were able to distinguish between different behavioural dimensions associated with interpersonal-related skills and those related to performance style, which loaded on separate components.

Arthur and Woehr, (2003) in their meta-analytical study, found mean dimension convergent validity was higher for AER compared to WER (.43 vs .29) and dimension discriminant validity was stronger for the AER approach compared to the WER (.48 vs .58)². Thus, evidence of construct validity was demonstrated to an extent for convergent validity, particularly in the AER approach, but poor discriminant validity in both the AER and the WER is evident due to high correlations found in both methods. In Kudisch et al. (1997), however, contrasting results with regard to the construct validity of ACs

² Strong correlations (i.e. high r) between dimensions from across exercises (AER) indicate strong convergent validity, whilst weak correlations (low r) between dimensions measured within the same exercise demonstrate strong discriminant validity.

were reported. In their study it was found that assessors were able to organise information with regard to both dimensions and exercises.

The above discussion showed that overall findings of factor analysis have been mixed; while some found AC ratings to be exercise-based, others have shown that ratings are dimension-based. Furthermore, depending on the analytical strategy used, findings of construct validity evidence are mixed. Several factors that affect the validity of AC have been identified in literature. For the purpose of this study conceptual and evidence-based factors will be discussed. It should be noted, however, that most of the discussion will be on evidence-based factors. Conceptual factors are mostly related to administrative factors, which are stipulated in the Guidelines and Ethical Considerations for Assessment Centre Operations (Joiner, 2000), whilst evidence-based factors are focused on design and implementation aspects of ACs (Chen, 2006). With regard to evidence-based factors, method variance, inadequate assessor training, assessor bias, assessor cognitive demands, and the nature of candidate's performance will be discussed.

a) *Method variance*

Lievens and Conway (2001) emphasised the importance of construct validity of ACs, particularly when the resultant feedback is used for employee developmental purposes. Dimensions such as problem-solving, planning and organising and interpersonal relations were initially deemed as building blocks of ACs (Arthur & Woehr, 2003). In support of this view Lance (2008) advocated the fact that traditionally, dimensions in AC represented stable behavioural categories that were meant to be distinct within exercises and consistent across exercises. He claimed that this was the problem associated with the construct validity dilemma. However, Lievens and Conway (2001) reported on a confirmatory analysis in which they found that exercises rather than dimensions were the important factors in ACs.

Silverman (in Heindl, 1997) also argued that rating assessees after every exercise may lead assessors to think in terms of exercises, while rating after

completion of all exercises may lead assessors to think more in terms of dimensions. Schneider and Schmitt (1992) claimed that there has been extensive research on the consistency of method variance, thus lack of discriminant validity, in contrast to the vast findings that support convergent validity of ACs. Greyling et al. (2003) in their study stated that this lack of discriminant validity is due to the method variance, which is shown by high correlations between different dimensions within the same exercises. Donahue et al. (1997) investigated the effects of a behavioural checklist on the construct validity of ACs and found that the construct validity of that AC was higher when using behavioural checklists. In the study it was therefore found that the use of a behavioural checklist improves the discriminant validity of an AC. Furthermore, Lance et al. (2000), in their study of three independent samples, using the MTMM approach, found support for discriminant validity with the highest correlation at .36, indicating little overlap between exercise factors.

Lievens and Conway (2001) in their study found that the within-exercise dimension ratings demonstrated equal proportions of exercise and dimension variation, thus no evidence of method variance. Notwithstanding this, Kleinmann and Koller (1997) found evidence of method variance where three exercises measured correlated moderately ($r=.37$ between group discussion and role play, $r = .23$ between group discussion and presentation, $r= .33$ between role play and presentation), while high correlations between dimensions measured in the same exercise were found. These findings demonstrate dimensional influences on behaviour-rating variance rather than exercise variance.

In literature it is alleged that these findings are due to misuse of ACs in that, although they were designed to rate individuals in dimensions measured in various exercises, the focus seems to have shifted to rating of exercises instead (Arthur et al., 2003; Haaland & Christiansen, 2002; Howard, 1997; Lance, 2008; Meriac et al., 2008). In support of this view, Lance (2008) stimulated much debate among researchers in the field of AC construct validity studies (e.g. Arthur et al., 2008; Connelly, Ones, Ramesh & Goff, 2008 ; Howard, 2008; Lievens, 2008; Melchers & Koenig, 2008; Moses, 2008; Rupp, Thornton & Gibbons, 2008.).

Lievens and Conway (2001) reported on the inherent focus on the task-based model of AC ratings with little support for the dimension-based model. In contrast to the views of Lievens and Conway (2001) and Lance (2008), Howard (2008) argued that dismissing dimensions would mean recklessness in that the task-based approach is only relevant to prescribed jobs, which would not be applicable because of the changing nature of jobs. Furthermore, the consequence of a task-based approach would be the generalising of outcomes and the approach will fail to allow for generalisation of human qualities. Arthur et al. (2008) pointed out that this approach would be in contention with the traditional AC theory, which is based on the interpretation of psychological constructs (dimensions), which are the underlying features of job performance that are elicited from various AC exercises. Furthermore, Melchers and Konig (2008) are of the opinion that changing the focus of AC to a task-based approach will not necessarily solve the AC construct validity dilemma but rather mean a change in the name.

Lance (2008) argued that behaviour-related dimensions should be relatively stable across all AC exercises and be different within exercises. However, the author believes that despite various design studies, this does not hold true, as exercises rather than dimensions are the central point of AC research. Thus, exercises account for variances in AC performance and assessee's performance is actually cross-situationally specific and undifferentiated within exercises. Consequently, the researcher suggests that dimensions should be totally abandoned and a task-based approach rather than a dimension-based approach should be adopted.

Rupp et al. (2008) however, argued that redesigning ACs away from dimensions will be a mistake because of the fact that performance in a specific job is multidimensional, encompassing job-relevant knowledge, skills and abilities, and can only be appropriately assessed through behavioural methods. Furthermore, some researchers claim that Lance's suggestion is only based on the exclusive use of MTMM methodology. It is suggested that the MTMM approach fails to test the above findings (Rupp et al., 2008).

Howard (2008) and Lievens (2008) pointed out that the MTMM approach is in fact inappropriate in that ACs do not measure traits but rather observable behaviour.

It is therefore concluded that variance in terms of the exercise effect is larger than variance attributable to dimensions that are assessed. It is also evident from the above discussion that research in construct validity of ACs has been into the structure of the AC process, providing better insights into the internal structure of ACs (Meriac et al., 2008; Thornton & Rupp, 2006). The above section focused on evidence-related factors, while the next paragraph focuses on conceptual issues that influence the construct validity of ACs, including the halo effect, assessor cognitive demands, inadequate assessor training and leniency.

b) Inadequate assessor training

Joiner (2000) specified that assessors are to receive extensive training in the process of observing and classifying dimensions and behaviours, as well as demonstrating skill in rating the candidates. Assessor training therefore plays an integral part in the AC process. Assessor training has the potential to improve both the reliability and construct validity of ACs (Schleicher, Day, Mayes, & Riggio, 2002). Essentially, the training is centred on observations and interpretation patterns of behaviours within certain exercises and guidelines are provided for exercises (Kaufmann et al., 1993). Furthermore, Joiner (2000) alleged that training should involve a combination of knowledge, skills, abilities and general information.

A proposed form of training is a new frame of reference training (FOR) which focuses on the elimination of idiosyncratic standards held by the assessor and replaces them with a common FOR relating to ratings (Jackson, Atkins, Fletcher & Stillman, 2005; Schleicher et al., 2002). Jackson et al. (2005) further stipulated that FOR principles should be included in the assessor training to facilitate a shared understanding of the dimensions being measured. FOR includes teaching crucial dimensions including job and

behaviour indicative per dimension, discussion behaviour indicative of various effectiveness levels within each dimension, providing practice evaluations with new FOR, and giving feedback on the accuracy of ratings (Schleicher et al., 2002).

The FOR approach was used in Arthur et al. (2003) and evidence of both convergent and discriminant validity was found. Pulakos (1984), cited in Jackson et al. (2005), found evidence of accurate ratings in assessors who were exposed to FOR training in comparison to ratings from those who had not received FOR training.

Lievens (2001) suggested that to increase construct validity of AC, the length of assessor training should be increased. Jackson et al. (2005) recommended the use of psychologists as opposed to managers, as the use of the managers demonstrates worse convergent validity than the use of psychologists. In addition, Sagie and Magnezy (1997) found evidence of the influence of AC construct validity in relation to the type of assessor used, viz. manager vs. psychologist. However, Lievens (1998) alleged that differential use of various dimensions will not be influenced by either the length or the number of refresher courses attended by assessors.

In closing, it can therefore be deduced that in order to improve the construct validity of ACs, assessors should be trained in observing, classifying, evaluating and rating behaviour. Furthermore, the type of assessor is also crucial in that, although psychologists are generally not hired as assessors, when using managers as assessors it is desirable to train them rigorously in an attempt to increase construct validity.

c) *Assessor bias*

The most common forms of assessor bias in an AC are leniency and halo effect. Leniency is a primary rating error which refers to the assessor's tendency to award ratings that are too high in terms of actual performance levels (Bartels & Doverspike, 1997). In fact, Bartels and Doverspike (1997) hypothesised that leniency in AC ratings is related to the personality characteristics of the assessor. The halo effect occurs when an assessor is overly influenced by a single favourable or unfavourable trait that clouds his/her judgment of the individual's other traits (Swan & Wilson, 2007). Thus, one particular aspect of the assessee's performance has an impact on the evaluation of other aspects of performance. Halo effect occurrences are twofold, firstly when assessors perceive one dimension as the most important for overall performance in a given exercise and secondly, when an assessee's performance in each exercise is consistent and assessors are efficient enough in rating such performance, which in actual fact will result in exercise method variance (Lance, Foster, and Thoresen & Gentry, 2004).

According to Greyling et al. (2003) method variance is due to the overall assessor halo effect, as they struggle to differentiate between dimensions measured in the same exercise. In addition, the researchers claimed that the halo effect originates from processes that affect assessors' ratings, such as social desirability, personality factors, physical attractiveness and the actor-observer effect. Causes of the halo effect are closely associated with the assessor ratings of dimensions after a single exercise as opposed to ratings of dimensions after the completion of all the exercises (Thornton, 1992). Consequently, the AER is more consistent and accurate than the WER. In an attempt at eliminating common rater bias, Kolk, Born, Van Der Flier (2004b) allocated AER assessors one exercise each to rate and within each exercise the assessor had only one dimension to rate. Findings showed improved discriminant validity of the AC process.

d) *Assessor cognitive demand*

Construct validity of ACs depends on the accuracy of ratings, which is the assessors' ability to observe, evaluate and interpret candidates' performance demonstrated accurately in each exercise. In an AC process, the assessor recognises and attends to relevant information about the candidate, organises and stores such information for later access, recalls the information in an organised manner when ratings are being made and integrates information about the candidate into summary judgments (Kauffman et al., 1993). Moreover, construct validity may be affected because dimensions of managerial performance are not necessarily mutually exclusive, therefore an overlap may exist (Heindl, 1997). However, the research points out that if there is a large overlap of dimensions, this could be an indication that dimensions should be reduced to a single dimension. A solution to this dilemma would be to choose dimensions that are conceptually distinct and unrelated (Lievens, 1998).

The above process demands cognitive complexity on the part of the assessor (Howard, 1997). Lowry (1997) claimed that vague definitions of dimension and their numbers in exercises may confuse assessors in terms of differentiating between them. In fact, the researcher pointed out that assessors in the public sector simply cannot do it. Thus, the rating process in ACs calls for cognitive ability. According to literature, assessors' capacity to process information is limited; hence they are not always able to meet the cognitive demands of the AC process (Lievens, 2002; Lievens & Conway, 2001).

In support of this view, Cadwell et al. (2003) alluded to the inherent limited capacity of human beings to process information and the way in which such processes challenge the number of items that can be processed effectively. In view of the above, assessors may suffer from cognitive overload that is caused by behavioural information exhibited to assessors at a fast pace by candidates. In fact, according to Thornton (1992) social judgement plays a crucial part in the recording, rating and evaluation of assessee's performance. Social judgment refers to an objective process in which an individual's behaviour is evaluated in order to structure impressions about the individual's

strengths and weaknesses (Thornton, 1992). Social judgment is a fundamental concept in the AC process, in which assessors are expected to gather factual information about AC participants and systematically integrate such information in order to form accurate judgments about them. There are two approaches to social judgement, namely schema-driven and behaviour-driven approaches which are explain below (Thornton, 1992; Thornton & Rupp, 2006).

Schema-driven theory is centred on the concept of schema, which are mental pictures, beliefs or models about a specific concept that people observe (Heindl, 1997; Thornton & Rupp, 2006). Fiske and Taylor (in Heindl, 1997) refer to schemata as consisting of prior knowledge of a particular concept. Schematic-driven theorists hold the view that people have a limited capacity to process large quantities of information that they are exposed to at a given time (Thornton, 1992). Behaviour-driven theory, which is also known as the rational model (Lievens & Klimoski, 2001), held the view that assessors have the ability to observe detailed behaviour of candidates while classifying numerous parts of information into different categories (Thornton & Rupp, 2006). This assists them to arrive at objective and accurate judgments (Lievens & Klimoski, 2001; Lord & Maher, 1990; Thornton & Rupp, 2006).

Applied in an AC context, Schema-driven theory claims that assessors are believed to possess a limited capacity regarding processing information that they observe during the AC process. Thus, assessors have a limited capacity to meet the cognitive demands of the AC process (Lievens & Klimoski, 2001). In this instance, Heindl (1997) pointed out that the dimensions that are observed by the assessor are only means for simplifying and categorising information. Thornton and Rupp (2006) claimed that using schema has a direct impact on the assessor's process of observing and classifying behaviour, resulting in biased observations due to such faulty classification schema. Moreover, schema-driven theorists are of the view that the assessor's prior knowledge or initial view held about the assessee has an impact on the observations and ratings, resulting in support of the initial impression.

In summary, the schema-driven approach is in contrast with the behaviour-driven approach in that the former is simplified by heuristic evaluation procedures, while the latter is optimised by the evaluation of all alternatives (Lord & Maher, 1990). Whilst the behaviour-driven theorists hold the view that raters are able to observe behaviour displayed by the candidates, schema-driven theorists believe that raters have pre-conceived beliefs and perceptions about candidates. Secondly, behaviour-driven theorists believe that raters have an ability to memorise specific behaviour and actions; schema-driven theorists believe that raters are unable to memorise and hence they resort to general impressions about behaviour they observe. Thus, according to the behaviour-driven model, raters have an extensive memory capacity; while in the schema-driven model raters only use a moderate capacity. Lastly, behaviour-driven theorists posit that raters make objective judgments, while in the schema-driven approach, raters make subjective judgments. The implication of social judgement is that the larger the number of performance dimensions, the more difficult it is for assessors to differentiate between dimensions, and this causes cognitive overload of the assessor.

Recommendations to attend to the assessor cognitive overload are offered in literature (Donahue et al., 1997; Kauffman et al., 1993; Lance et al., 2000; Lievens, 2001; Robie, Osburn, Morris, Etchegaray & Adams, 2000). A vast amount of research proposes collapsing dimensions to a more manageable number. Evidence of construct validity was reported in Arthur et al. (2003) study after collapsing dimensions into smaller distinct sets of dimensions. Collapsing dimensions leads to improvements in construct validity of ACs, reducing cognitive overload as well as improving the quality of ratings (Lievens & Conway, 2001). In addition, Robie et al. (2000) and Rupp et al. (2008) showed support for the reduction of dimensions and the implementation of behavioural checklists that categorise behaviour according to dimensions. When assessors were provided with behavioural checklists that categorised behaviour according to dimensions in Donahue et al. (1997), findings demonstrated better discriminant validity. However, assessors may find it arduous to search through a long list while observing behaviour (Howard, 1997). Lievens (1998) also supports the view of limiting the number of dimensions in ACs.

Furthermore, Lievens and Conway, (2001) proposed making dimensions much more transparent to assesses. Kleinmann and Koller (1997) showed that assesses are better oriented towards dimensions and consistently exhibit the accompanying behaviour, Consequently, this increases the opportunity for assessors to observe dimension-relevant behaviour, which results in improved construct validity (Lievens & Conway, 2001).

e) *The nature of candidate's performance in assessment centres*

According to Lievens and Conway (2001) the lack of construct validity of ACs cannot be blamed only on assessors, but also the candidate's real performance across exercises. As mentioned earlier, in ACs candidates' performance is measured repetitively through different exercises. Against this background, it appears therefore that Lievens and Conway's (2001) comment is relevant when studying the construct validity of ACs because some candidate's may perform well on individual exercises, while others are better at group exercises (Lievens et al., 2006; Lievens, 2001). Thus, consistent behaviour against AC exercises cannot always be expected. The above point builds on social and personality psychology, particularly, the interaction, situational and the trait approach, (Haaland & Christiansen, 2002; Jackson et al., 2005; Lievens et al., 2006; Lievens & Conway, 2001).

The interactionist approach is based on the premise that there is a continuous interaction between individuals and the situation they find themselves in (Haaland & Christiansen, 2002; Heindl, 1997). The situational approach is based on the premise that individual behaviour differs from one situation to another; but remains consistent in similar situations (Haaland & Christiansen, 2002; Heindl, 1997; Lance et al., 2004). Lastly, the trait approach views individuals as similar in some ways, as well as unique in some respects (Bergh & Theron, 1999). According to Tett and Burnett (2003) traits are relevant features of situations, which make it reasonable to expect the expression of one trait (dimension) rather than another. Based on these theories, one can say that AC exercises are not parallel measures of the same dimension as

initially thought, but rather unique situations evoking inconsistent responses from candidates. For example the role play exercises, in-basket and presentation exercises are dissimilar situations which tap on different constructs of the participants (Lievens et al., 2006). Thus, the relevance of behavior expressed will depend on the cues called forth by a relevant situation. The above discussion suggests that in studying construct validity of ACs, it is important to consider environmental issues and how these impact on candidates' performance in ACs.

In summary, the above discussion has shown little support for construct validity evidence. Methods such as factor analysis, MTMM and correlation studies have been used to study both convergent and discriminant validity. However, only a few studies have found evidence of discriminant validity, while findings that support convergent validity have been shown. It is clear from the above discussion that studies showed WER correlated higher than AER, which points to lack of discriminant validity of ACs.

2.5 CONCLUSION

Literature reviewed in this chapter explained the history of psychological test use in South Africa, as well as the developments that led to the introduction of the currently used CBA approach. Given the discrimination that was evident, the implementation of LRA and the EEA was therefore a means to protect previously disadvantaged groups from unfair labour practices and unfair discrimination. Subsequently, the practice of psychological testing shifted to focusing on considering the development of culturally fair tests. Moreover, scientific investigation of potential bias and problems associated with testing in a multicultural context is now the practice. This is closely related to the use of CBA including the AC methodology. As such the CBA approach has become one of the widely used HR intervention strategies. Regardless of the relevance and the objectivity of the competency-based approach, it does appear from the above discussion that there is a level of contention among researchers regarding the practice of ACs as means to assess identified competencies.

This is particularly obvious in the various views held on the construct validity evidence of ACs. Although they are reported to be highly predictive of future job performance, the literature review in this chapter reveals lack of discriminant validity among dimensions measured in the same exercise and poor convergent validity among similar dimensions across exercises.

This chapter therefore discussed literature on competencies and competency-based assessments, and conceptualised the nature and characteristics underlying AC methodology. Firstly, the history of assessments and the relevance of the competency approach to the South African context were discussed. Furthermore, the definition and development of competencies were discussed and various techniques used to identify competencies were provided. Competency-based assessments were discussed, together with the steps involved in implementing a CBA battery. Secondly, features and characteristics of ACs were provided, together with their use in industry. In addition, the design process of AC was provided. It was further shown how literature on ACs is divided. Empirical evidence pertaining to the predictive and construct validity of ACs was also provided and issues that affect both the discriminant and convergent validities of ACs were highlighted. The next chapter includes an article integrating the research objective, methodology and results. The last chapter in this dissertation explicates conclusions drawn from the findings of this study, limitations and provides recommendations for both practice and future research.

CHAPTER 3: ARTICLE
CONSTRUCT VALIDITY OF A MANAGERIAL ASSESSMENT
CENTRE

ABSTRACT

The issue of construct validity has become a contentious issue in the study of assessment centres (ACs) around the world. This study was aimed at investigating the construct validity of AC dimensional ratings by using a Pearson product moment correlation coefficient and Principal component analysis (PCA). The sample consisted of 138 individuals who participated in a two-day AC for selection as partners/directors in an auditing firm. Twenty-one dimensions were measured using six different exercises. The results of the correlation study showed some evidence of discriminant validity amongst some dimensions measured in the same exercises, and the PCA revealed some convergent validity among various dimensions measured across exercises. A further factor analysis of the 21 dimensions revealed two factors, namely performance and interpersonal dimensions. Implications and recommendations for the design of ACs are addressed.

The implementation of legislation governing employment and assessment practices in South Africa (e.g. the Employment Equity Act, Act 55 of 1998, the Labour Relations Act, Act 66 of 1995, and the Health Professions Act, Act 56 of 1974) brought about a focus on employing psychological assessment practices that are culturally fair, appropriate and transparent (Paterson & Uys, 2005; Potgieter & Van der Merwe, 2002). In fact, South African labour law requires job-relevant psychological assessment in the workplace, which includes the measurement of job-related competencies (Foxcroft & Roodt, 2001; Muchinsky, Kriek, & Schreuder, 2002). As a result, in South Africa the use of competency-based-assessment (CBA) as a strategy for integrating various HR functions and to curb unfair assessment practices in the workplace has risen (Potgieter & Van der Merwe, 2002). Similar to Boam and Sparrow (1992), Saunders (2000) emphasised that users regard the CBA approach in comparison to traditional paper and pencil testing as fair, objective and

transparent. With competencies being regarded as the basic building blocks of assessment centres (ACs) (Potgieter & Van der Merwe, 2002), ACs have also evolved as a means to promote fair and non-discriminatory assessment in response to the South African legislative frameworks.

Assessment Centres

The use of ACs emerged as a means to select capable military leaders during World War 1, particularly by the German government (Chen, 2006; Thornton & Rupp, 2006; Woodruffe, 1990). Owing to their ability to predict future job performance (Arthur, Day, McNelly & Edens, 2003; Gaugler, Rosenthal, Thornton & Bentson, 1987; Jansen & Stoop, 2001; Lance, 2008), ACs have been in use in many public and private sector organisations (Harel, Arditi-Vogel & Janz 2003). The most popular reasons for the use of ACs, according to Spychalski, Quinones, Gaugler and Pohley (1997), were selection (50%) and promotion (45.8%), development planning (39.2%) and training (34.4%). Bartram (2004) also noted that organisations use ACs to improve productivity and effectiveness. The growing use of ACs became prevalent in South African organisations in the early 90's. Charoux (1991) reported that participants regard the AC procedure as fair and appearing to reduce the risk of bias. More recently, Schlebusch and Roodt (2008) noted that ACs are ideal for selection due to their reported multicultural appropriateness, minimising the chances of unfair discrimination. Moreover, AC research has demonstrated less adverse impact compared to cognitive ability tests and interviews for different cultural groupings (Cascio & Aguinis, cited in Dean, Roth & Bobko, 2008; Harel et al., 2003; Thornton & Rupp, 2006).

AC is a standardised approach of evaluating behaviour observed by a number of trained assessors (Joiner, 2000; Thornton & Rupp, 2006; Woodruffe, 1990). In an AC, job performance behavioural dimensions (e.g. interpersonal skills, leadership skills, problem-solving) are measured through a combination of exercises such as psychometric tests and simulation exercises. Thus, repetitive information about participants' distinct work-related attributes is obtained (Haaland & Christiansen, 2002; Joiner, 2000), implying simultaneous measurement of various competencies, which are otherwise difficult to

measure (Thornton & Rupp, 2006). Based on the target job, simulation exercises used usually resemble real-life situations and may include role play, presentation exercises and/or case studies (Dayan, Fox & Kasten, 2002; Joiner, 2000; Woodruffe, 1990). Given the use of multiple dimensions and exercises observed by more than one assessor, the approach is termed the multi-method, multi-trait, multi-rater approach (Campbell & Fiske 1959; Shore, Thornton & Shore, 1990).

Competencies measured in an AC are usually derived from a competency development process, sometimes referred to as competency modelling/profiling (Thornton & Rupp, 2006). Competency modelling/profiling is a process through which competencies representing job proficiency are identified, resulting in a competency framework for a particular job (Saunders, 2000). Competency profiling relates closely to job analysis (Chen, 2006; Joiner, 2002; Lievens, 1999) it follows a top-down approach as opposed to the bottom-up approach followed in job analysis (Thornton & Rupp, 2006). Information from different assessors and different assessment methods is integrated to form an overall judgment of each participant, usually during a wash-up session (Heindl, 1997; Thornton & Rupp, 2006). Participants are then provided with detailed feedback on their strengths and weaknesses with regard to the dimensions measured (Lievens, 2002). It is then reasonable to state that the value of ACs depends in part on the quality of assessors, implying intensive assessor training as crucial to the fairness of the evaluation (Cadwell, Thornton & Gruys, 2003).

Validity studies on assessment centres

The increased use of ACs has led to much research in the United States and the United Kingdom to examine the predictive validity of ACs (Arthur et al., 2003; Gaugler, et al., 1987; Harel et al., 2003; Jansen & Stoop, 2001). In a meta-analytical study by Gaugler et al. (1987) AC validity outcomes as high as .40 were reported. In the meta-analytical study by Schmidt and Hunter (1998) ACs showed an increase of 24% in practical value compared to general cognitive ability (GMA) and when compared to the interview, ACs also showed a higher utility than interviews (Harel et al., 2003; Thornton & Rupp, 2006).

Harel et al. (2003) furthermore, reported on 17 reviews of meta-analytical studies, in which the AC method was ranked with the highest validity when job performance ratings are used as criteria.

Despite the predictive validity evidence of ACs, a lack of construct validity in ACs has been highlighted in numerous studies (Arthur, Woehr & Maldegen, 2000; Kleinmann & Koller, 1997; Lievens, 1998; Shore, et al., 1990). Construct validation is a process of investigating whether the test measures the constructs that it claims to measure (Haaland & Christiansen, 2002). The lack of construct validity in studies on ACs seems to be related to the structure and the composition of ACs, such as the dimensions, exercises, assessor characteristics, candidate's performance, rating approach and the integration approach used.

Several approaches have been used to study the construct validity of AC ratings. Shore et al., (1990) have suggested Campbell and Fiske's (1959) multi-trait multi-method (MTMM) matrix to determine construct validity. Here a set of traits are measured by a number of methods and the results are presented in a correlation matrix called the MTMM matrix (Greyling, Visser, & Fourie, 2003). According to this approach, when strong correlations occur between different methods measuring the same dimension, convergent validity is displayed and when the correlation is weak between different dimensions measured within one method, discriminant validity is demonstrated (Greyling et al., 2003; Shore et al., 1990). In an AC, dimensions should represent stable behavioural categories within exercises and be consistent across exercises (Lance, 2008; Lievens & Conway 2001). Findings that support convergent validity of ACs have been made, yet research evidence of the lack of discriminant validity in ACs is rife (Lievens & Conway, 2001; Schneider & Schmitt, 1992).

Within the MTMM approach, some researchers ascribe the lack of discriminant validity in ACs to method variance (Arthur, Day & Woehr , 2008; Greyling et al., 2003; Kleinmann & Koller, 1997; Lance, Foster, Thoresen & Gentry, 2004)

implying that although ACs were designed to rate individuals on competency dimensions measured in various exercises, the focus seems to have shifted to the rating of exercises instead (Arthur et al., 2003; Haaland & Christiansen, 2002; Howard, 1997; Lance, 2008; Meriac, Hoffman, Woehr & Fleisher, 2008). Consequently, Lance (2008) suggested a total dismissal of the dimension-based approach and the adoption of a task-based approach. However, Howard (2008) argued that the task-based approach is not sustainable, owing to the changing nature of jobs. Similarly, Melchers and Konig (2008) argued that the task-based approach will result in the generalising of outcomes and not the generalisation of human qualities. In addition, Arthur et al. (2008) are of the opinion that the task-based approach is in contention with the traditional AC theory. Lastly, Rupp, Thornton and Gibbons (2008) argued that redesigning ACs away from dimensions will be a mistake as performance in a specific job is multidimensional, encompassing job-relevant knowledge, skills and abilities, and can only be appropriately assessed through behavioural methods.

The occurrence of method variance has largely been attributed to the manner in which assessors derive their final ratings to make a judgment about the candidates' performance in an AC (Arthur et al., 2008; Greyling et al., 2003; Lance et al., 2004). This relates mostly to the rating approach, rating errors, information processing and cognitive demands faced by assessors, as well as assessor training. For example, Lance (2008) claimed that using the within exercise rating approach (WER) moderates findings of construct validity evidence, as opposed to using the across-exercise rating approach (AER). In fact, higher mean convergent validity for the AER approach compared to the WER approach (.43 vs. .29) was found. Yet, despite the lower discriminant validity for the AER approach compared to the WER approach (.48 vs. .58)³ discriminant validity was still low as found in a meta-analysis study by Arthur et al. (2008).

³ Strong correlations (i.e. high r) between the same dimensions from across exercises indicate strong convergent validity, whilst weak correlations (low r) between dimensions measured within the same exercise demonstrate strong discriminant validity.

Bartels and Doverspike (1997) held the view that the lack of discriminant validity in AC studies is a result of rating errors (such as the halo effect and leniency) due to bias and assessors' inability to differentiate between dimensions. The halo effect is associated with the assessor ratings of dimensions after a single exercise as opposed to ratings of dimensions after the completion of all the exercises (Bartels & Doverspike, 1997; Lievens, 2001; Thornton, 1992). Leniency refers to the assessor's tendency to award ratings that are too high in terms of actual performance levels (Bartels & Doverspike, 1997). According to literature, assessors' capacity to process information is limited; hence they are not always able to meet the cognitive demands inherent to the AC process (Lievens & Conway, 2001; Lievens, 2002). Because the AC process demands cognitive complexity on the part of the assessor, vague definitions of dimension and their numbers in exercises may confuse assessors in terms of differentiating between them (Howard, 1997; Lowry, 1997). Moreover, construct validity may be affected because dimensions of managerial performance are not necessarily mutually exclusive and sometimes tend to overlap (Heindl, 1997). In order to address the issue of method variance Schleicher, Day, Mayes, and Riggio (2002) rightly noted that the training assessors receive has the potential to improve both the reliability and construct validity of ACs.

Apart from the MTMM approach, factor analysis has also been used in research to study construct validity of ACs (Arthur & Woehr, 2003; Damitz, Manzey, Kleinmann & Severin 2003; Fleenor, 1996; Greyling et al., 2003; Kaufmann, Jex, Love & Libkuman, 1993; Kudisch, Ladd & Dobbins, 1997; Shore et al., 1990). Factor analysis has been used to group performance dimensions into general constructs and to establish construct validity through confirming related performance dimensions (Kaufmann et al., 1993). Thus, by using factor analysis, one can establish whether the underlying factor structure of ratings strongly represent the dimensions or the exercises in ACs. Fleenor (1996), for example, hypothesised that ratings would present dimensions rather than exercises, but found that the underlying factors represented exercises rather than dimensions. Similarly, Greyling et al. (2003), using principal axis factor analysis, confirmed that ratings were clustered according to exercises rather than dimensions. In addition, when the nine dimensions in

Greyling et al. (2003) were grouped to establish whether they supported a two-factor structure, this yielded two factors, interpersonal and problem-solving.

Notwithstanding this, some factor analytical studies have shown support for construct validity evidence of an AC. For example, Shore et al. (1990) found convergent validity of all interpersonal style and performance style dimensions, while discriminant validity was established for two interpersonal style dimensions and one performance style dimension. Similarly, Kudisch et al. (1997) found that assessors were able to organise information on both dimensions and exercises. Equally, Damitz et al. (2003), using PCA, showed assessors were able to distinguish between different behavioural dimensions associated with interpersonal-related skills and those related to performance style, which loaded on separate components.

Construct validity is fundamental to fair assessment (Taylor & Boeyens, 1991), yet the discussion above highlights the fact that construct validity of AC remains a contentious issue that should continuously be explored and confirmed. In the South African context, the assessment of diverse groups requires an even greater emphasis on the continuous investigation of issues such as construct validity to ensure consistent striving towards fair assessment practice.

For this reason the general aims of this study were firstly to extend construct validity research in the South African context by studying the construct validity of competency dimensions measured in a leadership assessment and development assessment centre (LADC) that was designed for selection of partners/directors in an auditing firm. Specifically, the study aims were to investigate the following:

- Determine whether different dimensions measured within exercises are significantly discriminated from one another (discriminant validity)
- Determine whether there are significant correlations between similar dimensions across exercises in the LADC (convergent validity)

- Whether the number of dimensions measured in the LADC can be collapsed by meaningfully grouping dimensions into a smaller numbers of dimensions.

METHOD

In this study data gathered over four years in an AC (the aforementioned LADC) were analysed from a quantitative-relational approach. This was a correlation study exploring the relationships between scores on various dimensions within and across different exercises in the LADC.

Participants

Similar to Greyling et al. (2003), the sample in this study consisted of 138 individuals who participated in a two-day LADC. This study however differs in that it was conducted in a different research and business context than that of Greyling et al. (2003). For the purpose of this study, ratings from an LADC of a South African auditing firm was collected, while Greyling et al. (2003) studied AC dimensions assessed in a team leader AC. Furthermore, in the current study data were gathered from LADCs spanning the period from 2006 to 2010, whilst in the Greyling et al. (2003) study, data were gathered from an AC spanning a period of 11 months. Thus, in the current study secondary data was used as the LADC was designed for a different purpose than for research. The Greyling et al. (2003) AC also included different dimensions from those used in the current study. The ages of participants ranged from 26 to 52 years. Their race and gender breakdown is summarised in Table 3.1 below.

Table 3.1: Race and gender breakdown of LADC participants (N=138)

	Asian	Black	Coloured	White	Total
Female	8	10	5	26	49
Male	15	20	5	49	89
Total	23	30	10	75	138

The assessment centre methodology

The AC methodology under study encompassed competency dimensions, the exercises and the training of assessors. Twenty-one dimensions were measured in six different exercises, as opposed to the nine dimensions in the Greyling et al. (2003) study as measured in six exercises, including verbal ability and numerical interpretation.

Dimensions

The 21 dimensions measured in the LADC (depicted in Table 3.2) were identified from a job profiling exercise. These competencies are aligned to the company's shared values, regarded as important for success and applicable to professionals at all levels. In the process, behavioural anchors were assigned to each competency dimension and rated on a 5-point Likert scale (1 - unacceptable, 2 - less than acceptable, 3 - acceptable, 4 - more than acceptable and 5 - very high level of ability). Thus, 1 represents a very negative display according to the criterion and 5 a very positive display of the dimensions.

Assessment centre exercises

The following exercises were used to measure the identified dimensions:

Self-presentation skills (SPS) - the candidates are required to prepare a ten-minute presentation of the practice office nomination exercise described below, using any presentation aids, to the panel.

Executive interview (EXEC) - This is a one-on-one interview conducted by an Executive Committee member as an interviewer. The interview guidelines are provided to the interviewer prior to the interview.

Applied Leader Skills Assessment (ALS) - The ALS comprises a number of emails that a company's partner/director might typically find in his or her inbox. Once candidates have had an opportunity to respond to the emails, they meet with the ALS panel to discuss their responses. The purpose of the ALS is to assess the candidates' strengths and development needs in a number of

competencies considered important for leaders. Furthermore, the ALS exercise is tailored to be relevant for each specific business unit.

Criteria-based interview (CBI) -This assessment comprises a panel interview focused on defined core competencies.

Psychometric assessments (PSYCH) - The psychometric assessment involved the completion of two psychometric instruments, namely the Occupational Personality Questionnaire (OPQ) published by SHL and the Cognitive Process Profile (CPP) published by Cognadev. Various validity research studies exist for both these tools and they are used by many South African organisations, for selection and development purposes, amongst others. Bartram (2005) for example report on the construct and criterion related validity for the OPQ. The OPQ indicates an individual's preferred style or typical way of behaving. The CPP gives an indication of an individual's cognitive or preferred problem solving style, the work environment suited to his/her cognitive functioning, learning potential and generic cognitive competencies. These two psychometric exercises are not directly linked to the LADC, but are rather used as aiding tools and measured independently from the LADC. As a result assessors identified dimensions on each of the PSYCH contributing to or inhibiting the competencies assessed in the LADC (see in table 3.2 below). Norm scores obtained on the PSYCH on the related dimensions were then adjusted to the 5-point Likert scale used to assess the other measures in the LADC.

Practice office nomination (PON) - Candidates are required to prepare and write a 1 500-word paper/report on any topic they choose of relevance to their career: the company's services, clients, the company's people, the firm, technical issues or their profession as a whole.

Table 3.2 below outlines an assessment matrix of the various competency dimensions as measured in various LADC exercises.

Table 3.2: Dimensions and LADC exercises

Dimension	Exercise Method					
	SPS	EXC	ALS	CBI	PSYCH	PON
Client orientation			X			X
Applied judgment and insight			X			X
Business acumen			X			X
Knowledge-up-to-datedness			X			X
Practice leadership			X			X
Winning business			X			X
Product and service knowledge			X			X
Creativity and innovation						X
Written communication	X	X				X
Oral communication	X	X				X
Personal impact	X	X				X
Personal integrity						X
Interpersonal skills				X	X	X
Achievement drive				X	X	X
Team Work				X	X	X
Respect for diversity				X	X	X
Negotiation				X	X	X
Resilience				X	X	X
Flexibility				X	X	X
Leadership				X	X	X
General judgement				X	X	X

Assessors and assessor training

Assessors included three partners/directors, of whom one is a psychologist, and three are business unit managers. In the Greyling et al. (2003) study the assessors were call centre managers and HR practitioners and they were exposed to one-day training. In this study assessors attended a four-hour training workshop, focusing on developing fundamental assessor skills such as observation, recording, classification and the rating of candidate's behaviour. Moreover, the assessor training model of Saunders (2000) informed the training framework. During the workshop, assessors were informed about how the competencies were identified and how the AC was designed as well as providing them with knowledge of the context in which the assessment was being conducted. The relevance and definition of each dimension was explained to the assessors.

Procedure

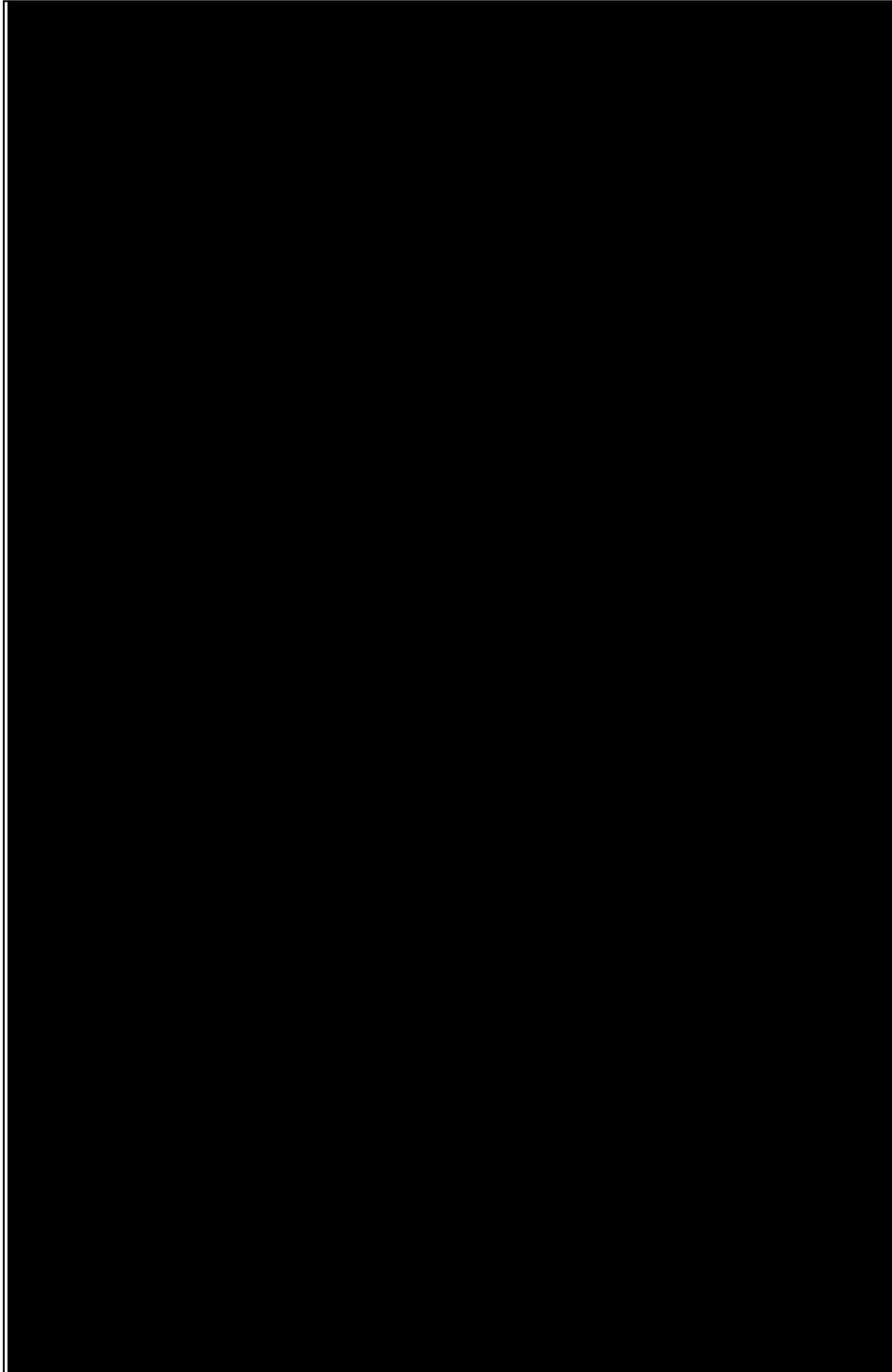
Individuals who participated in the LADC were managers of business units who were nominated by their senior executive managers for selection as partner/director. In each of the exercises described above, assessors were required to observe and record behaviour demonstrated by the participants.

During the AC, participating assessors were rotated, to ensure that each assessor observed a particular participant in only one exercise during each day of assessment. The observed behaviours were then matched with behavioural anchors associated with each dimension measured, on the 5-point Likert rating scale described above. After observing each exercise, assessors had to decide in which of the five categories the behaviour fell. Moreover; once the AC was completed, all the assessors met to validate and discuss discrepancies and reach consensus on final ratings, based on the combined ratings of all dimensions at the end of each exercise. Thus, the across exercise rating approach (AER) (Arthur et al., 2008; Lievens & Conway, 2001; Meriac et al., 2008) was used to derive ratings for each candidate's performance on each dimension. To ensure the smooth running of the AC, a project coordinator was made available as the AC administrator. Once all the final ratings had been obtained, assessors compiled feedback reports. Feedback to participants was then scheduled for two weeks after the completion of the AC. The feedback sessions involved the three partners and the psychologist.

RESULTS

This section presents results of the LADC discriminant and convergent validity research conducted, which are conveyed in a tabular format accompanied by a brief discussion of what is presented in each table. Descriptive statistics for the 21 dimensions measured in the six exercises using composite mean scores are shown in Table 3.3 below. The composite scores were attained by adding up all the item scores, and then divided by the number of respondents in the LADC, resulting in means that indicate values on a 5- point scale. In Tables 3.4 to 3.9 the questions regarding the discriminant validity of the various dimensions is addressed.

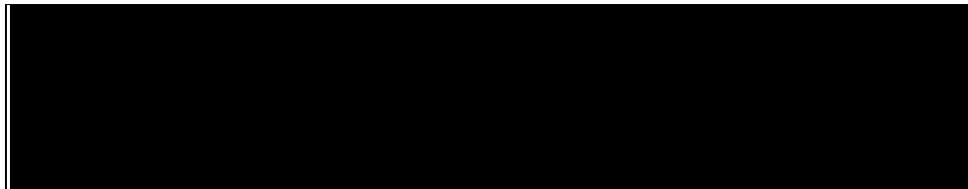
Table 3.3: Descriptive statistics for dimensions including composite scores



The first question of this study was to establish whether different dimensions measured within exercises are significantly discriminated from one another. Pearson product-moment correlation coefficients were calculated to obtain the correlations between dimensions in all LADC exercises separately. Tables 3.4 to 3.9 show the results of correlations among dimensions rated in the various exercises. In order to achieve discriminant validity, small correlations between the dimensions are expected.

The correlations between the three dimensions measured in the SPS shown in Table 3.4 below, ranged from medium to large correlations. In particular, a large correlation occurred between personal impact and oral communication ($r = .763$).

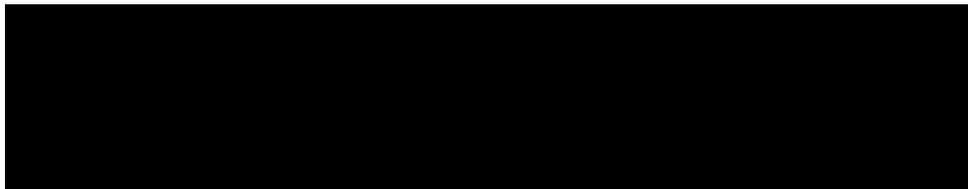
Table 3.4: Correlations between dimensions in SPS

A large black rectangular box redacting the content of Table 3.4.

** Correlation is significant at the 0.01 level (2-tailed)

Medium to large correlations occurred between the dimensions measured in the executive interview shown in Table 3.5 below. Interestingly, similar and statistically significant correlations occurred between personal impact and written communication as well as between personal impact and oral communication (for both $r = .748$).

Table 3.5: Correlations between dimensions in the executive interview

A large black rectangular box redacting the content of Table 3.5.

** Correlation is significant at the 0.01 level (2-tailed)

All of the correlations between the dimensions assessed in the ALS, shown in Table 3.6 below were statistically significant at the 0.01 level, ranging from

medium to large correlations, with the exception of the correlation between practice leadership and knowledge-up-to datedness ($r = .215$).

Table 3.6: Correlations between dimensions in ALS

	Client orientation	Applied judgement and insight	Business acumen	Knowledge-up- to- datedness	Practice leadership	Winning business	Product and service oriented
Client orientation	1.00						
Applied judgement and insight	.666**	1.00					
Business acumen	.657**	.570**	1.00				
Knowledge-up- to- datedness	.363**	.445**	.437**	1.00			
Practice leadership	.446**	.463**	.509**	.215*	1.00		
Winning business	.609**	.533**	.647**	.339**	.426**	1.00	
Product and service oriented	.385**	.355**	.447**	.401**	.373**	.485**	1.00

** Correlation is significant at the 0.01 level (2-tailed)

Only three of the correlations between dimensions measured in the CBI were statistically non-significant as shown in Table 3.7 below. Most of the correlations ranged from medium to large effect sizes. The highest correlation occurred between teamwork and interpersonal skills ($r = .628$).

Table 3.7: Correlations between dimensions measured in the CBI

* Correlation is significant at the 0.05 level (2-tailed)

Correlations between the nine dimensions rated in the PSYCH in Table 3.8 below ranged from statistically non-significant to the large correlation between leadership and negotiation ($r = .501$). Altogether 69% (25 out of 36) of the correlations between the dimensions were statistically non-significant. With the exception of the large correlations evident between leadership and negotiation,

the weak correlations in table 3.8 indicate evidence of discriminant validity amongst the 69% of dimensions measured in the PSYCH.

Table 3.8: Correlations between dimensions measured in the PSYCH

	Interpersonal skills	Achievement drive	Team work	Respect for diversity	Negotiation	Resilience	Flexibility	Leadership	General applied judgement
Interpersonal skills	1.00								
Achievement drive	-0.01	1.00							
Team work	.366**	0.10	1.00						
respect for diversity	0.12	-0.02	-0.01	1.00					
Negotiation	.186 [†]	0.14	0.11	0.05	1.00				
Resilience	0.11	-0.02	-0.01	.220 [†]	0.15	1.00			
Flexibility	.240**	0.00	.214 [†]	0.14	.240**	0.10	1.00		
Leadership	.354**	0.13	.432**	0.16	.501**	0.16	.189 [†]	1.00	
General applied judgement	-0.09	.263**	0.06	0.01	0.06	-0.05	-0.02	0.04	1.00

*Correlation is significant at the 0.05 level (2-tailed)

Table 3.9 below shows the correlations between the 21 dimensions measured in the PON. The correlations between dimensions in the PON ranged from statistically non- significant correlations to correlations statistically significant at the 0.01 level, with the largest correlations being between general judgment and applied judgment and insight ($r = .540$). Similarly, to table 3.8 above, the statistically non-significant correlations or weak correlations results show some evidence of discriminant validity.

Table 3.9: Correlations between dimensions measured in the PON

	Client orientation	Applied judgment and insight	Business acumen	Knowledge-up-to-datedness	Practice leadership	Winning business	Product and service knowledge	Creativity and innovation	Written communication	Oral communication	Personal impact	Personal integrity	Interpersonal skills	Achievement drive	Team work	Respect for diversity	Negotiation	Resilience	Flexibility	Leadership	General applied judgement			
Client orientation	1.00																							
Applied judgment and insight	.371**	1.00																						
Business acumen	.228*	.324**	1.00																					
Knowledge-up-to-datedness	0.14	.397**	.325**	1.00																				
Practice leadership	.260**	.399**	.262**	0.14	1.00																			
Winning business	.316**	.288**	.452**	0.12	.314**	1.00																		
Product and service knowledge	0.18	.293**	.277**	.310**	.260**	.252*	1.00																	
Creativity and innovation	0.16	.398**	.273**	.218*	.217*	.219*	.224*	1.00																
Written communication	.334**	.421**	.379**	.290**	.341**	.211*	0.19	.263**	1.00															
Oral communication	0.19	.289**	0.17	0.11	.326**	.224*	.223*	.204*	.402**	1.00														
Personal impact	-0.11	0.15	-0.01	-0.08	0.17	0.15	-0.03	0.01	-0.12	.338**	1.00													
Personal integrity	-0.11	0.08	-0.11	-0.02	0.07	-.201*	-0.07	-0.19	0.00	0.18	.334**	1.00												
Interpersonal skills	0.07	0.07	0.07	0.06	.225*	0.11	-0.08	.250*	0.11	.256**	0.11	-0.06	1.00											
Achievement drive	.238*	.203*	0.10	0.17	.351**	0.17	.210*	0.07	0.16	.368**	.210*	.286**	0.11	1.00										
Team work	.204*	.328**	.212*	0.17	.368**	0.19	.213*	0.12	.300**	.330**	.224*	0.15	.519**	0.11	1.00									
Respect for diversity	0.03	-0.05	0.14	-0.15	0.19	0.05	0.08	0.05	0.10	.330**	.203*	.382**	0.18	0.18	0.17	1.00								
Negotiation	.358**	.381**	.375**	0.10	.359**	.393**	0.19	.267**	.218*	.248*	0.09	-0.05	.292**	.350**	.357**	0.13	1.00							
Resilience	0.06	.310**	0.05	.242*	.314**	0.05	.225*	0.17	.259**	.364**	0.15	.262**	0.16	.402**	0.18	.265**	0.08	1.00						
Flexibility	0.18	.251*	.420**	0.13	0.18	.206*	.253**	.262**	.377**	.239*	-0.03	-0.05	.230*	0.03	.379**	0.16	.290**	.229*	1.00					
Leadership	.264**	.417**	.281**	0.12	.345**	.317**	.330**	.290**	.294**	.388**	.201*	0.06	.308**	.361**	.444**	0.07	.459**	.327**	.334**	1.00				
General applied judgement	.279**	.540**	.271**	.281**	0.17	0.20	0.13	.329**	.405**	.350**	0.13	0.07	0.07	0.05	.233*	0.06	.232*	.264**	.300**	.317**	1.00			

*Correlation is significant at the 0.05 level (2-tailed)

An important aim of the study was to investigate whether individual dimensions were highly correlated across the various LADC exercises, because the expectation is that constructs should be measured consistently irrespective of the method used for their measurement. These correlations would serve as an indication of convergent validity. Pearson product-moment correlation coefficients were used to calculate the correlations between the results obtained on various AC exercises for every dimension separately. Table 3.10 below presents correlations between the exercises used to measure all the dimensions.

Table 3.10: Correlations between dimensions across exercises (convergent validity)

Dimension	Exercise Method			
	SPS	Exec Interview	PON	
Written communication	SPS			
	Exec Interview	.319**		0.12
	PON	0.12	0.10	0.10
Oral communication	SPS			
	Exec Interview	0.16	0.16	-0.12
	PON	-0.12	0.03	0.03
Personal impact	SPS			
	Exec Interview	0.14	0.14	0.13
	PON	0.13	0.09	0.09
Interpersonal skills	CBI			
	PSYCH	.482**		0.09
	PON	0.09	-0.03	-0.03
Achievement drive	CBI			
	PSYCH	.482**		0.05
	PON	0.05	0.04	0.04
Team work	CBI			
	PSYCH	.389**		0.00
	PON	0.00	.203*	.203*
Respect for diversity	CBI			
	PSYCH	.396**		0.19
	PON	.240*	0.19	0.19
Negotiation	CBI			
	PSYCH	.469**		0.17
	PON	0.17	.282**	.282**
Resilience	CBI			
	PSYCH	.419**		0.09
	PON	0.09	-0.07	-0.07
Flexibility	CBI			
	PSYCH	.266**		0.14
	PON	0.14	0.03	0.03
Leadership	CBI			
	PSYCH	.515**		0.15
	PON	0.15	0.11	0.11
General judgement	CBI			
	PSYCH	.234**		0.19
	PON	0.19	0.12	0.12
Client orientation	ALS			
	PON	0.07	0.07	
Applied judgment and insight	ALS			
	PON	-0.04	-0.04	
Business acumen	ALS			
	PON	-0.06	-0.06	
Knowledge-up-to- datedness	ALS			
	PON	0.13	0.13	
Practice leadership	ALS			
	PON	-0.03	-0.03	
Winning business	ALS			
	PON	0.06	0.06	
Product and service knowledge	ALS			
	PON	-0.12	-0.12	

It was mentioned earlier that the expectation in AC construct validity research is that the same dimensions measured should correlate highly across exercises (i.e. strong convergent validity). The correlations shown in the table

3.10 above however, show somewhat different results with the exception of a large correlation between the CBI and PSYCH for the leadership dimension ($r = .515$). Less strong yet still statistically significant correlations between some dimensions measured across the CBI and PSYCH exercises are also evident (e.g. Interpersonal skills [$r = .482$], achievement drive [$r = .482$], teamwork [$r = .389$], respect for diversity [$r = .396$], negotiation [$r = .469$] and resilience [$r = .419$]). These correlations are supportive of convergent validity across these two exercises. A medium correlation also occurred between the executive interview and SPS for the written communication dimension ($r = .319$) as shown in Table 3.10 above.

The correlations of three dimensions (personal impact, interpersonal skills and oral communication) measured in the SPS, executive interview and the PON, are statistically significant. Moreover, correlations between the ALS and PON for the client orientation, applied judgment and insight, business acumen, knowledge-up-to-datedness, practice leadership, winning business, product and client service dimensions are also not statistically significant. Additionally, only a small correlation occurred between the PON and the PSYCH for the negotiation dimension ($r = .282$) and another small correlation occurred for the general applied judgment dimension measured in the CBI, PSYCH and the PON.

The AC ratings were then subjected to the Principal component analysis (PCA) which was followed by a varimax rotation using SPSS, to cluster dimensions together. The suitability of data for factor analysis was assessed prior to performing the PCA and the inspection of the correlation matrix revealed the presence of many coefficients of .5 and above. The PCA shown in Table 3.11, revealed the presence of 17 factors, which explained 76% of the variance. In Greyling et al. (2003) six factors were obtained from a principal axis factor analysis, explaining 67.01% of the variance. Given the nature of this study, the cut-off point for the factor matrices was .10; hence some values were omitted from the factor matrices.

Table 3.11: Rotated Factor Matrix ratings on the assessment centre exercises

	Component																	h2
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
SPS Oral communication	.303	.213	.792		-.122				.114									.813
SPS Personal impact	.222	.244	.759							.131			.119				-.116	.742
SPS Written communication	.199	.191	.659				.165				.206	.144		.138	.110	.310		.765
EXEC INTE Personal impact							.867			.103			.114	-.113	.107			.833
EXEC INTER Oral communication		.141		-.101			.862					.115						.821
EXEC INTER Written communication	.193	.185	.147	.126	-.171		.625		.199	-.158			-.109			.342		.742
ALS Business acumen	.188	.833												.127	.104			.795
ALS Client orientation		.798	.247		-.184			.146					-.128					.795
ALS Applied judgement and insight		.797						-.106						-.138		.117		.723
ALS Winning business	.175	.742	.196				.134				-.122			.132				.694
ALS Practice leadership	.123	.678	.120				.262		.168	.131		-.208		-.267		-.130		.759
ALS Product and service knowledge							.122					-.106				.817		.727
ALS Knowledge-up-to-datedness		.361	.124													.111		.828

Table 3.11: Rotated Factor Matrix ratings on the assessment centre exercises (continued)

	Component																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
CBI Achievement drive	.752	.118	.185		.117	-.178	.123	.145	.244				-.142	.111		-.102	.105
CBI Leadership	.746	.138	.255			.102	.179	.125	-.109	.158				.213	.136		
CBI Flexibility	.717		.214	.147		.115	.178	-.106						.113			-.182
CBI General Judgement	.692	.213	.124	.285		.186		-.121						-.222			
CBI Resilience	.679		.117	-.111				.117				.193	.183	-.208	.175	.252	.221
CBI Negotiation	.677				.250	-.212	-.103		.139			-.265	.146		.134	-.123	
CBI Team work	.636	.216			-.106		.225	.187	-.153	.387		.116			-.258		
CBI Interpersonal skills	.390	.127	.194				.192		-.439	.312	-.242	.208	.166	-.127	-.269	.185	-.162
CBI Respect for diversity	.309	.390		-.251			.271	.172	-.119		.318	.501					
PYSCH Team work	.155	.120	.159		.101				.110	.829							
PYSCH Leadership	.302	.117	.218			.153	.183			.452	-.310			.228	.342	-.175	.186
PYSCH Negotiation	.332	.133	.160			.389		.175	.140		-.487		.238	.200	.164	-.200	
PYSCH Respect for diversity			.113								-.130	.842			.158		
PYSCH Interpersonal skills		.152	.614						.269	-.425	.193	.167	-.142				
PYSCH Flexibility	.129	.291	.353			.344		-.269		.182	-.124	.107	.242			.195	
PYSCH General Judgement			.143	.123					.817			.103					
PYSCH Achievement drive	.385	.123					.213		.543		-.213	-.131		-.199	-.122	.296	
PYSCH Resilience	.209	.147			-.142				-.110		-.123	.114		.663	.197		
PON General judgement				.808	.170								.101				
PON Applied judgement and insight		-.122		.733		.229		.203		.153	.169			.109			.148
PON Written communication		-.134	.129	.526	.249	.154			.196	.101		-.238	-.293		.174		-.259
PON Oral communication			-.190	.443	.277	.132		.309	.182				.391		.234		-.209
PON Team work		-.126		.148	.761			.124		.197	.159		.133	-.147			
PON Flexibility	.154			.208	.725	.196	-.115	-.126	-.107	-.160	.100						
PON Interpersonal skills			-.167		.694			.205			-.175	.117		.374			
PON Leadership				.230	.499	.246	-.275	.408		.168			.199		-.109		-.150
PON Winning business	.155				.766					-.166	.134		.118	.129			
PON Business acumen				.307	.281	.622	-.120			-.131	.175	-.198			.241		.173
PON Client orientation				.332	.104	.583		.130	.145	.148			-.325				-.254
PON Negotiation			-.221	.178	.342	.544	-.164	.265		.273	-.111	.141			-.174		
PON Achievement drive					.239			.839									
PON Resilience	.103		.301	.284	.275	-.193		.620		-.169			.107				
PON Knowledge-up-to-datedness		-.153		.422				.157	.501		.323	-.198			-.145		.222
PON Product and service knowledge					.144	.241		.201	.119		.754			.114		-.140	
PON Personal impact			.114	.117	.108		.126			.113			.847				
PON Creativity and innovation	.103			.355	.131	.121	-.106			.123	.124			.736			
PON Personal integrity	-.135			.222	.128	-.262	.143	.306	-.125	.300			.182	-.471	.314		
PON Respect for diversity				.145	.330	-.101		.178	-.244		.221	.315		-.108	.502		-.127
PON Practice leadership	-.103	-.128		.263	.292	.319		.309		.209	.212			.137		-.360	-.100

As shown in Table 3.11 above, all SPS ratings loaded on the third factor, whilst the executive interview ratings loaded on the seventh factor. Ratings of the CBI loaded on the first factor, with the exception of respect for diversity, which loaded on the twelfth factor. ALS ratings loaded on the second factor, with the exception of product and client service, which loaded on the sixteenth factor. The PSYCH ratings loaded on various factors. Teamwork and leadership loaded on the tenth factor and flexibility on the third factor. Similarly, interpersonal skills loaded on the third factor, while negotiation loaded on the eleventh factor and respect for diversity on the twelfth factor. General judgment and achievement drive loaded on the eighth factor, while resilience loaded on the fifteenth factor. PON ratings also loaded on different factors. General judgment, applied judgment, written communication and oral communication loaded on the fourth factor. Teamwork, flexibility, interpersonal skills and leadership loaded on the fifth factor. Winning business, business acumen, client orientation and negotiation loaded on the sixth factor, while resilience loaded on the eighth factor. Knowledge currency loaded on the ninth factor and product and service on the tenth factor. Personal integrity and creativity and innovation loaded on the fourteenth factor, while respect for diversity loaded on the fifteenth factor. Personal impact loaded on the thirteenth factor and practice leadership on the sixteenth factor.

The third question of the research attempted to investigate whether the number of dimensions measured in the LADC under study can be collapsed by meaningfully grouping dimensions into a smaller number of dimensions. Therefore a composite score on every dimension across exercises was calculated for every individual. The intercorrelation matrix of the composite dimension ratings were subjected to PCA, which was followed by a varimax rotation using SPSS, to cluster dimensions together. PCA revealed the presence of two factors. The results of the factor pattern structure are presented in Table 3.12 below.

Table 3.12: Factor matrix of composite scores on 21 dimensions

Dimension	Component		h2
	1	2	
Client orientation	0.79		0.77
Applied judgment and insight	0.81		0.81
Business acumen	0.82		0.78
Knowledge- up-to-datedness	0.78		0.77
Practice Leadership	0.79		0.78
Winning business	0.76		0.71
Product and service knowledge	0.54		0.36
Creativity and innovation	0.73	0.52	0.80
Written communication		0.73	0.78
Oral communication		0.80	0.84
Personal impact		0.82	0.84
Personal integrity	0.68	0.61	0.83
Interpersonal skills	0.60	0.52	0.63
Achievement drive	0.72		0.62
Team work	0.73		0.68
Respect for diversity		0.89	0.91
Negotiation	0.60		0.60
Resilience		0.88	0.92
Flexibility		0.81	0.83
Leadership		0.87	0.92
General applied judgment	0.77		0.78

Client orientation, applied judgment and insight, business acumen, knowledge up-to-datedness , practice leadership, winning business, product and service knowledge, creativity and innovation, personal integrity, interpersonal skills achievement drive, teamwork, negotiation and general judgment loaded on the first factor, named interpersonal dimension. The second factor consisted of performance dimensions, and included written communication, oral communication, personal impact, respect for diversity, resilience, flexibility and leadership.

Similar to Greyling et al. (2003), the findings of the above factor structure relating to the final dimensions indicate that these dimensions are not necessarily indicative of completely separate concepts; thus only two factors are required to explain the assessor variations ratings. One can therefore

conclude that the findings of the present study support the AC construct validity dilemma that dimension ratings within the same exercises have a higher correlation than ratings of the same dimensions across different exercises.

DISCUSSION

The findings of the present study showed some evidence of discriminant validity, demonstrated by the statistically non-significant correlations between some dimensions measured in the PSYCH and between some dimensions measured in the PON. However; the majority of dimensions measured within various exercises showed no evidence of discriminant validity. The lack of discriminant validity amongst these dimensions is similar to the findings of Greyling et al. (2003), but the correlations in Greyling et al. (2003) ranged from small to medium, whereas the majority of correlations in the current study range from medium to large. The highest correlations in the present study are between oral communication and personal impact measured in both the SPS and the executive interview. Some evidence of construct validity is however apparent, shown by the correlation of the same dimensions across different exercises (convergent validity). Among the 21 dimensions measured in different exercises, ten are statistically significant, shown by medium correlations and one large correlation between the PSYCH and CBI for the leadership dimension. These findings are different from those of Greyling et al. (2003) in that of the nine dimensions they measured in six different exercises, only three (analytical thinking, motivating others and self-control) were statistically significant. However, only two medium correlations were found for the dimensions (analytical thinking and self-control) in the interview, role play, group exercise and case study.

These results were further supported by the factor analysis findings. PCA results derived from the AC ratings indicate levels of convergent validity of dimensions measured in some exercises. These results indicate that to some degree, assessors were able to differentiate similar dimensions measured in various exercises (e.g., in the CBI, PSYCH, ALS and PON). The findings of the correlation study are similar to previous research, particularly the within-

exercise correlations. Thus, although there is some evidence of convergent validity, there is very little evidence of discriminant validity for different dimensions. The above findings seem to be in support of Lievens's (1998) findings, in which ratings of the same dimensions across exercises had high correlations (i.e. stronger convergent validity), while ratings of different dimensions within the same exercise had high correlations (i.e. weak discriminant validity).

Construct validity of AC depends on the accuracy of ratings. This, according to Joiner (2000), is the assessor's ability to observe, evaluate and interpret candidates' performance demonstrated in each exercise accurately. According to Kauffman et al. (1993), in an AC process the assessor firstly recognises and attends to relevant information about the candidate. Secondly, the assessor organises and stores such information for later access and recalls the information in an organised manner when ratings are made. Lastly, the assessor integrates information about the candidate into summary judgments (Kauffman et al., 1993). Arriving at a particular rating of performance demonstrated by the participant is a complex cognitive process on the part of the assessor. Therefore, a large number of dimensions can make this process even more challenging. The findings of the present study emphasise the cognitive challenges that assessors were faced with during the rating process. This was due to the number of dimensions, which seemed to be too many (i.e. 21) and may have impacted negatively on the construct validity of the AC.

Another factor associated with the lack of discriminant validity in the present study is related to the definitions of dimensions measured. Thus, assessors may have struggled with long definitions of the dimensions under study. Consequently, assessors battled to differentiate between these dimensions. According to Lowry (1997) vague definitions of dimensions and their numbers in exercises may confuse assessors in terms of differentiating between them. This may result in a limited capacity of the assessor to process information and thus inability to meet the cognitive demands called for by the AC process. In addition, from the overlap evident among dimensions of this study one can

conclude that the high correlations among dimensions measured in the same exercise are a result of assessors being unable to differentiate between dimensions measured WER.

Moreover, assessors are to receive extensive training regarding the process of observing and classifying candidate's behaviour during the AC procedure, according to Joiner (2000). Assessor training thus not only implies reliability of the AC, but also influences its construct validity. In the present study, however, assessor training seems highly unlikely to have influenced these findings. To improve discriminant validity Lievens (2001) suggested the inclusion of the organisation's values, norms and personal qualities in assessor training. In the light of the above, the participating organisation's norms and values did not form part of assessor training, but the dimensions measured were arrived at through the organisation's values and strategies. These dimensions are viewed as qualities that sustain the organisation's competitive advantage. Therefore the impact of assessor training on the lack of construct validity in the present study is not ascribed to poor assessor training.

Another factor that could have had an impact on the lack of discriminant validity of the present AC could have been the type of assessors used. As mentioned earlier, of the four assessors, only one was a psychologist. Jackson, Atkins, Fletcher and Stillman (2005) recommended the use of psychologists as opposed to managers, as the use of managers demonstrates poor construct validity in relation to the use of psychologists. In Sagie and Magnezy (1997) evidence of the influence of AC construct validity in relation to the type of assessor used, manager versus psychologist was found. In the current study therefore, the assessors used were predominantly business unit managers and may have contributed to the lack of discriminant validity in this study.

Some of the assessors in the present study had knowledge of the firm's strategy and values, as well as experience in the target job and prior

knowledge of the candidates. Consequently, they make assumptions about the ideal candidate for the job. Thus, assessor schemata, including their characteristics and demographics, can be regarded as another contributory factor to the lack of discriminant validity of this study. Such schemata result in biased ratings. Heindl (1997) and Thornton and Rupp (2006) held the view that assessors will most likely pay attention to what is consistent with their assumptions and ignore contradictory information. Moreover, when Bartels and Doverspike (1997) studied assessor personality characteristics and their impact on the leniency of ratings, they found that assessors with high intelligence and sensitivity tended to be more lenient in their ratings. The transparency of dimensions to candidates in the present study is unknown. Kleinmann (1993) suggested that when there is lack of transparency regarding the dimension, the probability that participants will select incorrect action strategies increases, therefore influencing construct validity results of the particular AC. Consequently, the lack of discriminant validity in the AC under study could have been affected if the dimensions were not transparent to the participants.

There are a few limitations in the present study that are worth mentioning. Firstly, this is not a controlled study; therefore researchers were not involved in establishing the guidelines for AC administration and the scoring procedures. Consequently, the knowledge of the extent of standardisation procedures is limited. Lack of standardisation in the AC method and administration may have had an impact on the construct validity of ratings and the AC method itself (Kaufman et al., 1993). Secondly, the sample size (n=138) used in this study is relatively small, therefore findings cannot be generalised. Moreover, the sample used in this study used ratings from a single AC of one organisation and the dimensions reflected the organisation's values and strategies. Lastly, a small number of assessors were used and these findings can therefore not be generalised to other ACs and across organisations.

In summary, continued research is needed into the construct validity of ACs, particularly into the design and methodological characteristics as opposed to

the AC as a measurement instrument. A replication of the current study, using a larger sample is recommended. Additionally, meta analysis studies of AC construct validity is also recommended. As mentioned earlier, a large number of dimensions pose cognitive challenges to the assessors. To improve the construct validity of ACs it is recommended firstly that the dimensions be reduced to a much smaller number, especially when the AC is used for selection. Secondly, only a few conceptually distinct dimensions should be considered so that assessors can differentiate between them more clearly. There is supporting evidence in literature of improved construct validity findings when dimensions were grouped into conceptually related and distinct dimensions (Arthur & Woehr, 2003; Damitz et al., 2003; Shore et al., 1990). In fact, evidence of construct validity was reported in the work of Arthur et al. (2003). Lievens and Conway (2001) suggested that collapsing dimensions leads to improvement in the construct validity of ACs, reducing cognitive overload as well as improving the quality of ratings. To improve the standardisation of evaluation between assessors, more concrete definitions of dimensions should be provided.

Managerial assessors may experience more difficulty in providing different ratings on the given dimensions; therefore the use of psychologists instead of managers as assessors and in the development and design of an AC is suggested. Although having as many assessors as there are dimensions is costly, increasing the number of assessors may improve the construct validity of the AC under study. Lastly, it is suggested that dimensions be made much more transparent to the candidates. In Kleinmann and Koller's (1997) study it was found that candidates who were better oriented towards dimensions were able to exhibit accompanying behaviours more consistently. Similarly, Kleinmann et al. (2011) confirmed that candidates' ability to identify evaluation criteria in an AC improves their criterion related performance in the AC, by implication the overall validity of the AC. This increases the opportunity for assessors to observe dimension-relevant behaviour, which results in improved construct validity.

REFERENCES

- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment centre dimensions. *Personnel Psychology, 56*, 125-153.
- Arthur, W., Day, E. A. & Woehr, D. J. (2008). Mend it, don't end it; an alternative view of assessment centre construct – related validity evidence. *Industrial and Organisational Psychology, 1*, 105-111.
- Arthur, W. & Woehr, D. J. (2003). The construct-related validity of assessment centre ratings: A review and meta-analysis of methodological factors. *Journal of Management, 29*(2), 231-258.
- Arthur, W., Woehr, D.J. & Maldegen, R. (2000). Convergent and discrimination validity of assessment centre dimensions: A conceptual and empirical re-examination of the assessment centre construct-related validity paradox. *Journal of Management, 26*(4). 813-835
- Bartels, L. K., & Doverspike, D. (1997). Assessing the assessor: The relationship of assessor personality and leniency in assessment centre ratings. *Journal of Social Behavior and Personality, 12*(5), 179-190.
- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185-1203.
- Bartram, D. (2004). Assessment in organisations. *Applied Psychology: An International Review, 53*(2), 237-259.
- Bedell, M.R. (2001). Personality testing: controversial no more. Keynote address presented at the 4th annual conference of the Society of Industrial Psychology, Pretoria.
- Boam, R. & Sparrow, P. (1992). *Designing and achieving competency: A competency-based approach to developing people and organisations*. London: McGraw-Hill Book Company Limited.
- Cadwell, C., Thornton, G. C. & Gruys, M.L. (2003). Ten classic assessment centre errors: Challenges to selection validity. *Public Personnel Management, 32*(1), 73-88.

- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Charoux, E. (1991). Assessment centres offer a fair and accurate way to select and develop talent. *Management Today*, 7(2), 28-29.
- Chen, H. (2006). Assessment centre: A critical mechanism for assessing HRD effectiveness and accountability. *Advances in Developing Human Resources*, 8(2), 247-264.
- Chen, H. & Naquin, S.S. (2006). An integrative model of competency development, training design, assessment centre and multi-rater assessment. *Advances in Developing Human Resources*, 8(2), 265-282.
- Damitz, M., Manzey, D., Kleinmann, M. & Severin, K. (2003). *Applied Psychology: An international Review*, 52(2), 193-212.
- Dayan, K., Fox, S. & Kasten, R. (2002). Entry-level police candidate assessment centre: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, 55(2), 827-849.
- Dean, M.A., Roth, P.L. & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment centre ratings: A meta analysis. *Journal of Applied Psychology*, 93(3), 685-691.
- Employment Equity Act, number 55 of 1998. *Government Gazette*. Republic of South Africa.
- Fleener, J.W. (1996). Constructs and developmental assessment centres: Further troubling empirical findings. *Journal of Business and Psychology*, 10(3), 319-334.
- Foxcroft, C. & Roodt, G. (2001). *An introduction of psychological assessment in the South African context*. Cape Town: Oxford University Press.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. & Bentson, C. (1987). Meta-analysis of assessment centre validity. *Journal of Applied Psychology*, 72(3) 493-51.
- Greyling, L., Visser, D. & Fourie, L. (2003). Construct validity of competency dimensions in a team leader assessment centre. *SA Journal of Industrial Psychology*, 29(2), 10-19.

- Haaland, S. & Christiansen N.D. (2002). Implications of Trait-Activation Theory for evaluating the construct validity of assessment centre ratings. *Personnel Psychology*, 55, 137-163.
- Harel, G. H., Arditi-Vogel, A. & Janz, T. (2003). Comparing the validity and utility of behavior descriptions interview versus assessment centre ratings. *Journal of Managerial Psychology*, 18(2), 94-104.
- Health Professions Act, number 56 of 1974. Health Professions Council of South Africa. Pretoria.
- Heindl, W.A. (1997). *Assessment centres: A comparison of trait ratings and task ratings for predicting managerial performance*, Unpublished Masters dissertation. Pretoria: UNISA.
- Howard, A. (2008). Commentary: Making assessment centres work the way they are supposed to. *Journal of Industrial and Organisational Psychology*, 1, 98-104.
- Howard, A. (1997). A reassessment of assessment centres: Challenges for the 21st century. *Journal of Social Behavior and Personality*, 12(5), 13-52.
- Jackson, D.J.R., Atkins, S.G., Fletcher, R.B. & Stillman, J.A. (2005). Frame of reference training for assessment centres: effects on interrater reliability when rating behaviors and ability traits. *Public Personnel Management*, 34(1), 17-30.
- Jansen, W.P. & Stoop, B.A.M. (2001). The dynamics of assessment centre validity: results of a 7-year study. *Journal of Applied Psychology*, 86(4), 741-753.
- Joiner, D.A. (2002). Assessment centres: What's new? *Public Personnel Management*, 31(2), 179-183.
- Joiner, D. A. (2000). Guidelines for ethical consideration for assessment centre operations: International Task Force on Assessment Centre Guidelines. *Public Personnel Management*, 29(3), 315-331.
- Kaufman, J.R., Jex, S.M., Love, K. G. & Libkuman, T.M. (1993). The construct validity of assessment centre performance dimensions. *International Journal of Selection and Assessment*, 1(4), 213-223.

Kleinmann, M. (1993). Are rating dimensions in assessment centres transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*, 78(6), 988-993.

Kleinmann, M., Ingold, P.V., Lievens, F., Jansen, A., Melchers, K.G., & Konig, C. J. (2011). A different look at why selection procedures work: the role of candidates' ability to identify criteria. *Organisational Psychology Review*, 1(2), 128-146.

Kleinmann, M. & Koller, O. (1997). Construct validity of assessment centres: appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behavior and Personality*, 12(5), 65-84.

Kolk, N.J., Born, M., & Van Der Flier, H. (2004a). A triadic approach to the construct validity of the assessment centre. *European Journal of Psychological Assessment*. 20(3), 146-156.

Kriek, H.J. & Charoux, J.A.E. (1994). The assessment centre: Testing the fairness hypothesis. *Journal of Industrial Psychology*, 20(2), 21-25.

Kudisch, J.D., Ladd, R.T. & Dobbins, G.H. (1997). New evidence on the construct validity of diagnostic assessment centres: The findings may not be so troubling after all. *Journal of Social Behavior and Personality*, 6(3), 141-152.

Labour Relations Act, number 66 of 1995. Republic of South Africa.

Lance C.E. (2008). Why assessment centres do not work the way they are supposed to. *Journal of Industrial and Organisational psychology*, 1, 84-97.

Lance, C. E., Foster, M. R., Thoresen, J. D. & Gentry, W. A. (2004). Assessor cognitive processes in an operational assessment centre. *Journal of Applied Psychology*, 89(1), 22-35.

Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R. & Smith, D. E. (2000). Assessment centre exercise factors represent cross-situational specificity, not method bias. *Human Performance*, 13(4), 323-353.

Lievens, F. (2008). What does exercise-based assessment really mean? *Industrial and Organisational Psychology*, 1, 112-115.

- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centres: An examination of assessor and assessee effects. *Journal of Applied Psychology, 87*(4), 675-686.
- Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at troubling issues. *Journal of Organisational Behavior, 22*, 2003-221.
- Lievens, F. (1999). Development of a simulated assessment centre. *European Journal of Psychological Assessment, 15*(2), 117-126.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centres: A review. *International Journal of Selection and assessment, 6*(3) 141-142.
- Lievens, F. & Conway, J.M. (2001). Dimension and exercise variance in assessment centre scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*(6), 1202-1222.
- Lievens, F. & Klimoski, R. J. (2001). Understanding the assessment centre process: Where are we now? *International Review of Industrial and Organisational Psychology, 16*, 246-286.
- Lowry, P. (1997). The assessment centre process: New directions. *Journal of Social Behavior and Personality, 12*(5), 53-62.
- Melancon, C.S. & Williams, M. (2006). Competency-based assessment centre design: A case study. *Human Resources, 8*(2), 283-315.
- Melchers, K.G. & Konig, C.J. (2008). It is not yet time to dismiss dimensions in assessment centres. *Industrial and Organisational Psychology, 1*, 125-127.
- Meriac, J.P., Hoffman, B.J., Woehr, D.J. & Fleisher, M.S., (2008). Further evidence of validity of assessment centre dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*(9), 1042-1052.
- Meyer, T.S.P. (1996). *Creating Competitiveness through Competencies: Currency for the 21st Century*. Randburg: Knowledge Resources.
- Moses, J. (2008). Assessment centres work, but for different reasons. *Industrial and Organisational Psychology, 1*, 134-136.

- Mouton, J. (2005). *How to succeed in your Master's and Doctoral Studies: A South African Guide and Resource Book*. Pretoria: Van Schaik Publishers.
- Muchinsky, P., Kriek, H.J. & Schreuder, D. (2002). *Personnel Psychology*. Cape Town: Oxford, Southern Africa.
- Paterson, H. & Uys, K. (2005). Critical issues in psychological test use in South African workplace. *SA Journal of Industrial Psychology*, 31(3), 12-22.
- Potgieter, T.E. & Van der Merwe, R.P. (2002). Assessment in the workplace. A competency-based approach. *South African Journal of Industrial Psychology*, 28(1), 60-66.
- Reilly, R.R., Henry, S. & Smither, J.W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment centre dimensions. *Personnel Psychology*, 43, 71-84.
- Robie, C., Osburn, H.G., Morris, M.A., Etchegaray, J. M. & Adams, K.A. (2000). Effects of the rating process on the construct validity of assessment centre dimension evaluations. *Human Performance*, 13(4), 355-370.
- Rupp, D.E., Thornton, G.C. & Gibbons, A.M. (2008). The construct validity of the assessment centre method and usefulness of dimension as focal constructs. *Industrial and Organisational Psychology*, 1, 116-120.
- Sagie, A. & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology*, 70, 103–108.
- Saunders, E. (2000). *Assessing human competence: Practical guidelines for South African managers*. Johannesburg: Knowledge Resources.
- Schlebusch, S. & Roodt, G. (2008). *Assessment centres: Unlocking potential for growth*. Randburg: Knowres Publishing.
- Schleicher, D.J., Day, D.V., Mayes, B.T. & Riggio, R.E. (2002). A new frame for frame-of- reference training: Enhancing the construct validity of assessment centres. *Journal of Applied Psychology*, 87(4), 735-746.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection in personnel psychology: Practical and theoretical implications of 85 years of research. *Psychological Bulletin*, 12(2), 262-274.

Schneider, R.G., & Schmitt R.M. (1992). The relation of cognitive ability and personality traits to assessment centre performance. *Journal of Applied Social Psychology*, 30(7), 1474–1491.

Shore, T.H., Thornton, G.C., & Shore, L. (1990). Construct validity of two categories of assessment centre dimension ratings. *Personnel Psychology*, 43, 101- 116.

Spychalski, A.C., Quinones, M.A., Gaugler, B.B. and Pohley, K. (1997). A survey of assessment centre practices in organizations in the United States. *Personnel Psychology*, 50, 71-90.

Taylor, T.R. & Boeyens. J.C.A. (1991). The comparability of scores of blacks and whites on the South African Personality Questionnaire: An exploratory study. *South African Journal of Psychology*, 21(1), 1-11.

Thornton, G.C. (1992). *Assessment centres in human resource management*. Massachusetts: Addison-Wesley Publishing Company.

Thornton, G.C. & Rupp, D.E. (2006). *Assessment Centres in Human Resource Management. Strategies for prediction, diagnosis and development*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Thornton, G.C., Tziner, A., Dahan, M., Clevenger, J. P. & Meir, E. (1997). Construct validity of assessment centre judgments: Analyses of the behavioral reporting method. *Journal of Social Behavior and Personality*, 12(5), 109-128.

Woodruffe, C. (1990). *Assessment centres: Identifying and developing competence*. London: Institute of Personnel Management.

CHAPTER 4

CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

4.1 INTRODUCTION

The general aim of this research was to investigate the construct validity of a LADC. Empirical aims of the research were to investigate both the convergent and discriminant validity of the LADC; and whether the dimensions measured can be grouped into a smaller number of dimensions. These aims have been achieved and the results reported on in chapter 3. This chapter contains the conclusions, limitations and recommendations as the concluding phase of the study. Firstly, conclusions will be drawn regarding the construct validity of the LADC. Secondly, limitations, regarding generalisability of findings will be discussed. Lastly, recommendations for practice and future research will be made particularly in terms of the limitations imposed by the assessment centre operations and the sample size.

4.2 CONCLUSIONS

Construct validity of ACs depends on the accuracy of ratings, viz. assessors' ability to observe, evaluate and interpret candidates' performance demonstrated in each exercise (Joiner, 2000). In an AC process, the assessor recognises and attends to relevant information about the candidate, organises and stores such information for later access, recalls the information in an organised manner when ratings are made and integrates information about the candidate into summary judgments (Kaufman et al., 1993; Thornton, 1992). Previous research findings on the lack of AC construct validity have, however, puzzled researchers for a long time compared to the existence of predictive validity findings. This includes the lack of discriminant validity among dimension ratings of the same exercises.

The first question of the research was to determine whether dimensions measured within exercises are significantly discriminated from one another, thus, the discriminant validity. Although there was some evidence of discriminant validity between some dimensions measured in the PSYCH and PON, the high correlations found amongst the majority of different dimensions measured in the same exercises suggest a lack of evidence for discriminant validity for the LADC under study. Thus, the findings of this study support the widely held view that ACs lack the ability to discriminate amongst various dimensions within the same exercises (e.g. Arthur et al., 2000; Kleinmann and Koller, 1997; Lievens, 1998; Shore et al., 1990). In AC construct validity research, high correlations between the dimensions measured across exercise are expected. The correlations results in this study however, showed small to medium across exercise correlations amongst the majority of dimensions. Strong correlation was found however, for the leadership dimension measured in the CBI and PSYCH, whilst medium correlation for the interpersonal style dimension assessed in the CBI, PSYCH and PON exercises were found. These findings indicate very little convergent validity evidence of the LADC under study. The results of the PCA derived from the LADC ratings; indicate some level of convergent validity of some dimensions measured across different exercises. These results indicate that to some degree, assessors were able to differentiate between dimensions measured in various exercises (e.g., the CBI, PSYCH, ALS and PON). The above findings addressed the second question of this study, which sought to determine whether there are significant correlations between similar dimensions across exercises in the LADC (convergent validity).

The third question was to investigate whether the number of dimensions measured in the LADC can be collapsed by meaningfully grouping dimensions into a smaller number of dimensions. The findings of the PCA revealed two factors named interpersonal and performance dimensions. The two-factor structure indicate that assessors were able to differentiate between interpersonal and performance-related dimensions. These findings are consistent with those of Shore et al. (1990), where dimensions were classified according to interpersonal style and performance styles. The factor analysis results of their study supported two priori clusters of dimensions. Similarly, in the study of Damitz et al. (2003), PCA revealed that assessors

were able to distinguish between performance and interpersonally related dimensions.

The findings of correlations in this study were similar to previous research, particularly the within-exercises correlations. Whilst there was some evidence of convergent validity, there was very little evidence of discriminant validity for different dimensions. The above findings seem to support those of Lievens (1998), in which ratings of the same dimensions from across exercises correlated highly (evidence of stronger convergent validity), while ratings of different dimensions within the same exercise had high correlations (evidence of poor discriminant validity). Greyling et al. (2003) when investigating the construct validity of ACs, found no evidence of discriminant validity and very little convergent validity based on only two dimensions (i.e. motivating others and analytical thinking).

Joiner (2000) stipulated that assessors must classify and evaluate behaviour exhibited by the AC participants into meaningful categories of dimensions. This process according to Thornton (1992, p. 138), is the most controversial as assessors classify behaviour in a judgmental way. Robie et al. (2000) identified two major evaluation approaches, i.e. the within exercise rating (WER) approach, whereby each dimension was rated upon completion of each exercise, and the across-exercise rating (AER) approach. In the latter approach, assessors derive their ratings after all the exercises have been completed and ratings are based on performance in all exercises. Assessors' ability to rate participants is crucial and may have a significant impact on the construct validity, selection decision, feedback, scope of training and development of the candidates. The halo effect as a rating error has been associated particularly within exercise ratings, resulting in lack of discriminant validity. Kolk et al. (2004b) found in their study that rating errors are consistently responsible for the lack of AC construct validity. In fact, both in Kolk et al. (2004b) and Arthur and Woehr (2003), when assessors rated only one dimension the WER correlations decreased.

The rating approach used in the present study was the AER, which focused on final dimension ratings and therefore the possibility of bias and rating errors were limited. Heindl (1997) pointed out that dimensions of managerial

performance are not necessarily mutually exclusive and therefore overlapping may occur. Given the overlap that is evident in dimensions measured in the present study, one can therefore conclude that the high correlations among dimensions measured in the same exercise is a result of assessors being unable to differentiate between the dimensions measured. Furthermore, the lack of discriminant validity in the present study may be a result of the number of dimensions measured; implying that reducing the number of dimensions may cause a decrease in the correlations of WER dimension ratings. Because of a large number of dimensions that the assessor had to observe in the present study, assessors may have experienced cognitive overload.

The process of rating candidate performance demands cognitive complexity on the part of the assessor. Moreover, according to Lowry (1997), vague definitions of dimension and their numbers in exercises may confuse assessors in terms of differentiating between them. This will result in a limited capacity for the assessor to process information and consequently inability to meet the cognitive demands called for by the AC process. In addition, the assessors may have struggled with long definitions of the dimensions under study, resulting in assessors being unable to differentiate between the dimensions. To avoid these problems, Lievens (1998) suggested choosing dimensions that are conceptually distinct and unrelated. Some researchers have suggested reducing the number of dimensions in order to increase discriminant validities (e.g. Arthur & Woehr, 2003). Moreover; to overcome assessor information overload, Donahue et al. (1997); Kauffman et al. (1993); Lance et al. (2000); Lievens (2002) and Robie et al. (2000) proposed collapsing dimensions to a more manageable number. When dimensions were collapsed into smaller distinct sets of dimensions, evidence of construct validity was reported in Arthur et al. (2003) study. In fact, Lievens and Conway (2001) suggested that collapsing dimensions leads to improvements in construct validity of ACs, reducing cognitive overload as well as improving the quality of ratings.

The extent to which dimensions measured were transparent to participants in the current study is not known. Kleinmann (1993) suggested that when there is lack of transparency about dimensions, the probability that participants will

select incorrect action strategies increases, which will influence the construct validity results of the particular AC. According to Kleinmann et al. (2011) the ability to identify criteria (ATIC) being measured influences behavior and improves the validity of a selection procedure. In fact, in their study it was found that both the candidate's performance and the targeted dimensions improved when dimensions were transparent to the candidates (Kleinmann et al., 2011). Given the above, the lack of discriminant validity in the AC in this study could have been affected if the dimensions were not transparent to the participants.

Assessors are to receive extensive training on the process of observing and classifying candidates' behaviour during the AC. According to Joiner (2000), therefore training of assessors implies not only reliability of the AC, but also the impact it has on construct validity. Although it seems highly unlikely for training to have accounted for the lack of discriminant validity in the present study, their experience in assessment could have influenced assessor ratings. Among the assessors used in the present study, one was a psychologist while the rest came from various business units of the organisation. To improve discriminant validity, Lievens (2001) recommended the inclusions of the organisation's values, norms and personal quality in the assessor training. In the light of the above, the participating organisation's norms and values did not form part of the assessor training, but the dimensions measured were derived from the organisation's values and strategies. Thus, these dimensions are viewed as qualities that sustain the organisation's competitive advantage. Therefore; the impact of assessor training on the lack of construct validity in the present study cannot be explained by or ascribed to poor assessor training.

Another factor that could have had an impact on the lack of construct validity of the present AC is the type of assessors used. As mentioned above, only one of the four assessors was a psychologist. Jackson et al, (2005) suggested the use of psychologists as opposed to managers, as the use of managers demonstrates worse convergent validity than the use of psychologists. In addition, Sagie and Magnezy (1997) found evidence of the

influence of AC construct validity in relation to the type of assessor used: manager vs. psychologist.

As a consequence of the kind of assessors used in the current study, such as partners or directors and business unit managers, assessor schemata, including their characteristics and demographics, can be regarded as additional contributory factors to the lack of discriminant validity of this study. Such schemata resulting in biased ratings. Some of the assessors in the present study not only had knowledge of the firm's strategy and values and consequently assumptions about the ideal candidate for the job, but also experience in the target job and prior knowledge of the candidates. Their observations and ratings could therefore be influenced by their schemata. Heindl (1997) and Thornton and Rupp (2006) believed that assessors are most likely to pay attention to what is consistent with their assumptions and ignore divergent information. Moreover, when Bartels and Doverspike (1997) studied assessor personality characteristics and their impact on the leniency of ratings, they found that assessors with high intelligence and sensitivity tended to be more lenient in their ratings. The next section highlights the limitations of the present study.

4.3 LIMITATIONS

Various limitations to the present study were identified. These relate to the standardisation of the AC, the sample size, the age of data collected and assessors used. These limitations are highlighted in the next paragraphs.

The first limitation relates to procedural consideration of the AC under study. Since this is not a manipulated study, researchers were not involved in the setting of guidelines for AC administration and the scoring procedures. Therefore; the extent of standardisation procedures is unknown. The lack of standardisation in the AC method in organisations, assessors and administration may have an impact on the construct validity of ratings and the AC method itself (Kaufman et al., 1993). Secondly, the sample size (n=138) used in this study is relatively small. Because a larger sample is necessary to

generalise research findings, particularly in AC construct validity, the results of the present study cannot be generalised. Moreover, the sample used in this study used ratings from a single AC of one organisation and the dimensions reflected the organisation's values and strategies. Moreover, only a small number of assessors were used. These findings therefore, cannot be generalised to other ACs and across other organisations. Given the second limitation of this study, these findings should be interpreted with caution and replication studies using larger samples are necessary in order to generalise such findings.

A third limitation of this study relates to the differences in age of the data used and the assessors. The data collected and analysed ranged from 2006 - 2010. Although the dimensions are the same, it is not known whether the training method used was the same for all assessors and very little is known about the AC participants. In the next section recommendations for practice and future research are presented.

4.4 RECOMMENDATIONS

Recommendations are firstly made in relation to the users of ACs and secondly for future research. Continued research is needed into the construct validity of ACs, particularly the design and methodological characteristics as opposed to the AC as a measurement instrument. Therefore, the following recommendations are made in terms of the assumption that the AC construct validity dilemma has dire implications if the results are used for training or development, as well as a basis on which selection decisions are made. Given the above, the following paragraphs provide recommendations for both HR and future research.

4.4.1 Recommendations for practice

It is important to mention that organisations should design and implement ACs that accurately reflects the constructs that they are intended to measure.

Recommendations for practice are made in terms of the dimensions, assessors, simulation exercises and rating process, including the observation, evaluation and integration process.

4.4.1.1 Assessment centre dimensions

As mentioned earlier, a large number of dimensions pose cognitive challenges to the assessors. To improve the construct validity in AC practice, firstly, a smaller number of dimensions are recommended, especially when the AC is used for selection. Secondly, having only a few conceptually distinct dimensions so that assessors can differentiate between them more clearly is recommended. The above recommendations will assist in reducing the cognitive demands that assessors experience during observation and classifying of behaviours demonstrated by the candidates. In fact, Thornton (1992) suggests the use of no more than seven dimensions. Furthermore, organisations should ensure that the chosen dimensions are clearly defined and relatively unrelated. In the light of legislation and regulations (e.g. the LRA, EEA and Health Professions Act) governing the practice of assessments in South Africa, it is important to ensure that assessments are valid. Therefore, in relation to assessments, organisations should ensure that the dimensions chosen are job-relevant and that they measure the constructs that they claim to measure.

4.4.1.2 Assessors

Recommendations are made here on the type of assessors used and the training of assessors. Firstly, a healthy balance between psychologists and managers as assessors is recommended, as well as the involvement of a psychologist in designing the AC. Although having as many assessors as there are dimensions is costly; increasing the number of assessors may improve the construct validity of the AC under study. Assessor training is important in improving the construct validity of ACs to limit exercise variance. The organisations should focus on the quality of such training of assessors. Assessors should be trained and familiarised with the dimensions measured and the purpose of an AC should be articulated clearly to assessors. In fact,

Arthur and Woehr (2003) found improved construct validity evidence when the purpose of an AC was made clear to the assessors.

4.4.1.3 Rating process

As mentioned earlier, assessments in South Africa occur in a multicultural environment, therefore assessors should be sensitised in observing, evaluating and rating behaviours demonstrated by candidates in order to avoid rating errors (e.g. leniency and the halo effect) to ensure fair and unbiased judgments of candidates. A rotational system for assessors is further recommended to minimise rating biases. Lastly, an across-exercise rating approach is recommended to improve both the convergent and the discriminant validity of ACs.

4.4.2 Recommendations for future research

Firstly; a replication of similar studies is recommended as well as continuous research in the construct validity of ACs. Further research is required into assessor training approaches in order to standardise observations, evaluation and the integration processes. Secondly, research in the use of a behavioural checklist and its impact on construct validity is necessary. In fact, Donahue et al. (1997) investigated the effect of using behavioural checklists versus using graphic rating scales on the construct validity of ACs, using MTMM and confirmatory factor analysis, and found improved discrimination validity results.

As stated earlier, assessment practices in South Africa are regulated because of the multicultural environment in which the practice occurs. Therefore, a third research implication is continuous validity studies of ACs, pertaining in particular to the sources of gender differences in observed dimension ratings, namely; to investigate latent differences in constructs that are being measured. For example, Goldstein et al., (1998), Pulakos et al. (1996) when studying the impact of subgroup differences in the construct validity of ACs, found significant differences in overall assessment ratings which were in

favour of women participants. Similarly, Dean et al. (2008) found significant latent mean differences favouring women on the overall communication and interaction dimensions.

4.5 CONCLUSION

This chapter addressed the last part of this study. Firstly, conclusions were formulated; in which there was no evidence of discriminant validity, while some level of convergent validity was found. Limitations pertaining to the standardisation of the AC, the sample size, the age of data collected and assessors were highlighted. Lastly, recommendations for the practice of ACs and for future research were made.

REFERENCES

- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment centre dimensions. *Personnel Psychology, 56*, 125-153.
- Arthur, W., Day, E. A. & Woehr, D. J. (2008). Mend it, don't end it; an alternative view of assessment centre construct – related validity evidence. *Industrial and Organisational Psychology, 1*, 105-111.
- Arthur, W. & Woehr, D. J. (2003). The construct-related validity of assessment centre ratings: A review and meta-analysis of methodological factors. *Journal of Management, 29*(2), 231-258.
- Arthur, W., Woehr, D.J. & Maldegen, R. (2000). Convergent and discrimination validity of assessment centre dimensions: A conceptual and empirical re-examination of the assessment centre construct-related validity paradox. *Journal of Management, 26*(4). 813-835
- Bartels, L. K., & Doverspike, D. (1997). Assessing the assessor: The relationship of assessor personality and leniency in assessment centre ratings. *Journal of Social Behavior and Personality, 12*(5), 179-190.
- Bartram, D. (2005). The great eight competencies: A criterion-centric approach to validation. *Journal of Applied Psychology, 90*(6), 1185-1203.
- Bartram, D. (2004). Assessment in organisations. *Applied Psychology: An International Review, 53*(2), 237-259.
- Bedell, M.R. (2001). Personality testing: controversial no more. Keynote address presented at the 4th annual conference of the Society of Industrial Psychology, Pretoria.
- Bergh, Z. C. & Theron, A. L. (1999). *Psychology in the work context*. Western Cape: International Thomson Publishing.
- Biesheuvel, S. (1959). *Race, Culture and Personality*. South African Institute of Race Relations. Johannesburg.
- Biesheuvel, S. (1959). *The measurement of intelligence and aptitudes of African Peoples*. National Institute for Personnel Research, C.S.I.R. Johannesburg.

- Biesheuvel, S. (1954). The measurement of occupational aptitudes in a multi-racial society. *Occupational Psychology*, 52, 289-296.
- Boam, R. & Sparrow, P. (1992). *Designing and achieving competency: A competency-based approach to developing people and organisations*. London: McGraw-Hill Book Company Limited.
- Bowler, M.C. & Woehr, D. (2006). A meta-analytical evaluation of the impact of dimension and exercise factors on assessment centre ratings. *Journal of Applied Psychology*, 91(5), 1114-1120.
- Burnett, M. & Dutsch, J.V. (2006). Competency-based training and assessment centre: Strategies, technologies, process and issues. *Advances in Developing Human Resources*, 8(2), 141-143.
- Cadwell, C., Thornton, G. C. & Gruys, M.L. (2003). Ten classic assessment centre errors: Challenges to selection validity. *Public Personnel Management*, 32(1), 73-88.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81-105.
- Carrel, M.R., Elbert, N.F., Hatfield, R.D., Grobler, P.A., Marx, M. & Van Der Schyf, S. (1996). *Human resource management in South Africa*. Cape Town: Pearson Education South Africa.
- Casio, W. F. (1998). *Applied psychology in human resource management*. London: Prentice Hall.
- Charoux, E. (1991). Assessment centres offer a fair and accurate way to select and develop talent. *Management Today*, 7(2), 28-29.
- Charoux, E. (1990). Assessment centres: A route to black advancement. *HRM*, 15(3), 20-28.
- Charoux, E. (1990). *Coaching the Black Manager*. Pretoria: Juta and Company.
- Cheetham, G. (1996). Towards a holistic model of professional competence. *Journal of European Industrial Training*, 20(95), 20-30.

- Chen, H. (2006). Assessment centre: A critical mechanism for assessing HRD effectiveness and accountability. *Advances in Developing Human Resources*, 8(2), 247-264.
- Chen, H. & Naquin. S.S. (2006). An integrative model of competency development, training design, assessment centre and multi-rater assessment. *Advances in Developing Human Resources*, 8(2), 265-282.
- Claassen, N.W.C. (1997). Cultural differences, politics and test bias in South Africa. *European Review of Applied Psychology*, 47(4) 297-307.
- Cockerill, T., Hunt, J. & Schroder, H. (1995). Managerial competencies: fact or fiction? *Business Strategy Review*, 6(3), 1-12.
- Connelly, B. S., Ones, D. S., Ramesh, A. & Goff, M. (2008). A pragmatic view of assessment centre exercises and dimensions. *Industrial and Organisational Psychology*. 1, 121-124.
- Damitz, M., Manzey, D., Kleinmann, M. & Severin, K. (2003). *Applied Psychology: An international Review*, 52(2), 193-212.
- Dayan, K., Fox, S. & Kasten, R. (2002). Entry-level police candidate assessment centre: An efficient tool or a hammer to kill a fly? *Personnel Psychology*, 55(2), 827-849.
- Dean, M.A., Roth, P.L. & Bobko, P. (2008). Ethnic and gender subgroup differences in assessment centre ratings: A meta analysis. *Journal of Applied Psychology*, 93(3), 685-691.
- De Lange, M., Fourie, L. & Van Vuuren, L.J. (2003). Reliability of competency-based, multi-dimensional, multi-rater performance dimensions. *Journal of Industrial Psychology*, 29(2), 39-48.
- Donahue, L. M., Donald, M., Truxillo, J .M. & Gerrity, C.M. (1997). Assessment centre construct validity and behavioral checklists: Some additional findings. *Journal of Social Behavior and Personality*, 12(5), 85-108.
- Easterby-Smith M., Thorpe R. & Holman. (1996). Using repertory grids in management. *Journal of European Industrial Training*, 20(3), 3–30.
- Employment Equity Act, number 55 of 1998. *Government Gazette*. Republic of South Africa.

- Erasmus, P.F., & Arumugam, S. (1997). ABLE: The key to adequate job performance. *People Dynamics*, 16(9), 38-41.
- Fick, M.L. (1929). Intelligence tests results of poor white, native (Zulu), coloured and Indian school children and the educational and social implications. *South African Journal of Science*, 25, 904-920.
- Fleenor, J.W. (1996). Constructs and developmental assessment centres: Further troubling empirical findings. *Journal of Business and Psychology*, 10(3), 319-334.
- Fletcher, S. (1997). *Competence-based assessment techniques*. London: Koga Page.
- Foxcroft, C. (2004). Planning a psychological test in the multicultural South African context. *SA Journal of Industrial Psychology*, 30 (4), 8-15.
- Foxcroft, C. & Roodt, G. (2001). *An introduction of psychological assessment in the South African context*. Cape Town: Oxford University Press.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. & Bentson, C. (1987). Meta-analysis of assessment centre validity. *Journal of Applied Psychology*, 72(3) 493-51.
- Gregory, J. R. (2004). *Psychological testing, history, principles and applications*. Boston: Pearson International.
- Greyling, L., Visser, D. & Fourie, L. (2003). Construct validity of competency dimensions in a team leader assessment centre. *SA Journal of Industrial Psychology*, 29(2), 10-19.
- Goldstein, H. W., Yusko, K. P., Braverman, E. P., Smith, D. B., & Chung, B. (1998). The role of cognitive ability in the subgroup differences and incremental validity of assessment centre exercises. *Journal of Personnel Psychology*, 51, 357-374.
- Goldstein, H., Yusko, K.P. & Nicolopoulos, V. (2001). Exploring Black-White subgroup differences of managerial competencies. *Personnel Psychology*, 54, 783-807.

- Haaland, S. & Christiansen N.D. (2002). Implications of Trait-Activation Theory for evaluating the construct validity of assessment centre ratings. *Personnel Psychology*, *55*, 137-163.
- Hagan, C. M., Konopaske, R., Bernardin, H.J. & Tyler, C. (2006). Predicting assessment centre performance with 360-degree, top-down, and customer-based competency assessment. *Human Resource Management*, *45*(3), 357-390.
- Hagger, P., Gonczi, A. & Athanasou, J. (1994). General issues about assessment of competence. *Technology and Evaluation in Higher Education*, *19*(1), 3-16.
- Harel, G. H., Arditi-Vogel, A. & Janz, T. (2003). Comparing the validity and utility of behavior descriptions interview versus assessment centre ratings. *Journal of Managerial Psychology*, *18*(2), 94-104.
- Harter, J.K. & Schmidt. F. L., (2008). Conceptual versus empirical distinctions among constructs: Implications for discriminant validity. *Industrial and Organisational Psychology*, *1*, 36-39.
- Health Professions Act, number 56 of 1974. Health Professions Council of South Africa. Pretoria.
- Heindl, W.A. (1997). *Assessment centres: A comparison of trait ratings and task ratings for predicting managerial performance*, Unpublished Masters dissertation. Pretoria: UNISA.
- Heinsman, H., De Hoogh, A.H.B., Koopman, P.L. & Van Muijen, J.J. (2007). Competencies through the eyes of psychologists: A closer look at assessing competencies. *International Journal of Selection and Assessment*, *15*(4), 412-427.
- Hermelin, E., Lievens, F. & Robertson, I.T. (2007). The validity of assessment centres for the prediction of supervisory performance ratings: A meta-analysis. *International Journal of Selection and Assessment*, *15*(4) 405-411.
- Howard, A. (2008). Commentary: Making assessment centres work the way they are supposed to. *Journal of Industrial and Organisational Psychology*, *1*, 98-104.

Howard, A. (1997). A reassessment of assessment centres: Challenges for the 21st century. *Journal of Social Behavior and Personality*, 12(5), 13-52.

Hunter, J.E., & Hunter, R. (1984). Validation and utility of alternate predictors of job performance. *Psychological Bulletin*, 96, 72-98.

Jackson, K.J.R., Barney, A.R., Stillman, J.A. & Kirkley, W., (2007). When traits are behaviors: The relationship between behavioral responses and trait-based overall assessment centre ratings. *Human performance*, 20(94), 415-432.

Jackson, D.J.R., Stillman, J.A. & Atkins, S.G. (2005). Rating tasks versus dimensions in assessment centres: A psychometric comparison. *Human Performance*, 18(3), 213-241.

Jackson, D.J.R., Atkins, S.G., Fletcher, R.B. & Stillman, J.A. (2005). Frame of reference training for assessment centres: effects on interrater reliability when rating behaviors and ability traits. *Public Personnel Management*, 34(1), 17-30.

Jansen, W.P. & Stoop, B.A.M. (2001). The dynamics of assessment centre validity: results of a 7-year study. *Journal of Applied Psychology*, 86(4), 741-753.

Joiner, D.A. (2002). Assessment centres: What's new? *Public Personnel Management*, 31(2), 179-183.

Joiner, D. A. (2000). Guidelines for ethical consideration for assessment centre operations: International Task Force on Assessment Centre Guidelines. *Public Personnel Management*, 29(3), 315-331.

Jubb, R. & Robotham, D. (1997). Competencies in management development: Challenging myths. *Journal of European Industrial Training*, 21(5), 171-175.

Juceviciene, P. & Lepaite, D. (1995). *Competence as derived from activity: the problem of their level correspondence*. Retrieved 29 September 2009, from: <http://www.education.ktu/evaco/index.html>.

Kaufman, J.R., Jex, S.M., Love, K. G. & Libkuman, T.M. (1993). The construct validity of assessment centre performance dimensions. *International Journal of Selection and Assessment*, 1(4), 213-223.

Khumalo, R. (1999). The management of human resources in successful companies - an African context. *Journal of Industrial Psychology*, 25(1), 1-6.

Kim, S. & Hong, J. (2005). The relationship between salesperson competencies and performance in the Korean Pharmaceutical Industry. *Management Review*, 16(2), 259-268.

Kleinmann, M. (1993). Are rating dimensions in assessment centres transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*, 78(6), 988-993.

Kleinmann, M., Ingold, P.V., Lievens, F., Jansen, A., Melchers, K.G., & Konig, C. J. (2011). A different look at why selection procedures work: the role of candidates' ability to identify criteria. *Organisational Psychology Review*, 1(2), 128-146.

Kleinmann, M. & Koller, O. (1997). Construct validity of assessment centres: appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behavior and Personality*, 12(5), 65-84.

Kolk, N.J., Born, M., & Van Der Flier, H. (2004a). A triadic approach to the construct validity of the assessment centre. *European Journal of Psychological Assessment*. 20(3), 146-156.

Kolk, N.J., Born, M. & Van Der Flier, H. (2004b). Impact of common rater variance on construct validity of assessment centre dimension judgments. *Human Performance*, 15 (9), 325-337.

Kriek, H.J. & Charoux, J.A.E. (1994). The assessment centre: Testing the fairness hypothesis. *Journal of Industrial Psychology*, 20(2), 21-25.

Kudisch, J.D., Ladd, R.T. & Dobbins, G.H. (1997). New evidence on the construct validity of diagnostic assessment centres: The findings may not be so troubling after all. *Journal of Social Behavior and Personality*, 6(3), 141-152.

Labour Relations Act, number 66 of 1995. Republic of South Africa.

- Lance C.E. (2008). Why assessment centres do not work the way they are supposed to. *Journal of Industrial and Organisational psychology*, 1, 84-97.
- Lance, C.E., Foster, M.R., Nemeth, Y.M., Gentry, W.A., & S., Drollinger, (2007). Extending the nomological network of assessment centre construct validity: prediction of cross-situationally consistent and specific aspects of assessment centre performance. *Human Performance*, 20(40), 345-362.
- Lance, C. E., Foster, M. R., Thoresen, J. D. & Gentry, W. A. (2004). Assessor cognitive processes in an operational assessment centre. *Journal of Applied Psychology*, 89(1), 22-35.
- Lance, C. E., Newbolt, W. H., Gatewood, R. D., Foster, M. R., French, N. R. & Smith, D. E. (2000). Assessment centre exercise factors represent cross-situational specificity, not method bias. *Human Performance*, 13(4), 323-353.
- Lange, D.M., Fourie, L., & Van Vuuren, L.J. (2003). Reliability of competency-based, multi-dimensional, multi-rater performance ratings. *Journal of Industrial Psychology*, 29(2), 39-48.
- Lawler, E. E. (1994). From job-based to competency-based organisations. *Journal of Organisational Behavior*, 15(4), 3-15.
- Lievens, F. (2008). What does exercise-based assessment really mean? *Industrial and Organisational Psychology*, 1, 112-115.
- Lievens, F. (2002). Trying to understand the different pieces of the construct validity puzzle of assessment centres: An examination of assessor and assessee effects. *Journal of Applied Psychology*, 87(4), 675-686.
- Lievens, F. (2001). Assessors and use of assessment centre dimensions: A fresh look at troubling issues. *Journal of Organisational Behavior*, 22, 2003-221.
- Lievens, F. (1999). Development of a simulated assessment centre. *European Journal of Psychological Assessment*, 15(2), 117-126.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centres: A review. *International Journal of Selection and assessment*, 6(3) 141-142.

- Lievens, F., Chasten, C. S., Day, E. A. & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment centre convergent and discriminant validity. *Journal of Applied Psychology, 91*(2), 247-258.
- Lievens, F. & Conway, J.M. (2001). Dimension and exercise variance in assessment centre scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86*(6), 1202-1222.
- Lievens, F., De Fruyt, F. & Van Dam, K. (2001). Assessor's use of personality traits in descriptions of assessment centre candidates: A five factor model perspective. *Journal of Occupational Psychology, 74*, 623-636.
- Lievens, F. & Klimoski, R. J. (2001). Understanding the assessment centre process: Where are we now? *International Review of Industrial and Organisational Psychology, 16*, 246-286.
- Lievens, F. & Thornton, G.C. (2005). *Assessment centres: Recent developments in practice and research*. London: Blackwell Publishing.
- Lievens, F. & Van Keer, E. (2001). Short research note: A comparison of different models. *Journal of Occupational and Organisational Psychology, 74*, 373-378.
- Lord, R.G. & Maher, K.J. (1990). Alternative information-processing models and their implications for theory, research, and practice. *Academy of Management Review, 15*, 9-28.
- Lowry, P. (1997). The assessment centre process: New directions. *Journal of Social Behavior and Personality, 12*(5), 53-62.
- Magwaza, S. (1995). Humanising and democratising assessments. *People Dynamics, 13*(5), 12-15.
- Mauer, K.F. (2000). Psychological test use in South Africa. Retrieved 24 June 2009: <http://sunsite.wits.ac.za/conference/psychology/pai1>.
- Mbokazi, B., Visser, D. & Fourie, L. (2004). Management perceptions of competencies essential for middle managers. *South African Journal of Industrial Psychology, 30*(1), 1-9.

- McClelland, D.C. (1973). Testing for competence rather than for "intelligence." *American Psychologist*, 28, 1-14.
- Melancon, C.S. & Williams, M. (2006). Competency-based assessment centre design: A case study. *Human Resources*, 8(2), 283-315.
- Melchers, K.G. & Konig, C.J. (2008). It is not yet time to dismiss dimensions in assessment centres. *Industrial and Organisational Psychology*, 1, 125-127.
- Meriac, J.P., Hoffman, B.J., Woehr, D.J. & Fleisher, M.S., (2008). Further evidence of validity of assessment centre dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93(9), 1042-1052.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*. 35(11), 1012-1027.
- Meyer, T.S.P. (1996). *Creating Competitiveness through Competencies: Currency for the 21st Century*. Randburg: Knowledge Resources.
- Meyer, T.S.P. (1995). A framework for the use of competencies for achieving competitive advantage. *South African Journal of Business Management*, 27(4), 1-9.
- Mirabile, R.J. (1997). Everything you wanted to know about competency modelling. *Training and Development*, 51, 73-77.
- Monk, A. (2001). *The development of a competency model for auditors working in a professional services environment*. Unpublished Masters dissertation, Pretoria: Unisa.
- Moses, J. (2008). Assessment centres work, but for different reasons. *Industrial and Organisational Psychology*, 1, 134-136.
- Mouton, J. (2005). *How to succeed in your Master's and Doctoral Studies: A South African Guide and Resource Book*. Pretoria: Van Schaik Publishers.
- Muchinsky, P., Kriek, H.J. & Schreuder, D. (2002). *Personnel Psychology*. Cape Town: Oxford, Southern Africa.

- Naquin, S.S. & Holton, E. (2006). Leadership and managerial competency models: A simplified process and resulting model. *Advances in Developing Human Resources*, 8(2), 144-165.
- Nobre, M. (2005). *The 32i (Occupational Personality Questionnaire Version 1) as a predictor of theft in a financial institution*. Unpublished Masters dissertation. Stellenbosch: University of Stellenbosch.
- Olivier, B.H. (1994). *The assessment centre as a predictor of managerial training results*. Unpublished Masters dissertation. Pretoria: UNISA.
- Paterson, H. & Uys, K. (2005). Critical issues in psychological test use in South African workplace. *SA Journal of Industrial Psychology*, 31(3), 12-22.
- Potgieter, T.E. & Van der Merwe, R.P. (2002). Assessment in the workplace. A competency-based approach. *South African Journal of Industrial Psychology*, 28(1), 60-66.
- Pulakos, D.E., Schmitt, N. & Chan, D., (1996). Models of job performance ratings: An examination of ratee race, ratee gender, and rater level effects. *Human Performance*, 9(2), 103-119.
- Reilly, R.R., Henry, S. & Smither, J.W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment centre dimensions. *Personnel Psychology*, 43, 71-84.
- Robie, C., Osburn, H.G., Morris, M.A., Etchegaray, J. M. & Adams, K.A. (2000). Effects of the rating process on the construct validity of assessment centre dimension evaluations. *Human Performance*, 13(4), 355-370.
- Rupp, D.E., Thornton, G.C. & Gibbons, A.M. (2008). The construct validity of the assessment centre method and usefulness of dimension as focal constructs. *Industrial and Organisational Psychology*, 1, 116-120.
- Sagie, A. & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories, and assessment centre construct validity. *Journal of Occupational and Organizational Psychology*, 70, 103–108.
- Saunders, E. (2000). *Assessing human competence: Practical guidelines for South African managers*. Johannesburg: Knowledge Resources.

- Schippmann, J. S., Ash, R.A., Batista, M., Carr, L., Eyde, L., Hesketh, B., Kehoe, J., Pearlman, K., Prien, E., & Sanchez, J. I. (2000). The practice of competency modelling. *Personnel Psychology*, *53*, 703-740.
- Schlebusch, S. & Roodt, G. (2008). *Assessment centres: Unlocking potential for growth*. Randburg: Knowres Publishing.
- Schleicher, D.J., Day, D.V., Mayes, B.T. & Riggio, R.E. (2002). A new frame for frame-of- reference training: Enhancing the construct validity of assessment centres. *Journal of Applied Psychology*, *87*(4), 735-746.
- Schmidt, C. (2006). Validity as an action concept in IO Psychology. *Journal of Industrial Psychology*. *32*(4), 59-67.
- Schmidt, F.L. & Hunter, J.E. (1998). The validity and utility of selection in personnel psychology: Practical and theoretical implications of 85 years of research. *Psychological Bulletin*, *12*(2), 262-274.
- Schneider, R.G., & Schmitt R.M. (1992). The relation of cognitive ability and personality traits to assessment centre performance. *Journal of Applied Social Psychology*, *30*(7), 1474–1491.
- Schroder, H.M., (1989). *Managerial competence – the key to excellence*. Iowa: Kendall/Hunt.
- Semark, T. M. & South, P. (1996). A framework for the use of competencies for achieving competitive advantage. *South African Journal of Business Management*, *27*(4), 1-11.
- SHL, (2007) *Defining competencies that underpin successful job performance*, online competency profiler, [Brochure] Pretoria.
- SHL. (2002). Construct validity study, study number V016. Critical reasoning test battery. Retrieved 7 September 2008 on <http://www.shl.research.co.za> (internet document).
- SHL. (2001). *Guidelines for best practice in occupational assessment in South Africa*. Pretoria.
- SHL. (1993). *Inventory of management competencies*. London

- Shore, T.H., Thornton, G.C., & Shore, L. (1990). Construct validity of two categories of assessment centre dimension ratings. *Personnel Psychology*, 43, 101- 116.
- SIOPSA, (2005). *Guidelines for the validation and use of assessment procedures for the workplace*. Pretoria.
- Spangenberg, H.H., Esterhuysen, J.H., Briedenhann, J.E., & Calitz, C.J. (1989). Validation of an assessment centre against BARS: an experience with performance related criteria. *Journal of Industrial Psychology*, 15(2), 1-10.
- Spencer, L.M. & Spencer, S.M. (1993). *Competence at work: models for superior performance*. New York: John Wiley and Sons, Inc.
- Spychalski, A.C., Quinones, M.A., Gaugler, B.B., & Pohley, K. (1997). A survey of assessment centre practices in organizations in the United States. *Personnel Psychology*, 50, 71-90.
- Swan, W.S and Wilson, L.E. (2007). *Ready to Use Performance Appraisals*. New Jersey: John Wiley & Sons, Inc.
- Tate, W. (1995). *Developing Managerial Competence: A critical guide to methods and materials*. England: Gower.
- Taylor, I. (2007). *A practical guide to assessment centres and selection methods*. London: Kogan Page.
- Taylor, T.R. (1990). *Are you testing fairly*. Human Science Research Council. Pretoria.
- Taylor, T.R. & Boeyens, J.C.A. (1991). The comparability of scores of blacks and whites on the South African Personality Questionnaire: An exploratory study. *South African Journal of Psychology*, 21(1), 1-11.
- Taylor, T.R., & Radford, E.J. (1986). Psychometric testing as an unfair labour practice. *South African Journal of Psychology*, 16, 79-86.
- Terre Blanche, M. & Durrheim, K. (2004). *Research in Practice: Applied methods for the social science*. Cape Town: UCT Press.
- Tett, R. P. & Burnett, D.D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology*, 88(3), 500-517.

- Tett, R.P. & Guterman, H.A. (2000). Situation trait relevance, trait expression, and cross situational consistency: Testing a principle of trait activation. *Journal of Research in Personality, 34*, 397-423.
- Theron, C. (2009). The diversity-validity dilemma: In search of minimum adverse impact and maximum utility. *South African Journal of Industrial Psychology, 35*(1), 1-13.
- Theron, D, & Roodt, G. (2000). Mental models as moderating variables in 360 degree competency assessments. *Journal of Industrial Psychology, 26*(2), 14-19.
- Thornton, G.C. (1992). *Assessment centres in human resource management*. Massachusetts: Addison-Wesley Publishing Company.
- Thornton, G.C. & Rupp, D.E. (2006). *Assessment Centres in Human Resource Management. Strategies for prediction, diagnosis and development*. New Jersey: Lawrence Erlbaum Associates, Publishers.
- Thornton, G.C., Tziner, A., Dahan, M., Clevenger, J. P. & Meir, E. (1997). Construct validity of assessment centre judgments: Analyses of the behavioral reporting method. *Journal of Social Behavior and Personality, 12*(5), 109-128.
- Trafford, V. & Leshem, S. (2008). *Stepping Stones to Achieving Your Doctorate by Focusing on Your Viva from the Start*. Boston: McGraw-Hill Open University Press.
- Van Der Vijver, F.J.R., & S., Rothmann (2004). Assessment in multicultural groups: the South African case. *SA Journal of industrial psychology, 30*(4), 1-7.
- Van Heerden, G. (2005). *The relationship between the Cognitive Process Profile (CPP) the Meyer-Briggs Type Inventory (MBTI)*. Unpublished Masters dissertation. UNISA: Pretoria.
- Van Rooyen, J. (2008). Training manual: Assessment centre methodology. Johannesburg.

Vathanophas V & J. Thai-Ngam. (2007). Competency Requirements for Effective Job Performance in the Thai Public Sector. *Contemporary Management Research*, 3(1), 45-70.

Weiss, T.B. & Hurtle, F. (1997). *Reengineering Performance Management*. St Lucie: Boca Raton, FL.

Whiddett, S. & Hollyforde, S. (2003). *A practical guide to competencies; how to enhance individual and organisational performance*. . London: Chartered Institute of Personnel Development.

Woodruffe, C. (1990). *Assessment centres: Identifying and developing competence*. London: Institute of Personnel Management.

Appendix A

Definitions of dimensions measured in the LADC

Dimension	Definition
1. <i>Client orientation</i>	Degree to which the candidate demonstrates orientation and identification with the client's needs and understands these needs.
2. <i>Applied judgment and insight</i>	The capacity to apply general and technical knowledge, insight and consideration of implications when analysing situations and making decisions.
3. <i>Business acumen</i>	General business alertness and insight so that person can be an effective business advisor to clients. Ability to understand correctly, interpret and apply financial, commercial and industrial vagaries and dynamics of commercial and industrial business operations.
4. <i>Knowledge-up-to-datedness</i>	Degree of taking a positive interest and action in developing personal technical knowledge and skills, either generally or specifically, in order to be able to provide excellent client service.
5. <i>Practice leadership</i>	Ability to coordinate and manage client, human, financial, information, technology and other resources to achieve the business units' goals to the benefit of clients.
6. <i>Winning business</i>	Degree of strength of marketing mentality, instinct, relation to and identification with

Dimension	Definition
	the marketing and selling of services.
<i>7. Product and service knowledge</i>	Knowledge of the firm's products and services and applicability thereof and willingness to promote these products and services to meet the client's needs
<i>8. Creativity and innovation</i>	The ability to create new and imaginative approaches to work-related issues. Identification of new approaches and willingness to question traditional assumptions.
<i>9. Written communication</i>	The ability to provide information that is accurate and reliable and to write reports or opinions precisely and concisely.
<i>10. Oral communication</i>	Degree to which the candidates demonstrates the ability to present orally, interactively to generate and maintain interest and attention of others.
<i>11. Personal impact</i>	Degree to which one is magnetic and attractive to others as a personality, regardless of whether one is extroverted or introverted in approach.
<i>12. Personal integrity</i>	The extent to which the candidate applies professional and personal standards of conduct in all he or she does and is consistent in maintaining these standards.
<i>13. Interpersonal skill</i>	Ability to interact and relate to others comfortably, easily and with respect both in work/business and social situations.
<i>14. Achievement drive</i>	The extent to which the person commits to personal and work goals with

Dimension	Definition
	determination and energy, constantly strives to improve self and performance and is realistically ambitious.
<i>15. Teamwork</i>	Degree to which candidate demonstrates, regardless of personal preference, the ability to generate teamwork approach and work effectively as part of a team.
<i>16. Respect for diversity</i>	Degree to which the candidate demonstrates awareness of and respect for diversity; is open to culturally challenging ideas and different perspectives.
<i>17. Negotiation</i>	Ability to attain objectives within defined parameters through flexibility of skills, as well as ensure other party/parties are reasonably satisfied with the outcome.
<i>18. Resilience</i>	Ability to tolerate and cope with mental and emotional stress and frustration in pursuit of goals or duties of position.
<i>19. Flexibility</i>	Mental and emotional ability to adapt to and adjust to changing demands and conditions
<i>20. Leadership</i>	Ability to inspire and motivate others to commitment and involvement; provide a vision; create a climate of mobility and positiveness.
<i>21. General applied judgment</i>	Ability to comprehend, analyse and make judgments, solve problems and jump from one issue to another without losing track of things.