

**The evaluation of a frame-of-reference training programme for
assessors of assessment centres**

G. Mulder,

Mini-dissertation submitted in partial fulfilment of the requirements for the degree
Magister Commercii in Industrial Psychology at the
North-West University (Potchefstroom Campus)

Supervisor: Dr. L.I. Jorgensen

Assistant Supervisor Prof. D. Meiring

May 2012

Potchefstroom

COMMENTS

The reader is reminded of the following:

- The editorial style as well as the references referred to in this mini-dissertation follow the format prescribed by the Publication Manual (6th edition) of the American Psychological Association (APA). This practice is in line with the policy of the Programme in Industrial Psychology of the North-West University (Potchefstroom) to use APA style in all scientific documents as from January 1999.
- The mini-dissertation is submitted in the form of a research article. The editorial style specified by the South African Journal of Industrial Psychology (which agrees largely with the APA style) is used, but the APA guidelines were followed in constructing tables.



1 Mei 2012

Hiermee verklaar ek, me *Cecilia van der Walt*, dat ek die taalversorging van die skripsie van me *Gerdi Mulder*, getitel *The Evaluation of a Frame-Of-Reference Training Programme for Assessors of Assessment Centres*, behartig het.

ME CECILIA VAN DER WALT

HOD, BA
Plus Taalversorging en vertaling op Hons-vlak,
Plus Akkreditering by SAVI vir Afrikaans
Registrasienommer by SAVI: 1000228

Epos-adres: ceciliavdw@lantic.net

Selnommer: 072-616 4943 (S)
Faksnommer: 086 578 1425

DECLARATION

I, Gerdi Mulder, hereby declare that this mini-dissertation titled, “The evaluation of a frame-of-reference training programme for assessors of assessment centres” is my own work and that the views and opinions expressed in this work are those of the author and relevant literature references as shown in the references.

I further declare that the content of this research will not be handed in for any other qualification at any other tertiary institution.

Gerdi Mulder

ACKNOWLEDGEMENTS

The author would like to thank:

- My family, friends, and everyone special in my life. Thank you for all the support, motivation and encouragement throughout this study and throughout my academic career.
- Dr Jorgensen, my supervisor, who motivated and inspired me throughout the entire study. Thank you for keeping me focused when I was distracted and for keeping me motivated through the highs and lows of this study. Thank you for believing in me and for standing by me through all the obstacles that we encountered. You are truly an inspiration with your optimism and faith.
- Prof. Meiring, who acted as my co-supervisor for this study. Thank you for providing me with opportunities and guidance when I needed it most. Your input and support is truly appreciated.
- Jesus Christ, who provided me with hope, energy and strength when I needed it most. To Him all the glory and praise for this study.
- Special thanks to all the participants of this study. Thank you for being dedicated throughout the training and providing me with valuable feedback and encouragement. This research would not have been possible without your interest and energy.
- Dr Alewyn Nel and Dr Carin Hill and Me. Marike Krugell who helped with the statistical analysis as well as interpretation of the data. I appreciate your investment and dedication.

TABLE OF CONTENTS

List of tables	viii
List of figures	ix
Abstract	x
Opsomming	xii

CHAPTER 1: INTRODUCTION

1.1	Problem Statement	1
1.2	Research objectives	7
1.2.1	General objective	7
1.2.2	Specific objective	7
2.1	Research approach	8
2.2	Research method	8
2.2.1	Literature review	8
2.2.2	Research participants	9
2.2.3	Measuring instruments	9
2.2.4	Research procedure	9
2.2.5	Statistical analysis	10
2.2.6	Ethical considerations	11
3	Chapter overview	11
4	Chapter summary	11
	References	12

CHAPTER 2: RESEARCH ARTICLE 1

15-43

CHAPTER 3: CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

3.1	Conclusions	44
3.2	Limitations of this research	47
3.3	Recommendations	49
3.3.1	Recommendations for the organisation	49
3.3.2	Recommendations for future research	50
	References	52

LIST OF TABLES

Article 1

Table	Description	Page
Table 1	Characteristics of the participants ($N=22$)	25
Table 2	The content, objectives and methodology of the FOR training programme for assessors of assessment centres	26
Table 3	The Cronbach's alphas (α) between the pre- and the post-test for the experimental and control group for the AC	30
Table 4	The difference between the pre- and post-test scores for the experimental group for the AC	31
Table 5	The difference between the pre- and post-test scores for the control group for the AC	33

LIST OF FIGURES

Article 1

Figure	Description	Page
Figure 1	The comparison of the pre- and post-test rating for the one-on-one simulation by the experimental group	32
Figure 2	The comparison of the pre - and post-test rating for the group discussion simulation by the experimental group	33

ABSTRACT

Title:

The evaluation of a frame-of-reference training programme for assessors of assessment centres

Key terms:

Assessment centres, assessors, training of assessors, frame-of-reference training

Assessment centres are one of the most effective selection processes. However, the biggest issue facing assessment centres is that of construct validity (Collins et al., 2003; Guion, 1998). A certain aspect that could affect the construct validity of assessment centres is assessor training; to improve the consistency of their judgments (Pell, Homer & Roberts, 2008). The assessors' levels of expertise play a significant role in the validity of the whole process (Jones & Born, 2008). It can therefore be said that the group of people that has the biggest impact on the whole assessment process, are the assessors (Schlebusch, 2008). When graduating, Industrial Psychology postgraduate students should be able to assist in any assessment setting in a variety of settings and organisations (HPCSA, 2010).

By implementing Frame-of-Reference training as an intervention for assessors, the construct as well as criterion validity could be influenced significantly (Lievens & Conway, 2001 & Schleicher, Day, Mayes & Riggio, 2002). Although international studies exist on Frame-of-Reference and assessor training, currently no such research exists for the South African context. The general aim of this research was to determine the effect of a Frame-of-Reference training programme for assessors of an assessment centre.

For this purpose a purposive sample of Industrial Psychology Honours students was used. They were randomly divided into a control group and an experimental group. Both groups received the same pre- and post-test, in the form of having to evaluate independent role-players based on predetermined criteria, by viewing approved recordings of typical assessment centre simulations. The Frame-of-Reference training programme was conducted over a three-day period. Practical sessions were also hosted for assessors to practice and

receive feedback on their newly obtained skills. The experimental group received the training between the pre- and post-tests. The comparison group only received the training after the post-test was completed, this ensured fair research practices. A quantitative research design was thus implemented.

Descriptive statistics, Cronbach alpha coefficients, Wilcoxon Signed Ranks and Mann-Whitney U -test were implemented to analyse the data. The descriptive statistics and paired t -tests confirmed that during the one-on-one and group discussion the ratings of the experimental group seemed to be statistically different and significant. However, the same result could not be reported for the presentation simulation. Overall, the frame-of-reference training had a positive impact on the assessor skills of the participants.

OPSOMMING

Titel:

Die evaluering van 'n verwysingsraamwerk-opleidingsprogram (frame-of-reference training programme) vir assessors van assesseringsentra

Sleuteltermes:

Takseersentrum/*assessment centres*, assessors/*assessors*, opleiding van assessors/*training of assessors*, verwysingsraamwerk-opleiding/*frame-of-reference training*

Takseersentrums is een van die effektiëste keuringsprosesse. Die grootste vraagstuk waardeur takseersentrums in die gesig gestaar word, is dié van konstrukgeldigheid (Collins et al., 2003; Guion, 1998). 'n Bepaalde aspek wat die konstrukgeldigheid van takseersentrums kan beïnvloed, is assessor-opleiding; om die konsekwentheid van hul takserings te verbeter (Pell, Homer & Roberts, 2008). Die assessors se kundigheidsvlak speel 'n betekenisvolle rol by die geldigheid van die hele proses (Jones & Born, 2008). Daar kan dus gesê word dat die groep mense wat die grootste impak op die hele assesseringsproses het, die assessors is (Schlebusch, 2008). Wanneer Bedryfsielkundestudente hul graad ontvang het, behoort hulle op nagraadse vlak daartoe in staat te wees om in enige assesseringomgewing in 'n verskeidenheid omgewings en organisasies te kan help (HPCSA, 2010).

Deur Verwysingsraamwerk-opleiding as 'n intervensie vir assessors te implementeer kon die konstruk sowel as kriteriageldigheid aansienlik beïnvloed word (Lievens & Conway, 2001 & Schleicher, Day, Mayes & Riggio, 2002). Alhoewel internasionale studies oor Verwysingsraamwerk en assessorsopleiding bestaan, bestaan daar tans geen sodanige navorsing vir die Suid-Afrikaanse konteks nie. Die algemene doel van hierdie navorsing was om die effek van 'n Verwysingsraamwerk-opleidingsprogram vir assessors van 'n takseersentrum te bepaal.

Vir hierdie studie is 'n doelbewuste steekproef bestaande uit Bedryfsielkunde Honneursstudente gebruik. Hulle is in 'n vergelykende groep en 'n eksperimentele groep ingedeel. Beide groepe het dieselfde voor- en natoets in die vorm van die evaluering van

onafhanklike rolspelers, gebaseer op voorafbepaalde kriteria, deur na goedgekeurde opnames van tipiese takseersentrum-simulasies te kyk. Die aanbieding van die Verwysingsraamwerk-opleidingsprogram het oor drie dae gestrek. Praktiese sessies is ook vir assessors gehou om te oefen en terugvoer te ontvang oor hul nuut verworwe vaardigheid. Die eksperimentele groep het die opleiding tussen die voor- en natoets ontvang. Die vergelykingsgroep het die opleiding eers nadat die na-toets afgehandel was, ontvang. Dit het regverdige navorsingspraktyke verseker.

Beskrywende statistiek, Cronbach alfa koëffisiënte en gepaarde *t*-toetse is geïmplementeer om die data te analiseer. Die beskrywende statistiek en gepaarde *t*-toetse het bevestig dat, tydens die een-tot een- en groepsbesprekings die waardebevestigings van die eksperimentele groep geblyk het statisties betekenisvol vir sekere bevoegdhede te wees. Geen bevoegdhede wat in die aanbiedingsimulasie gewys is, het egter 'n betekenisvolle verskil in waardebevestiging van die voor-toets na die na-toets getoon nie.

CHAPTER 1

INTRODUCTION

This mini-dissertation is presented in the form of an article regarding *The evaluation of a frame-of-reference training programme for assessors of assessment centres*. The article focuses on the effect and content of a frame-of-reference training programme on assessors of an assessment centre. This programme is specifically aimed at enabling postgraduate Industrial Psychology students to assist in various and diverse assessment settings by providing them with frame-of-reference training for assessment centres. Key words utilised in this research include *assessment centres, assessors, training of assessors, frame-of-reference training*. In this chapter, the problem statement and the research objectives (including the general and specific objectives) are discussed. Following this, the research method is explained and an overview is given of the chapters.

1.1 Problem statement

The Health Professions Council of South Africa (2010) expects a graduate Industrial Psychology student registered as a psychometrist to be able to assist in any assessment procedure in various diverse settings in organisations. In the field of personnel psychology, one of the most effective assessment techniques being implemented is that of an Assessment Centre (AC) (Lievens & Thornton, 2005). Dilchert and Ones (2009) and Lievens and Thornton (2005) point out that this technique is very appropriate for processes such as selection, recruitment as well as talent identification. The origins of ACs date back to World War II (WW II) (1939-1945) when the American and English armies respectively developed a simulation that would enable them to identify potential talented spies and officers (Howard, 2009). In 1974 the first assessment centre was introduced into South Africa by Douglas Bray and Bill Byham of the USA. The AC was used as a selection tool in the Edgars group and ever since then AC grew from strength to strength and were employed in various industries in South Africa (Meiring, 2008). The Assessment Centre Study Group (ACSG) was founded for the purpose of providing an opportunity for practitioners to annually exchange ideas and explore new ventures by viewing new research material.

The International Task Force on Assessment Centre Guidelines (2010) notes that the main objective of an Assessment Centre (AC) is to serve as a tool during selection processes for

identifying the most appropriate candidate for a certain position. These guidelines indicate the manner in which the most appropriate candidate can be identified, namely by, “during the process, employ[ing] multiple techniques and multiple assessors to produce judgements regarding the extent to which a participant displays selected competencies” (p.3).

Schlebusch (2008) further identifies various criteria in the South African context that should be present for a true AC. The most important characteristics are that multiple simulations should be utilised and various candidates should be observed during the different simulations. Multiple and competent observers should also then observe these candidates and they should observe, classify and evaluate behavioural constructs that should be exposed during the various simulations. Participants should also be notified about the main objective of an AC, being for selection. These simulation and other assessment procedures are specifically designed to extract certain behaviour in a candidate, in turn making it possible to observe and evaluate these behaviours. These observations are done by a group of assessors that have the task of observing and evaluating these candidates and award an appropriate rating which results in a recommendation for appointment (Goodstone & Lopez, 2001; Schlebusch, 2008).

Furthermore Lievens (2009), elaborates on the subject of assessors by stating that trained observers should be appointed to observe and evaluate these displayed behaviours by applicants taking part in the AC. These participants can be classified as assesseees that can be viewed as individuals that are measured in terms of their competencies by means of an assessment centre (International Task Force on Assessment Centre Guidelines, 2010; Lievens, Tett & Schleicher, 2009; Schleicher, Day, Mayes & Riggio, 2002).

Thornton and Mueller-Hanson (2004) maintain that one of the most commonly made mistakes in ACs is appointing either poorly trained assessors or unqualified assessors, although a variety of research has shown that ACs display good criterion-related validity (Arthur, Day, McNelly & Edens, 2003), predictive validity (Thornton, Murphy, Everest & Hoffman, 2000) and even, if expert assessors are being used, good inter-rater reliability (Lievens, 2002). However, one aspect that can be seen as a significant challenge of ACs is that of displaying good construct validity (Guion, 1998; Lievens, 2009). Various researchers Jones and Born (2008) and Pell, Homer and Robertson (2008) have speculated over the challenge of construct validity, assessor expertise and consistency in assessor judgements that could have a significant effect on the construct validity of the AC process. As stated

previously, the main objective of an AC is to select the most appropriate candidate for a certain position, and this is done by observing these candidates in a series of work-like simulations assessing them on a set of competencies.

The role of the assessor in an AC

The International Task Force on Assessment Centre Guidelines (2010) and Goodstone and Lopez (2002) postulate that an assessor can be seen as an individual that is trained to observe, record and classify behaviour and, from these observations, makes accurate judgements. The International Task Force for Assessment Centre Guidelines (2010) defines a recording of behaviour as follows: “A systematic procedure must be used by assessors to record specific behavioural observations accurately at the time of observation. This procedure might include techniques such as handwritten notes, behavioural observation scales or behavioural checklists” (p.4).

Schlebusch (2008) and Schleicher et al. (2002) further state that for these assessors to be able to adhere to the above-stated responsibilities they would require specialised training to develop these expert competencies. The South African Qualifications Authority (2001) also agrees with this statement in stating that any person who observes or assess with the intention to make a judgement that will affect qualifications of candidates needs to be trained. The International Task Force for Assessment Centre Guidelines (2010) recommends at least two days of training for assessors. It can thus be derived, according to the role of an assessor in an AC as well as the meaning of construct validity, that the level of expertise and consistency in assessor judgements are crucial to the success of the AC process (Goodstone & Lopez, 2002).

Gaugler and Thornton (1989) have been advocating that the scepticism regarding the accuracy of assessor judgments can be ascribed to the limited cognitive abilities of assessors. The reasoning behind this perspective concerns two major areas, namely that assessor evaluations could be inaccurate due to the huge amount of pressure that coincides with being an assessor, the other reason being that in a team of assessors there might be a discrepancy in the schemas being used in evaluations and eventually in the integration of their ratings. Jones and Born (2008) found that if an assessor should feel familiar or comfortable with certain behaviours, they would rate emotively, meaning that they would react positively to the assessee displaying these behaviours and rate on their emotions rather than on the behaviours

of the assessee. This phenomenon can directly be related back to illustrating poor construct validity in the example of assessors not evaluating the desired competencies, but rather their own emotions with regard to the assessee. The importance of validity in ACs are thus emphasised in this research (Jones & Born, 2008).

Literature indicates that an intervention that could prevent emotive ratings as well as other phenomena with construct validity is that of assessor training (Holmboe, 2004; Lievens, 1998). Lievens et al. (2009) make significant mention of the discussion of assessor training by stressing the importance of sufficient and adequate assessor training. This statement is supported by Schlebusch (2008), who also states that specific care should be taken with assessor training as not only the validity but also the reliability can be drastically influenced by the quality of the assessor training. Jones and Born (2008) state that assessor training can be very beneficial to the overall AC process. In a recent study concerning an international survey of assessment centre practices (2008), results indicated that out of 397 participant organisations globally, 47% indicated that there are 5 or less trained assessors in their organisation. Wills and Alexander (2000) claims that organisations should afford time for assessors to attend training. Seeing that in the American private sector ACs reached 300% return on investment at a stage (Joiner, 2004), this could be very beneficial to the organisation in the longer term.

Schleicher, Day, Mayes and Riggio (2002) remind us that it is generally known and emphasized that by paying attention to assessor training potential arises for the increase of construct validity of ratings. Schmitt, Schneider and Cohen (1990) also support the assumption that training could be a very indicative process for the strength of exercise factors. Both the South African Quality Authority (SAQA) (2001) and the International Task Force on Assessment Centre Guidelines (2010) incorporated certain standards and competencies to be met before an assessor can be registered. This should also give an indication that there is something to the arguments of training being able to influence the validity of assessment processes. Although there are numerous predictions and assumptions with reference to the training of assessors, there is, as Eurich, Krause, Cigularov and Thornton (2009) and Schleicher et al. (2002) profess, very little, if any, empirical research on various strategies of assessor training.

Training of assessors

Goodstone and Lopez (2001) report on research that indicated that 87,7% of organisations base their assessor training on common errors (such as halo, leniency and central tendency) made during assessment procedures. Although this should form part of the training Goodstone and Lopez (2001) speculate whether this approach is indeed the most effective in trying to improve construct validity and better rating accuracy in ACs. In the South African context certain criteria have been set for the training of assessors (Schlebusch, 2008). It is recommended that a trainee assessor should partake in an AC as an assessee (Schlebusch, 2008), after which the trainee assessor should attend at least two ACs as an assessor (although their input will not be considered in the final decision). Lecture room training can then be provided to the trainee assessor on how to observe, record, classify and evaluate behaviours where after the trainee assessor should act as an assistant assessor for at least two ACs under supervision of an expert assessor (Schlebusch, 2008; International Task Force on Assessment Centre Guidelines, 2010). Only once the expert assessor, other members of the assessor team and the AC administrator have all agreed that the trainee assessor is adequately trained and experienced, can the trainee assessor be classified as being a competent assessor (Schlebusch, 2008). Lastly Schlebusch (2008) and Eurich et al. (2009) agree that for assessors to be seen as competent they should be able to accurately recognise and rate manifested behaviour.

From the above-mentioned it is clear that assessor training is imperative for the AC process and many approaches can be followed for training assessors effectively and adequately. The International Task Force of Assessment Center Guidelines (2010) recommends that the training of assessors should be included in any AC process. However, Lievens et al. (2009) mention Frame-of-Reference (FOR) training that proves evidence of increasing inter-rater reliability and criterion validity as well as evidence for an increase in dimension differentiation. Jackson, Atkins, Fletcher and Stillman (2005); Lievens (2002); Lievens et al. (2009) and Schleicher et al. (2002) further advocate FOR training in stating that this training approach equips assessors with a mutual understanding, or frame of reference, with regard to pre-determined dimensions being measured in that specific AC. The initiative to train assessors using FOR was sparked by the success of FOR in the field of performance appraisal (Lievens, 2002).

Frame-of-Reference (FOR) training

Another perspective for looking at FOR training can be to define it as providing assessors with a mutual performance model to be implemented during an AC (Lievens et al., 2009). By implementing FOR training, certain principles have to be present in order to ensure the mutual understanding or frame of reference among the assessors. Firstly the dimensions (constructs/competencies) being evaluated have to be defined, behavioural examples of these dimensions have to be provided and discussed, opportunity to practice practical evaluations of behavioural constructs should be provided and lastly feedback should be given to the trainee assessors regarding their practice evaluations (Bernardin, Buckley, Tyler & Wiese, 2000; Melchers, Leinhardt, Von Aardburg & Kleinmann, 2011; Sulsky & Kline, 2007). In FOR training a deliberate schema-driven approach is implemented and exercised in order to reach a goal where assessors trade in their pre-existing prejudices for alternative schemata provided by FOR. Research indicates that by implementing FOR, the cognitive load of assessors should be relieved, which in turn should allow the assessor to rate more accurately and effectively (Lievens, 2001; Schlebusch, 2008). As mentioned previously, this could have a significant influence on the construct validity of an AC (Goodstone & Lopez, 2001; Schleicher et al., 2002).

This argument is supported by research, where various authors have reported that FOR presented higher discriminant validities, criterion validities and rating accuracy (Lievens, 2002, Schleicher et al., 2002). This is supported by Lievens (2002; 2009) and Thornton (2005) who stress the advantages and importance of FOR training in increasing assessors' effectiveness. Schleicher et al. (2002) further argue that FOR would improve the legal defensibility of an AC. Furthermore, Jackson et al. (2005) state that FOR training improves both the theoretical and practical knowledge as well as the experience of the competency being observed and evaluated in an AC among the group of assessors. The implication of these findings is that FOR training should be incorporated into assessor training. This could also link back to performance-related areas as well as organisational requirements relevant to each AC implemented (Lievens, 2001). Research on FOR training in the field of ACs is still lacking (Lievens et al., 2009).

Therefore, from the afore-mentioned, it can be derived that effective FOR assessor training could increase construct, predictive and content validity. As indicated above, international

studies have been done on assessor as well as FOR training. However, no research for the South African context has been reported. By implementing FOR training for graduate psychometrist students, it could enable them to assist in any diverse assessment procedure.

Based on the problem statement, the following research questions arise:

- How are assessment centres and assessment centre assessors conceptualised in the literature?
- What are the content and methodology related to a frame-of-reference training programme for assessors?
- What are the effects of a frame-of-reference training programme for assessors in assessment centres?

1.2 RESEARCH OBJECTIVES

Based on the research questions, the following research objectives are presented.

1.2.1 General objective

The general objective of this research is to evaluate a training programme for assessors of an assessment centre.

1.2.2 Specific objectives

The specific objectives of this research are:

- to conceptualise assessment centres and assessment centres assessors from the literature;
- to investigate the content and methodology for a frame-of-reference training programme for assessors; and
- to evaluate the effects of a frame-of-reference training programme for assessors of an assessment centre.

2 RESEACRH DESIGN

2.1 Research approach

In this study a quantitative research design will be implemented. Quantitative research can be seen as a systematic process which uses numerical data in an objective way in order to be able to explain certain relationships or explore possible new relationships between variables (Maree, 2007). With quantitative research the researcher collects numerical data with the objective of making conclusions with regard to various relationships between theory and research (Bryman & Bell, 2011).

This research fall within the field of experimental research. Experimental research can be defined as an experiment that allows for manipulation in order for the researcher to investigate and solve the “cause-and-effect” question (Maree, 2007). A classic experimental research design will be implemented by establishing two groups, namely the control group and experimental group, and incorporating a pre-test-post-test design. In practice this can be translated to the control and experimental group taking part in the same pre-test and post-test. Between the pre-test and the post-test the experimental group receive the FOR training, programme in this instance. Members of the control group only receive the training after they have taken part in the post-test.

2.2 Research method

2.2.1 Literature review

The literature review focuses on assessment centres and assessors in general. A complete review that focuses on current practices, availability and effective use of assessor training in assessment centres is done in phase 1. These sources include:

- Article databases, which include EBSCOHOST, ScienceDirect, Emerald, Sabinet Online and SAePublications.
- Relevant textbooks.
- Internet-based search engines such as Google Scholar and Google.

- Journal articles from various publications such as: *Personnel Psychology*; *International Journal of Selection and Assessment*; *Industrial and Organisational Psychology*; *Research in Personnel and Human Resources Management*; *Journal of Applied Psychology*.

2.2.2 Research participants

The population consists of postgraduate Industrial Psychology students ($N = 22$) and the stratified random sampling technique is utilised in order to divide the population into an experimental and control group. This correlates with the above-mentioned research design. Purposive sampling is normally implemented where the required population is specifically identified for being information rich and not necessarily simply random (Byram & Bell, 2011; Maree, 2007; Struwig & Stead, 2007).

2.2.3 Measuring instruments

The measuring methodology utilised in this research concerns an observation and rating of a typical AC simulation. The entire population group will be requested to observe pre-recorded video material of three typical AC simulations. Three candidates (role-players) were subjected to the simulations, which were then recorded for pre-test-post-test purposes. The population will subsequently be requested to award a rating to each candidate separately on nine competencies. Once the FOR training is completed, the participants will observe the same AC simulation and again award a rating as part of the post-test. The rating received from the population group is analysed after the pre-test and post-test to investigate the effect of the FOR training programme. The comparison between the experimental and control group ratings for the pre-test as well as the comparison between these groups for the post-test will then indicate the effect of the training programme.

2.2.4 Research procedure

The first action in the research procedure is to obtain approval from the NWU Ethics Committee. When approval is obtained, both the experimental and control groups are invited to an information session. In this information session the participants are informed about the research aim, objectives as well as the training programme and procedure to be followed. After the information session, the participants are afforded the opportunity of deciding

whether they wish to participate in the research. If they choose to do so, their informed consent is obtained where after the confirmed participants are randomly divided into an experimental and a control group, as is stated in the pre-test-post-test design model. The next step is for all the participants to form part of the pre-test observation as well as the pre-test focus group where the experimental group and control group form part of two separate focus groups. The pre-test observation requires the participants to observe pre-recorded video-material of three typical AC simulations and assess and rate three candidates for each simulation on pre-determined competencies. The exact same procedure is followed for the post-test with the same video-material that needs to be observed, in order to ensure standardisation of the data collected.

Between the pre-test and the post-test, the experimental group will receive the FOR training programme. The training programme is a three (3) day programme consisting of workshops correlating with the FOR principles. After the training programme is administered to the experimental group the post-test is administered. Only once the post-test data has been collected, will the control group undergo the FOR training programme.

2.2.5 Statistical analysis

SPSS (SPSS Inc., 2009) is carried out to analyse the data and statistics obtained. Means, standard deviations, skewness and kurtosis, generally known as descriptive statistics, are used to analyse the data. The Wilcoxon Signed Rank Test as well as the Mann-Whitney *U*-test will also be utilised to be able to determine the effect of the FOR training programme (Palant, 2010). Cronbach's alphas are also utilised to observe the internal consistency and reliability of the AC utilised in this research. These statistical methods indicate the effect of the training programme on the rating differences between the control group and the experimental group and the accuracy thereof.

2.2.6 Ethical considerations

Any research carried out has to be conducted in a morally ethic manner and to follow a code of moral guidelines (Struwig & Stead, 2007). Any researcher conducting any form of research has to be familiar with this concept. In this particular research the principles of confidentiality and anonymity, freedom of participation as well as honest collecting and reporting of data is implemented and followed.

Ethical codes and guidelines are designed to protect the participants. This is evident from the American Psychological Association's (2007) five guidelines provided in their code of conduct. They state that a) only a qualified and competent researcher should be allowed to conduct research, b) when conducting any form of research, honesty, integrity, fairness and respect are principles that should be present at all times, c) the researcher should be held responsible for any actions taken during the research process, d) the participants' privacy, cultural preferences, racial heritage, gender and rights should be taken into account to ensure no discrimination, e) the research should not harm any participants and should be conducted in their best interest at all times.

3 CHAPTER OVERVIEW

The differences with regard to the control and experimental groups' scores are examined in Chapter 2. Chapter 3 discusses the conclusions, limitations and recommendations of this study.

4 CHAPTER SUMMARY

In Chapter 1 the problem statement, research objectives, measuring instruments as well as the research method were discussed, after which a brief overview of the chapters that will follow is explained.

REFERENCE LIST

- American Psychological Association. (2010). *Ethical Principles of Psychologists and Code of Conduct 2010 Amendments*. Retrieved Feb 4, 2012 from <http://www.apa.org/ethics/code/index.aspx>
- An International Survey of Assessment Centre Practices. (2010). *The Global Research Questionnaire*. Surrey: Assessment & Development Consultants Ltd.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.
- Bernarding, H. J., Buckley, M. R., Tyler, C. L., & Wiese, D. S. (2000). A reconsideration of strategies for rater training. *Research in personnel and human resources management, 18*, 221-274.
- Bryman, A., & Bell, E. (2011). *Business research methods*. Oxford, UK: Oxford University Press.
- Dilchert, S., & Ones, D.S. (2009). Assessment centre dimensions: Individual differences correlates and meta-analytical incremental validity. *International Journal of Selection and Assessment, 17*(3), 254-270.
- Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. III, (2009). Assessment Centers: Current practice in the United States. *Journal of Business & Psychology, 24*, 387-407.
- Gaugler, B. B., & Thornton, G. C. (1989). Number of assessment centre dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology, 74*, 611-618.
- Goodstone, M. S., & Lopez, F. E. (2001). The Frame of Reference Approach as a solution to an Assessment Center dilemma. *Consulting Psychology Journal: Practice and Research, 53*(2), 96-107.
- Guion, R. M. (1998) *Assessment, measurement, and prediction for personnel decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Health Professions Council of South Africa. (2010). Health Professions Act 1974 (Act 56 of 1974). Retrieved 23 Jan, 2012, from <http://www.hpcsa.co.za>.
- Holmboe, E. (2004). Faculty and the observation trainees' clinical skills: problems and opportunities. *Academic Medicine, 79*(1), 16-22.

- Howard, A. (2009, 18 March). *The changing face of assessment centers: what we learned and what we missed*. The 29th annual Assessment Centre Study Group of South Africa conference, Stellenbosch, Western Cape Province, South Africa.
- International Task Force on Assessment Center Guidelines. (2000). *Guidelines and ethical considerations for assessment center operations*. San Francisco, CA: Development Dimensions International
- Jackson, D. J. R., Atkins, S. G., Fletcher, R. B., & Stillman, J. A. (2005). Frame of reference training for assessment centers: effects on interrater reliability when rating behaviours and ability traits. *Public Personnel Management, 34*(1), 17-30.
- Joiner, D. A. (2004, Jun 20-23). Assessment Centre trends: assessment centre issues and resulting trends. Paper presented at the 28th annual meeting of IPMAAC, Seattle, Washington, United States of America.
- Jones, R. G., & Born, M. P. (2008). Assessor Constructs in Use as the Missing Component in Validation of Assessment Center Dimensions: A critique and directions for research. *International Journal of Selection and Assessment, 16*(3), 229-238.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection and Assessment, 6*, 141-152.
- Lievens, F. (2002). An examination of the accuracy of slogans related to assessment centres. *Personnel Review, 31*, 86-102.
- Lievens, F. (2009). Assessment Centers: A tale about dimensions, exercises and dancing bears. *European Journal of Work and Organisational Psychology, 18*(1), 102-121.
- Lievens, F., & Thornton, G. C. III. (2005). Assessment centres: Recent developments in practice and research. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell Handbook of Personnel Selection* (pp. 243-264). Malden, MA: Blackwell.
- Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment Centers at the crossroads: toward a reconceptualization of assessment center exercises. *Research in Personnel and Human Resources Management, 28*, 99-152.
- Maree, K. (2007). *First steps in research*. Pretoria: Van Schaik Publishers.
- Meiring, D. (2008). Assessment centres in South Africa. In S. Schlebusch, & G. Roodt (Eds.), *Assessment Centres: Unlocking potential for growth* (pp. 21-31). Randburg, South Africa: Knowres Publishing (Pty) Ltd.
- Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology, 64*, 53-87.

- Pallant, J. (2010). *SPSS Survival Manual. A step by step guide to data analysis using SPSS* (4th ed). Berkshire, England: McGraw-Hill
- Pell, G., Homer, M. S., & Roberts, T. E. (2008). Assessor training: its effect on criteria-based assessment in a medical context. *International Journal of Research & Method in Education*, 31(2), 143-154
- Salkind, N. J. (2009). *Exploring research* (pp. 243-251). Saddle River, NJ: Pearson Education Inc.
- Schlebusch, S. (2008). Before the Centre. In S. Schlebusch, & G. Roodt (Eds.), *Assessment Centers: Unlocking potential for growth* (pp. 176-196). Randburg, South Africa: Knowres Publishing (Pty) Ltd.
- 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.
- Schmitt, N., Schneider, J. R., & Cohen, S. A. (1990). Factors affecting validity of a regionally administered assessment center. *Personnel Psychology*, 43, 1-12.
- South African Qualifications Authority (SAQA). (2001). Criteria and guidelines for assessment of NQF registered unit standards and qualifications. Retrieved 4 Feb, 2012 from <http://www.sqa.org.za>
- SPSS Inc. (2008). *SPSS 16.0 for Windows*. Chicago, IL.
- Struwig, F. W., & Stead, G. B. (2001). *Planning, designing and reporting research*. Cape Town, South Africa: Pearson Education South Africa.
- Sulsky, L. M., & Kline, T. J. B. (2007). Understanding frame-of-reference training success: a social learning theory perspective. *International Journal of Training and Development*, 11(2), 121-131.
- Thornton, G. C., III, Murphy, K. R., Everest, T. M., & Hoffman, C. C. (2000). Higher cost, lower validity and higher utility: Comparing the utilities of two tests that differ in validity, costs, and selectivity. *International Journal of Selection and Assessment*, 8, 61-75.
- Thornton, G. C., III & Mueller-Hanson, R. (2004). *Developing organisational simulations: A guide for practitioners and students*, Mahwah, New Jersey: Lawrence Erlbaum Associates.

CHAPTER 2

RESEARCH ARTICLE

THE EVALUATION OF A FRAME-OF-REFERENCE TRAINING PROGRAMME FOR ASSESSORS OF ASSESSMENT CENTRES

ABSTRACT

Orientation;

The use of assessment centres (ACs) has drastically increased over the past decade. However, ACs are constantly confronted with the lack of construct validity. One aspect of ACs that could improve the construct validity significantly is that of assessor training. Unfortunately untrained or poorly trained assessors are often used in AC processes. Literature indicates that a specific technique that can be used to train assessors is that of Frame-of-Reference (FOR) training.

Research purpose:

The purpose of this research was to evaluate a frame-of-reference training programme for assessors of an assessment centre.

Research design, approach and method:

A quantitative research design was implemented, utilising a randomised pre-test-post-test comparison group design. The population group consisted of Industrial Psychology postgraduate students at a South African university. The entire population consisted of 22 postgraduate students, 11 formed the experimental group. The remaining 11 students formed the control group. Three typical AC simulations were utilised as the pre- and post-test, with the ratings gathered from both groups in the pre- and post-test were statistically analysed to determine the effect of the FOR training programme.

Main findings:

The data indicated that there was a significant increase in the familiarity of the participants with the one-on-one simulation and with the group discussion simulation.

Practical implications:

This indicates that if implemented correctly, a FOR training programme for assessors of ACs could have a significant effect.

Key words: assessment centres, assessors, training of assessors, frame-of-reference training.

INTRODUCTION

The popular use of Assessment Centres (ACs) has over the years drastically increased at international level in various applied industries (International Task Force on Assessment Centre Guidelines, 2010; Krause & Gebert, 2003). Currently, this assessment practise is implemented in, amongst others, educational, military, industrial and government organisations. It is widely accepted that ACs are mostly used in the field of personnel psychology for processes such as recruitment, selection and identification of managerial potential and talent (Dilchert & Ones, 2009; Lievens & Thornton, 2005). Lievens and Thornton (2005) emphasise the efficacy and importance of the implementation of ACs in personnel selection and promotion. Although for a long time ACs were solely used at international level, in 1974 this technique started establishing itself in South Africa as a popular assessment technique (Meiring, 2008). Major companies incorporated ACs as a means of assessment, which led to a need for practitioners to exchange ideas in a constructive manner, and henceforth the Assessment Centre Study Group (ACSG) was founded (Meiring, 2008). Since 1970, the main aim of the ACSG is to hold annual conferences to promote new research, insights and teaching of ACs in a constructive and effective manner.

Thornton and Rupp (2006) explain that an AC can be seen as a combination of work-like exercises as well as other assessment type procedures specifically designed to activate certain behaviour in candidates in order for those behaviours and skills to be evaluated and observed. Schlebusch (2008) claims that the main aim and purpose of an AC is to select the most appropriate participant to be appointed in a position or programme and also states that one of the criteria for an AC is that participants should be informed that results will influence the decision of appointment. Some specific features that should also be present for a true AC are: a job analysis should be carried out; multiple simulations and assessment instruments should be utilised; multiple and competent observers and role-players should be present; behavioural and not psychological constructs should be observed; behaviour should be noted and classified; data integration should take place and efficient feedback should be provided to participants (Schlebusch, 2008).

Although ACs are one of the more costly techniques used for assessment, Eurich, Krause, Cigularov and Thornton (2009) argue that ACs have good predictive validity (Thornton, Murphy, Everest & Hoffman, 2000) and criterion-related validity (Arthur, Day, McNelly & Edens, 2003). Furthermore, depending on the expertise level of the assessors, ACs also

indicate evidence of good inter-rater reliability (Lievens, 2002). Moreover, Joiner (2004) states that in the American private sector ACs, at some point, reached a 300% return on investment (ROI).

Thornton and Mueller-Hanson (2004) state that although ACs consistently demonstrate criterion validity, the construct validity is still lacking significantly. Collins et al. (2003) mention in their study that evidence against construct validity, such as constant low construct validity in certain dimensions, has in fact been reported. The issue of construct validity can be seen as one of the biggest challenges that ACs have to conquer (Guion, 1998). In a study done by Lievens (2009) he also mentions the significant issue of construct validity and that he feels ACs have to overcome the “lack of evidence to measure the constructs (dimensions) they are reported to measure” (p.104). It can thus safely be said that over the years the biggest unresolved problem that still remains in the practice of ACs is that of construct validity.

The consistency of assessor judgments is one specific aspect of ACs that influence or contributes to the construct validity (Pell, Homer & Roberts, 2008). The main aim of an assessor in an AC is to observe a candidate’s behaviour and assign a rate accordingly which results in the candidate being appointed in a specific post (Goodstone & Lopez, 2001). Therefore the assessor’s expertise plays a significant role in the construct validity of the process (Jones & Born, 2008).

Assessors in Assessment Centres

The general aim of ACs is the evaluation of various competencies, and for this reason a team of assessors is needed to assess and observe these competencies (Schlebusch, 2008). According to the International Task Force on Assessment Centre Guidelines (2010), the definition of an assessee is that of “an individual whose competencies are measured by an assessment centre” (p.10). This corresponds with previous research (Lievens, Tett & Schleicher, 2009; Schleicher, Day, Mayes & Riggio, 2002). Goodstone and Lopes (2001) confirm this by stating that an assessor’s task is ultimately that of performance appraisal; thus the essential part of any AC process is that of a trained assessor observing a candidate’s behaviour and accordingly appoints a rating to it.

The importance of validity in ACs is clear from findings from Jones and Born (2008) who found that assessors react more positively to behaviours and situations they are familiar with and therefore give emotive ratings. Schlebusch (2008) argues that up until now South African research has been reactive rather than proactive and that research on ACs, and specifically assessor training for the South African context, is limited. It is clear that although many issues contribute to the construct validity debate, one crucial element is that of assessors and their training.

Lievens (2009) asserts that trained observers should be used to observe participants in a typical job-related setting, whilst paying attention to various determined dimensions. Observing and evaluating participants are thus carried out by observers or individuals otherwise known as assessors. Schlebusch (2008) defines the group of assessors as the individuals who “have the greatest impact on the whole assessment process”. Literature indicates that two of the most common mistakes made in any AC is that of firstly, using unqualified assessors and secondly, using poorly trained assessors (Thornton & Meuller-Hanson, 2004). Both Holmboe (2004) and Lievens (1998) found that the training of assessors could possibly have a significant effect on the construct validity of ACs. That the focus of assessor training should rather be on the quality of the training than the quantity (length) has been supported by research (Jackson, Atkins, Fetcher & Stillman, 2005). Schlebusch (2008) supports this statement in stating that not only the validity but also the reliability of an AC can be influenced by the quality of assessor training, and therefore specific care should be taken to ensure that they are indeed competent.

Training of Assessors

The main aim of training observers is to develop certain abilities that enable them to accurately and effectively rate participants' behaviour (Schlebusch, 2008). Lievens et al. (2009) stress the fact that sufficient training for assessors is critical. For these assessors to be able to rate accurately, Schlebusch (2008) says, some of the skills relevant to observing, noting, classifying and evaluating participants' behaviour during exercises or simulations have to be developed. They should thus be able to record detailed behaviour and reactions accordingly and precisely. Schlebusch (2008) indicates steps that should ideally be followed for an individual who wishes to be classified as a competent assessor. Jones and Born (2008) claim that the levels of assessor expertise significantly affect the validity of ACs and can be very beneficial to the AC process.

Schlebusch (2008) recommends that for the South African context, an assessor in training should attend an AC as a participant. After completion, the individual should then attend an AC as an assessor (although their inputs will not be considered at that time). When individuals have attended two ACs, the International Task Force on Assessment Centre Guidelines (2010) advises they undergo lecture room training where after they should act as assistant assessors twice under the supervision of a qualified competent assessor (Schlebusch, 2008). Only once the expert assessor, AC administrator and other members of the assessor team ultimately agree, can the individual be declared a competent assessor (International Task Force on Assessment Centre Guidelines, 2010; Schlebusch, 2008).

Lievens et al. (2009), however, claim that evidence exists for another technique, namely frame-of-reference training, which could increase inter-reliability, dimension differentiation and even criterion-validity. Jackson et al. (2005) suggest that frame-of-reference (FOR) training should be implemented in the training of assessors to ensure a shared understanding of dimensions being measured.

Frame-of-reference training

Frame-of-reference (FOR) training specifically focuses on developing a mutual understanding or frame of reference amongst assessors (Lievens, 2002; Lievens et al., 2009; Schleicher et al., 2002). The purpose of developing this mutual understanding is to equip all assessors with the same performance model that they can utilise as a tool while observing during an AC (Lievens et al., 2009). This mutual understanding can be reached by defining the dimensions (constructs/competencies) being evaluated, providing and describing appropriate behavioural examples of the dimensions (constructs/competencies) being evaluated, providing opportunities for practising evaluations practically, and finally providing feedback to assessors relating to their evaluations (Bernarding, Buckley, Tyler & Wiese, 2000; Melchers, Lienhardt, Von Aardburg & Kleinmann, 2011; Sulsky & Kline, 2007). The ultimate goal of FOR training is thus to assist assessors in their tasks of observing and evaluating behaviours and to then categorise these observations into accurate and appropriate performance dimensions.

Lievens (2002; 2009) and Thornton and Rupp (2006) have on numerous occasions emphasized the importance and advantage of FOR training in increasing the effectiveness of assessors. Jackson et al. (2005) state that an explanation for this could be the fact that FOR training promotes an improved theoretical as well as practical understanding of relevant behaviour amongst assessors. This understanding can be linked to certain areas related to performance and organisational requirements that each AC demands. FOR training should therefore be specifically designed for a certain AC (Lievens, 2002). As an example, in an AC where listening skills would be observed, in the training the specific listening skills required for the AC will be defined and discussed in detail. After which a practical example of the listening skills required will be illustrated or discussed. Certain skills that could appear as listening skills but are not necessarily required for this AC will also be discussed. The aim of this process is to equip the assessors with a mental picture of the competency they will observe during the AC and to eliminate the possibility of assessors using their own mental pictures of how a certain competency manifests. Lievens et al. (2009) further indicate that research on comprehensive training approaches such as FOR training is lacking.

Schleicher et al. (2002) believe that the implementing of FOR training for assessors can be viewed as an intervention that will have a significant influence on both the construct validity and the criterion-validity of ACs. FOR training is recognised as a well-known term in the field of performance appraisal, mostly because of the evidence that FOR has a significant effect on the increase of assessors' reliability and accuracy (Lievens, 2009; Schleicher et al., 2002). Lievens and Thornton (2005) point out that FOR training not only trains assessors to distinguish between behaviours and dimensions in accordance with a specific framework, it also aims to reduce the cognitive load by implementing a unified scoring framework.

Lievens (2002; 2009) and Schleicher et al. (2002) claim that if the FOR training approach is followed it should lead to more accurate results by educating assessors to use more effective and appropriate schemas (frames of reference). This argument is supported by research where various authors have reported that FOR training presented higher discriminative validities, criterion validities and rating accuracy (Lievens, 2002 & Schleicher et al., 2002). The implication of these findings indicate on a practical level that the principles of FOR training should be incorporated into assessor training seeing that there is evidence that FOR trained assessors possess better abilities in using different dimensions accurately (Lievens, 2002).

Schleicher et al. (2002) also argue that FOR increases overall validity as well as legal defensibility and therefore this approach should be implemented and followed.

After completing an Honours degree in Industrial Psychology a student can register as a psychometrist with the Health Professions Council of South Africa (HPCSA). The HPCSA states that a registered psychometrist should be able to participate in assessment procedures in diverse settings and organisations. The scope of practice for assessments (HPCSA, 2010) mentions that during any assessment, observers have to declare their limits to their evaluations and not misuse the assessment technique or results. By training graduate psychometrist students in FOR training, their ability to participate in diverse assessments and settings could be enhanced.

From the discussion above it is clear that by focussing on effective assessor training, more specifically FOR training, construct validity as well as predictive and content validity could increase. It has however been speculated that FOR could also influence convergent validity. However, currently there is no conclusive evidence yet proving this. Although international studies exist on assessor training as well as FOR training, currently no such research exists for the South African context.

Research objectives

Based on the discussion above, the objectives (general and specific) of this research were:

General objective

The general objective of this research is to evaluate a training programme for assessors of an assessment centre.

Specific objectives

- to conceptualise assessment centres and assessment centres assessors from the literature.
- to investigate the content and methodology for a frame-of-reference training programme for assessors; and
- to evaluate the effects of a frame-of-reference training programme for assessors of an assessment centre.

Expected contribution of the study

The expected contribution of this study is to design an assessor training programme using Frame-of-Reference training for the South African context, specifically by incorporating this programme into the selection process for forthcoming Honours students at universities. The principle of FOR training has continuously proven that it not only improves assessor accuracy but also criterion and discriminative validities of ACs (Lievens & Conway, 2001 & Schleicher et al., 2002). Various authors further claim that assessor judgements are one of the reasons for low construct validity (Jones & Born, 2008; Pell et al., 2008). Jackson et al. (2005) state that presenting assessors with FOR training produced more accurate ratings than did rater and error training for assessors. Currently there is only one South African source providing a framework for assessor training (Schlebusch, 2008). As previously mentioned, construct validity of ACs is seen as one of this process' strongest critique (Buckett, 2010; Collins et al., 2003; Guion, 1998; Lievens, 2009). The aim of this study, however, will be to evaluate the effect a FOR training programme has on assessors' ratings. This will be determined by using focus groups and interviews. The effect FOR training has on an AC as a process will be determined in a potential future PhD study. Seeing that FOR training can improve construct validity of ACs and currently no such programme exists for the South African context, the contribution of this study will focus on providing such a programme, in particular to selection processes for Honours students in South Africa.

RESEARCH DESIGN

Research approach

A quantitative design was implemented for this research. According to Maree (2007) "quantitative research is a process that is systematic and objective in its ways of using numerical data from only a selected subgroup of a universe (or population) to generalise the findings to the universe that is being studied" (p.145). According to Bryman and Bell (2011) with a quantitative research design the researcher collects numerical data in order to be able to draw conclusions with regard to relationships between theory and research.

This research also fell within the field of experimental research. Maree (2007) describes experimental research as an experiment that can be manipulated as well as controlled in order for the researcher to be able to answer a "cause-and-effect" question. A classic experimental research design was implemented where two groups were established. The dividing of the

group into two provides the basis for manipulation of the independent variable (Bryman & Bell, 2011; Struwig & Stead, 2007). Salkind (2009) further states that classic experimental allows the researcher to extensively explore the effect of the independent variable (FOR training programme) on the dependent variable (participants' knowledge of the subject). In this study a randomised pre-test-post-test control group design was implemented (Salkind, 2009). This particular research utilised two groups of participants namely the comparison group and the experimental group. Both the comparison and the experimental group received a pre-test and a similar post-test in the form of observing typical AC simulations. The independent variable, namely the frame-of-reference training programme, was administered to the experimental group between the pre- and post-test, but the comparison group did not receive the FOR training programme. The comparison group only received the training programme after the post-test. This provided the opportunity for the comparison group to receive the training and it ensured fair research practices.

Research method

This section presents the research participants, measuring battery, research procedure, statistical analysis and ethical considerations. It should also be mentioned that research question 1 is answered in the literature review

Research participants

The population consisted of postgraduate students at a tertiary institution. Purposive sampling was used to obtain a population of 22 students ($N=22$) included in the study. The sample size was governed by data saturation and was determined by the number of participants willing and accessible to participate (Burns & Grove, 1987). The method of purposive sampling is used in incidents where the sampling is not necessarily focused on being random but rather done with a specific outcome in mind and the goal of providing a sample of information-rich participants (Bryman & Bell, 2011; Maree, 2007; Struwig & Stead, 2007).

Table 1

Characteristics of the Participants (N=22)

Item	Category	Frequency	Percentage
Gender	Male	7	32%
	Female	15	68%
Ethnicity	Caucasian	20	91%
	Indian	1	4.5%
	African	1	4.5%
Age	20-22 years	11	50%
	23-25 years	11	50%
Language	Afrikaans	20	91%
	English	1	4.5%
	Xhosa	1	4.5%
Qualification level	Undergraduate students	22	100%

From the table above can be derived that the population used in this research was predominantly Caucasian and Afrikaans speaking. More than half of the population was female and all participants were undergraduate students (currently completing their Honours degree) between 20 and 25 years of age

Measuring instruments

Data was collected by means of ratings of nine competencies of a typical AC simulation. During this process the participants were requested to evaluate independent role-players based on nine competencies whilst viewing a DVD recording of a typical AC simulation. During their evaluation they were asked to award a rating to the role-player on the various competencies. The ratings received from the experimental and the control group were compared and analysed after the pre- and post-test. The effect of the FOR training programme on the practical understanding and skills of the participants to observe behaviour accurately was determined by comparing the results of the pre- and post-test respectively.

Research procedure

In order to statistically and ethically gather data the research project obtained approval from the NWU Ethics Committee. Once approval was granted, all participants were invited to an information session during which the researches' aim and the procedure were explained to them. The participants' consent was first obtained, where after they were randomly divided into the control group and the experimental group. This is in accordance with the pre-test-post-test control group design (De Vos, Strydom, Fouché & Delport, 2005). The schedule for the pre-test simulations was then drawn up.

The entire group was subjected to a pre-test assessment. During this pre-test the participants had to evaluate role-players in an AC based on nine predetermined competencies. Next, the experimental group was subjected to the FOR training programme, whilst the comparison group received no training. The training programme mainly consisted of a series of workshops dedicated to the development of interviewing and assessor skills. The contents of the FOR training programme is illustrated in the table 2. The programme was presented by means of two prior-recorded ACs during which the participants were taught FOR principles. Once the training programme had been presented, the entire group underwent the post-test. This assessment consisted of the same DVD recording of the AC as with the pre-test. The comparison group only underwent the training programme after the post-test had been administered. The ratings of the experimental and control group were then compared after the post-test to measure the extent of the FOR training programme. During the pre- and post-test, the same video material was viewed by participants. This ensured standardisation of the collected data.

Table 2 depicts the content and methodology of the FOR training programme:

Table 2

The content, objectives and methodology of the FOR training programme for assessors of assessment centres

Workshop	Title	Objective	Method
Day 1 Session 1	Basic Interviewing and Facilitation skills	Transferring practical and theoretical knowledge of managing a basic facilitation process	<ul style="list-style-type: none">• Lectures• Role play
Day 1 Session 2	Introduction to ACs and competencies	Manifest a comprehension of basic AC principles and practices	<ul style="list-style-type: none">• Group work• Discussions

		Manifest an understanding of competencies and how to identify them	<ul style="list-style-type: none"> • Lectures
Day 2 Session 1	Practical work	To observe competencies in role-players' behaviour and evaluate accordingly	<ul style="list-style-type: none"> • Video material • Group discussions
Day 2 Session 2	Feedback	Provide feedback on evaluations by expert assessors	<ul style="list-style-type: none"> • Video material • Group discussions • Individual coaching session
Day 3 Session 1	Conclusive	Transferring of knowledge	<ul style="list-style-type: none"> • Lecturing • Group discussion

Statistical analysis

In this study, SPSS (2009) was utilised to determine non-parametric statistics, namely the Mann-Whitney *U*-test and the Wilcoxon Signed Ranks test. First the Mann Whitney *U*-test was implemented with the experimental and control groups by comparing the medians, to determine whether the two groups were at the same level prior to the FOR training programme being implemented. This non-parametric technique is preferred for data measured according to a category or a ranking, as well as for small samples (Pallant, 2010), which is the case in this study. Next, the Wilcoxon Signed Ranks test was used to determine the difference in the experimental group between the pre- and post-test. This technique is used with repeated measures. In other words, to measure the participants during the two different occasions (Pallant, 2010). Effect sizes were calculated for the results of both the Mann Whitney *U*-test and the Wilcoxon Signed Ranks test. This was done by dividing the *z*-value by the square root of *N* (=22). The guidelines, as set by Cohen (1988), were used to determine the effect size, namely .1=small effect, .3=medium effect and .5=large effect.

Cronbach's alpha coefficients were also used to determine the internal consistency and reliability of the ratings received. These statistics were utilised to effectively observe the effect of the training programme on the rating difference and accuracy between the experimental group and control group.

Ethical considerations

In order to conduct this research, the researcher must first possess a thorough knowledge of applicable ethics and receive proper ethical authorisation and permission for the NWU Ethics

Committee. Struwig and Stead (2007) claim that research ethics are put in place to ensure that all research is conducted in a morally ethical manner by following a code of moral guidelines. It is thus important for all activities conducted in the research to be able to conform to the principle of “do no harm”. The American Psychological Association (2007) provides five guidelines in their code of conduct. These guidelines stipulate that a) a researcher qualified and competent to carry out their specific research, b) honesty, integrity, respect and fairness should be present in all research activities, c) the researcher should be willing to take responsibility for his/her actions, d) at all times during the research the participants’ rights, privacy, cultural preferences, gender and racial heritage should be respected in order to ensure that no discrimination takes place during the research, e) the research project should never harm any participants and be in their best interest at all times.

In this research the ethical considerations that were considered are privacy, confidentiality and anonymity, termination of participants’ involvement, and the provision of research attention. All the necessary provisions were made to ensure that all participants had the right to privacy to ensure the protection of their integrity and dignity. Regarding confidentiality and anonymity, all participants were ensured of total anonymity; meaning that no names will be mentioned and in the case of confidentiality being breached, incriminating record will be destroyed. During the data collection period, all participants had the right to terminate their involvement, even if they initially agreed to form part of the study. Research assistance was also provided in the form of feedback being given to participants by relevant researchers.

RESULTS

The following section gives an account of the results of the study. First, the content and methodology of the FOR training programme will be reported, and then the Cronbach’s alpha will be investigated. Finally, the non-parametric statistics will be reported.

The training programme was conducted over a three- (3) day period. An existing training programme was adjusted accordingly to accommodate frame-of-reference training. From the table it can be seen that the first workshop focused on basic facilitation skills such as listening, objective attending and paraphrasing. The second workshop focused on informing the participants of the principles of an AC as well as in-depth discussions of the competencies being assessed. During the second day of the training programme, practical exercises were

conducted on rating the 4 candidates (role-players) taking part in two separate one-on-one simulations on the competencies (strategic perspective, interpersonal skills, leadership, conflict management, judgement, self-confidence, assertiveness, persuasive communication and performance under pressure). On the third day of the training programme, two workshops were presented; first focussing on practical exercises on rating the same candidates participating in two different presentation simulations on the discussed dimensions. The final workshop focused on practical exercises on rating 2 group discussion simulations consisting of 4 candidates each.

This study consisted of two groups; an experimental group and a control group. In order to answer the third objective of this study, the first step was to determine the internal consistency of the AC (One-on-One, Presentation and Group Discussion simulations) between the Experimental group and Control group, as reported in Table 3.

Table 3

The Cronbach's alphas (α) between the pre- and the post-test for the experimental and control group for the AC

Simulation	Pre-test			Post-test		
Cronbach's alpha (α): Experimental group						
	Candidate 1	Candidate 2	Candidate 3	Candidate 1	Candidate 2	Candidate 3
One-on-One	0.892	0.733	0.725	0.912	0.813	0.936
Presentation	0.935	0.876	0.873	0.911	0.759	0.574
Group Discussion	0.941	0.908	0.876	0.928	0.800	0.855
Cronbach's alpha (α): Control group						
	Candidate 1	Candidate 2	Candidate 3	Candidate 1	Candidate 2	Candidate 3
One-on-One	0.883	0.958	0.871	0.853	0.807	0.737
Presentation	0.880	0.948	0.857	0.810	0.927	0.808
Group Discussion	0.926	0.857	0.545	0.932	0.900	0.830

In the table above, the internal consistency, by reporting the Cronbach's alphas, for the experimental and control groups for the pre- and post-test, is illustrated. From the table it can be derived that the internal consistency for the experimental group from the pre-test (across all three simulations) ranges between 0.725 and 0.941, and for the post-test, between 0.574 and 0.936. Similarly, for the control group the Cronbach's alphas for the pre-test (across all three competencies) range from 0.545 to 0.958 and for the post-test between 0.737 and 0.932.

Next, the significant differences in terms of the rating of the nine competencies of the AC were determined between the experimental and control groups prior to the FOR training

programme. A Mann-Whitney *U*-test revealed no significant differences in the assessment of the one-on-one, presentation and group discussion simulations, between the experimental and control groups. For the three simulations utilised in the AC, the Mann-Whitney *U*-test ranged between 31 and 60, the *z*-value ranged between -1.94 and -0.03, the *p*-value ranged between .052 and .974, and lastly the correlations coefficient ranged between -.41 and -.01.

The next step was to investigate the difference between the pre- and post-test scores for the experimental group for the nine competencies of the AC (one-on-one, presentation and group discussion simulations). These results are reported in Table 4.

Table 4

The difference between the pre- and post-test scores for the experimental group for the AC

Simulation	Pre-test – Post-test	<i>z</i> -value	<i>p</i>	<i>R</i>
One-on-One	Candidate 1	-1.02 ^b	.306	-.22
	Candidate 2	-1.16 ^c	.247	-.25
	Candidate 3	-2.81 ^b	.005*	-.60
Presentation	Candidate 1	-1.70 ^b	.090	-.36
	Candidate 2	-0.87 ^b	.385	-.19
	Candidate 3	-1.53 ^b	.126	-.33
Group Discussion	Candidate 1	-2.36 ^b	.018*	-.50
	Candidate 2	-1.65 ^b	.099	-.35
	Candidate 3	-2.80 ^b	.005*	-.60

^a Wilcoxon Signed Ranks Test

^b Based on positive ranks

^c Based on negative ranks

* Statistically significant: $p \leq .05$

Practically significant correlation: $r \geq .10$ (small effect); $r \geq .30$ (medium effect); $r \geq .50$ (large effect)

The Wilcoxon Signed Ranks Test indicated a statistically significant reduction in assessing the One-on-One simulation for Candidate 3 after the FOR training programme, $z = -2.81$, $p = .306$, with a large effect size ($r = -.60$) (See figure 1). The median score on the aforementioned decreased from the pre-test ($Md = 7.56$) to the post-test ($Md = 6.78$). Similarly, a significant reduction was found for Candidate 1 ($z = -2.36$, $p = .018$, $r = -.50$: large effect)

and Candidate 3 ($z = -2.80, p = .005, r = -.60$: large effect) in the assessment of the Group Discussion simulation. The median score for Candidate 1 decreased from the pre-test ($Md = 6.67$) to the post-test ($Md = 5.22$), while for Candidate 3, the median score on the aforementioned decreased from the pre-test ($Md = 6.89$) to the post-test ($Md = 6.22$) (See figure 2).

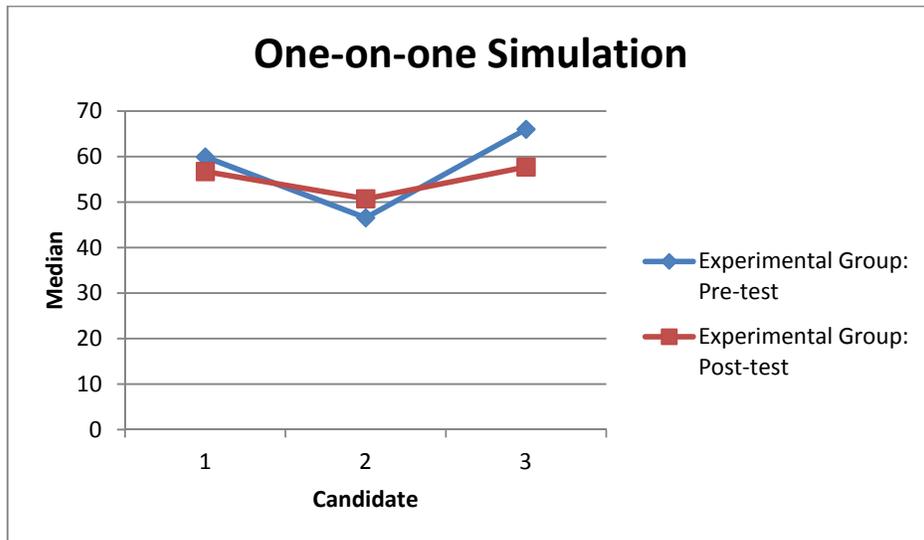


Figure 1

The comparison of the pre- and post-test rating for the one-on-one simulation by the experimental group

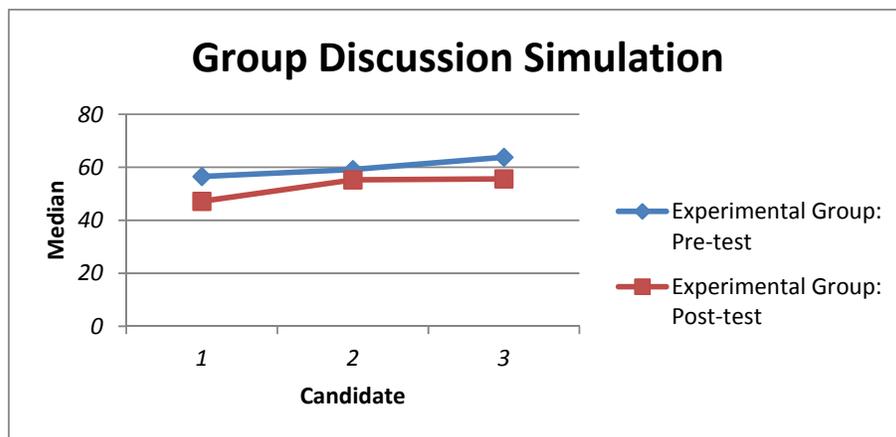


Figure 2

The comparison between the pre- and post-test ratings for the group discussion simulation for the experimental group

Next, the difference between the pre- and post-test scores for the control group for the nine competencies of the AC (one-on-one, presentation and group discussion simulations) is reported in Table 5.

Table 5

The difference between the pre- and post-test scores for the control group for the AC

Simulation	Pre-test – Post-test	z-value	P	r
One-on-One	Candidate 1	-1.29 ^b	.197	-.28
	Candidate 2	-0.31 ^b	.755	-.07
	Candidate 3	-0.98 ^b	.327	-.21
Presentation	Candidate 1	-1.88 ^b	.060	-.40
	Candidate 2	-0.15 ^b	.878	-.03
	Candidate 3	-1.03 ^b	.305	-.22
Group Discussion	Candidate 1	-0.18 ^b	.859	-.04
	Candidate 2	-0.31 ^c	.759	-.07
	Candidate 3	-1.65 ^b	.100	-.38

^a Wilcoxon Signed Ranks Test

^b Based on positive ranks

^c Based on negative ranks

^{*} Statistically significant: $p \leq .05$

Practically significant correlation: $r \geq .10$ (small effect); $r \geq .30$ (medium effect); $r \geq .50$ (large effect)

Table 4 reveals that, unlike the experimental group, the Wilcoxon Signed Rank Test for the control group shows no statistically significant differences between the pre- and post-test for the AC (One-on-One, presentation, group discussion simulations).

The following step involved the determination of the differences in the rating of the nine competencies of the AC between the experimental and control groups after the FOR training programme had been implemented within the experimental group. A Mann-Whitney *U*-test revealed significant difference in the assessment of the One-on-one simulation for Candidate 2 between the experimental group and the control group ($U = 25$, $z = -2.34$, $p = .019$, $r = -.50$: large effect). Additionally, a significant difference in the assessment of the Presentation simulation for Candidate 3 between the experimental group and the control

group ($U = 28$, $z = -2.14$, $p = .032$, $r = -.46$: medium effect) was found. All of the remaining differences between the experimental and control groups were non-significant.

DISCUSSION

This study focused on evaluating a frame-of-reference (FOR) training programme for assessors of an assessment centre. The main aim of the programme was to improve the evaluation and assessing skills of Industrial Psychology students doing their Honours degree.

According to the Health Professions Council of South Africa (HPCSA), a postgraduate student in the field of Industrial and Organisational Psychology should be able to assist in various assessment procedures over a diversity of settings and organisations. Hence the FOR training programme had the objective of improving the students' basic facilitation skills, their familiarity with the simulations implemented to measure the pre-determined competencies as well as their experience and familiarity with the competencies and the process of assessing. The results indicate that the FOR training programme increased the participants' familiarity with the one-on-one simulation and partially with the group discussion simulation. However, no significant results can be reported for the participants' ability on assessing the presentation simulation. This conclusion is supported by research that indicates that assessor expertise plays an important role in the accuracy of assessors (Jones & Born, 2008; Holmboe, 2004, Lievens, 2009 & Lievens, 1998). During the FOR training the least amount of attention was given to the presentation simulation. The partially significant results indicated by the group discussion simulation are confirmed in previous research (Melchers, Kleinmann & Prinz, 2010).

Concerning the first objective of this study, namely to conceptualise assessment centres and assessors of assessment centres from the literature, literature indicates that assessment centres is a process that utilises multiple work-like exercises to measure multiple dimensions, which is pertinent to effective performance in a specific position (Hoffman, Melchers, Blair, Kleinmann & Ladd, 2011; Thornton & Rupp, 2005). Assessors of an assessment centre can be viewed as the individuals who, through the multiple work-like exercises, observe, classify and evaluate the multiple dimensions displayed by the candidates taking part in the multiple exercises (Goodstone & Lopez, 2001; Hoffman et al. 2011; International Task Force on Assessment Centre Guidelines, 2010; Schlebusch, 2008). Training of these assessors focuses on developing certain abilities of the assessors to equip them in respect of observing,

assessing and classifying candidates' behaviour effectively and accurately (Schlebusch, 2008). A technique often used for assessor training is frame-of-reference (FOR) training (Jackson, Atkins, Fetcher & Stillman, 2005; Lievens, Tett & Schleicher, 2009). Frame-of-reference training aims at developing a mutual understanding between the assessors with regard to the competencies required for the specific assessment centre (Lievens, 2001; Lievens et al., 2009; Schleicher et al, 2002).

The second objective of this study was to investigate the content and methodology of a frame-of-reference training programme for assessors of assessment centres. This resulted in the compilation of a three-day training programme consisting of five (5) separate workshops. An already existing assessor training programme by Herman Spangenberg (1997) was adapted for the required context. The first day of the training consisted of two workshops. The first workshop focused on basic facilitation skills such as listening, objective attending and paraphrasing. A helping skills programme providing training in facilitation skills by Du Preez and Jorgensen (2012) was utilised and adapted to fit to the context and specifically the interviewing purposes of the assessor training programme. Research indicates that these competencies are important for assessors to be effective in assessing assessment centres (An international survey of assessment centre practices, 2008).

The second workshop focused on informing the participants of the principles of an AC as well as in-depth discussions of the competencies being assessed. The discussion concerning the principles of an AC focused on the objective of an AC, the reasons for using an AC, characteristics of an AC, different simulations that can be implemented in an AC as well as the role and duties of the assessor. Specifically the simulations utilised in the specific training and certain skills such as Observe, Record, Classify and Evaluate (ORCE) necessary for assessing were also discussed. The competencies being utilised in the training were also defined and discussed in terms of certain behavioural indicators of these competencies, and the scoring sheet was also discussed and explained extensively. The competencies utilised in this study were strategic perspective, interpersonal skills, leadership, conflict management, judgement, self-confidence, assertiveness, persuasive communication, performance under pressure, adaptability, ability to follow instructions, information usage, oral communication and technical and professional knowledge. This corresponds with previous studies on the requirements of a FOR training programme as well as research on the content of typical

assessor training (Melchers, Lienhardt, Von Aardburg & Kleinmann, 2011; Bernardin, Buckley, Tyler & Wiese, 2000; Schlebusch, 2008 Sulsky & Kline, 2007).

During the second day of the training programme, practical exercises were conducted on rating the 4 candidates (role-players) taking part in two separate one-on-one simulations on the competencies (strategic perspective, interpersonal skills, leadership, conflict management, judgement, self-confidence, assertiveness, persuasive communication and performance under pressure). The content of this training further corresponds with studies regarding FOR training (Bernardin et al., 2000; Melchers, Lienhardt, Von Aardburg & Kleinmann, 2011; Sulsky & Kline, 2007). Seeing that this was an entirely new simulation to the candidates (they had previously been exposed to presentations and group discussion simulations during their studies) the most time was spent on the one-on-one simulation.

On the third day of the training programme two workshops were presented; first focussing on practical exercises on rating the same candidates taking part in two different presentation simulations on the discussed dimensions. The competencies observed in the presentation simulation were: ability to follow instructions, information usage, oral communication, technical and professional knowledge, judgement, self-confidence and performance under pressure. The last workshop focused on practical exercises on rating two group discussion simulations consisting of four candidates each. The competencies being observed in the group discussion simulation were strategic perspective, interpersonal skills, leadership, judgement, self-confidence, assertiveness, persuasive communication, performance under pressure and adaptability. This concurs with research on the principles and content of a FOR training programme which indicate the providing of practical exercises and providing feedback on these practice ratings (Bernardin et al., 2000; Melchers, Lienhardt, Von Aardburg & Kleinmann, 2011; Sulsky & Kline, 2007).

Concerning the results of the third objective, to evaluate a frame-of-reference training programme for assessors of assessment centres; it was found that the experimental group improved in their evaluation and assessing skills.

Firstly the internal consistency for the AC was investigated. The ratings for the competencies of the AC for the experimental group in the pre-rest all showed high reliabilities. Similarly, in

the post-test for this group, all the competencies showed high reliabilities. However, the presentation simulation for Candidate 3 had a relatively low reliability.

The ratings for the competencies of the AC for the control group in the pre-test all showed high reliabilities as well. However the Group discussion simulation for Candidate 3 had a relatively low reliability. For the one-on-one simulation, a significant increase in internal consistency could be reported for all three candidates. This could indicate that with the post-test the reliability of the AC increased, indicating a positive effect of the FOR training programme. However, no significant increases could be reported for the presentation simulation. A decrease could in fact be reported for the presentation simulation during the post-test. The reliabilities for the presentation simulation were still relatively high. This could indicate that although there was a decrease for the post-test, the measure was still reliable, meaning that the FOR training programme, out of the three simulations, showed the least amount of improvement for the presentation simulation. However, the reliabilities for the post-test for this group are all within the acceptable range. This indicates that the experimental group had a good reliability for the pre-test, which made an increase in the post-test even more significant(Lievens, 2002; 2009; Schleicher et al., 2002).

The results of the pre-test showed that no significant differences in the rating of the AC existed between the experimental and control groups which could have had an influence on the training programme. This indicates that the two groups were at the same level concerning knowledge of FOR training prior to the programme being implemented.

After the experimental group had received the FOR training, the results indicated that the participants improved in their rating of Candidate 2 for the One-on-One simulation. This is an indication that the way in which the participants rated this simulation were similar (the ratings became closer between the participants). The same result was found for the Group Discussion simulation for two of the three Candidates. Furthermore, the results for the control group indicated no statistically significant differences between the pre- and post-test for the AC. It can therefore be concluded that the FOR training programme did in fact improved the mutual understanding and definition of the competencies being assessed for each candidate with the experimental group. Previous studies regarding FOR training support this finding by stating that FOR training promotes a mutual understanding between assessors (Jackson et al., 2005; Lievens, 2001; Lievens et al., 2009; Schleicher et al., 2002; Thornton & Rupp, 2005).

A study done by Melchers, Kleinmann and Prinz (2010) found that a group discussion simulation is one of the most difficult simulations to evaluate. Since there are multiple candidates as well as multiple dimensions that have to be evaluated simultaneously, this could cause cognitive overload. For the experimental group to then have improved most on the group discussion simulation could be an indication that the FOR training programme had a significant influence on the experimental group.

No constant differences in the ratings of certain competencies could be reported during the presentation simulation for the experimental group. One could speculate that a possible reason could be the fact that, during the FOR training, the least amount of time was spent on the rating of the presentation simulation since the participants claimed to feel more confident in the rating of the content quicker than with the other simulations. Previous research (Goodstone & Lopez, 2001; Lievens, 2001; Schlebusch, 2008; Hoffman et al. 2011 & International Task Force on Assessment Centre Guidelines, 2010) supports the finding that the amount of training for a specific simulation could have an effect on the reliability of the rating for that simulation.

Another observation made during the training on the presentation simulation was that the scenarios used in the training were not always feasible to truly measure the competencies for an Honours student. Possible explanations for this phenomenon could be that the scenarios for the simulation did not make it possible for the candidates to portray the competencies as defined for an Industrial Psychology Honours student. The scenarios required specific expertise on other subjects, not Industrial Psychology-related. The presentation was also not delivered in front of an audience but only recorded. This made it difficult for candidates to portray the industrial-specific competencies required for the simulation. Therefore there was some difficulty to truly practise the ratings for Honours student competencies.

The results further indicated significant differences for two simulations between the experimental group and the control group from the pre-test to the post-test, namely in the One-on-one simulation for Candidate 2, and of the Presentation simulation for Candidate 3. This implies that the two groups measured differently on only two simulations (ideally it would have been preferred that they would measure differently on more simulations; therefore showing a significant improvement for the participants compared to the control

group). One possibility is that these results can be explained as a stochastic error. Kahane (2008) explains a stochastic error as “variables or processes that are inherently random (i.e. not deterministic or exact)” (p. 218). Literature does support this finding: a similar finding was reported by Melchers, Lienhardt, Von Aardburg and Kleinmann (2011) indicating that with FOR training, in some instances, the training had shown a stronger effect on some aspects than on others.

Overall, there is an indication that among the experimental group there was more of a mutual understanding of the competencies after the FOR training. This indicates that there indeed was improvement in the assessing skills of graduate Industrial Psychology students in assessment centres. In conclusion it can be stated that the FOR training programme had an effect on the reliability of the ratings awarded by the assessors. As stated in previous studies, accurate rating by assessors could have a significant effect on the construct validity of the AC, and therefore this FOR training could have extensive positive and practical implications for various AC processes.

REFERENCE LIST

- American Psychological Association. (2010). *Ethical Principles of Psychologists and Code of Conduct 2010 Amendments*. Retrieved May 4, 2011, from <http://www.apa.org/ethics/code/index.aspx>.
- An International Survey of Assessment Centre Practices. (2010). The Global Research Questionnaire. Surrey: Assessment & Development Consultants Ltd.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.
- Bernardin, H. J., Buckley, M. R., Tyler, C. L., & Wiese, D. S. (2000). A reconsideration of strategies for rater training. *Research in personnel and human resources management, 18*, 221-274.
- Bryman, A., & Bell, E. (2011). *Business research methods*. Oxford, UK: Oxford University Press.
- Burns, N., & Grove, S. K. (1987). *The practice of nursing research*. Philadelphia, PA: WB Saunders.
- Cohen J. (1988). *Statistical power analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct of assessment centre evaluations? *International Journal of Selection and Assessment, 11*(1), 17-29.
- De Vos, A. S., Strydom, H., Fouche, C. B., & Delpont, C. S. L. (2005). *Research at Grass roots for the social sciences and human services professions* (3rd ed.). Pretoria: Van Schaik publishers.
- Dilchert, S., & Ones, D. S. (2009). Assessment centre dimensions: individual differences correlates and meta-analytical incremental validity. *International Journal of Selection and Assessment, 17*(3), 254-270.
- Du Preez, J. & Jorgensen, L.I. (2012). The evaluation of a helping skills training programme for intern-psychometrists. *Journal of Psychology in Africa*, in press.
- Eurich, T. L., Krause, D. E., Cigularov, K., & Thornton, G. C. III. (2009). Assessment centres: Current practice in the United States. *Journal of Business & Psychology, 24*, 387-407.

- Goodstone, M. S., & Lopez, F. E. (2001). The frame of reference approach as a solution to an assessment centre dilemma. *Consulting Psychology Journal: Practice and Research*, 53(2), 96-107.
- Guion, R. M. (1998). *Assessment, Measurement, and Prediction for Personnel Decisions*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hauenstein, N.M.A., & Foti, R.J. (1989). From laboratory to practice: Neglected issues in implementing frame-of-reference rater training. *Personnel Psychology*, 42, 359-378.
- Health Professions Council of South Africa. (2010). Health Professions Act 1974 (Act 56 of 1974). Retrieved 8 June, 2011, from <http://www.hpcs.co.za>.
- Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centres. *Personnel Psychology*, 64, 351-395.
- Holmboe, E. (2004). Faculty and the observation trainees' clinical skills: problems and opportunities. *Academic Medicine*, 79(1), 16-22.
- International Task Force on Assessment Centre Guidelines. (2010). *Guidelines and ethical considerations for assessment centre operations*. San Francisco, CA: Development Dimensions International.
- 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 behaviours and ability traits. *Public Personnel Management*, 34(1), 17-30.
- Joiner, D. A. (2004, Jun 20-23). Assessment Centre trends: assessment centre issues and resulting trends. *Paper presented at the 28th annual meeting of IPMAAC*, Seattle, Washington, United States of America.
- Jones, R. G., & Born, M. P. (2008). Assessor constructs in use as the missing component in validation of assessment centre dimensions: A critique and directions for research. *International Journal of Selection and Assessment*, 16(3), 229-238.
- Kahane, L.H. (2008). *Regression basics* (2nd ed.). CA: Sage.
- Krause, D. E., & Gebert, D. (2003). A comparison of assessment centre practices in organisations in German speaking regions and the United States. *International Journal of Selection and Assessment*, 11, 297-312.
- 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. (2002). An examination of the accuracy of slogans related to assessment centres. *Personnel Review*, 31, 86-102.
- Lievens, F., & Thornton, G. C. III. (2005). Assessment centres: recent developments in practice and research. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell Handbook of Personnel Selection* (pp. 243-264). Malden, MA: Blackwell.
- Lievens, F. (2009). Assessment centres: a tale about dimensions, exercises and dancing bears. *European Journal of Work and Organisational Psychology*, 18(1), 102-121.
- Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centres at the crossroads: toward a reconceptualization of assessment centre exercises. *Research in Personnel and Human Resources Management*, 28, 99-152.
- Maree, K. (2007). *First steps in research*. Pretoria: Van Schaik Publishers.
- Meiring, D. (2008). Assessment centres in South Africa. In S. Schlebusch, & G. Roodt (Eds.), *Assessment Centres: Unlocking potential for growth* (pp. 21-31). Randburg, South Africa: Knowres Publishing (Pty) Ltd.
- Melchers, K. G., Kleinmann, M., & Prinz, M. A. (2010). Do assessors have too much on their plates? The effects of simultaneously rating multiple assessment centre candidates on rating quality. *International Journal of Selection and Assessment*, 18(3), 329-341.
- Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology*, 64, 53-87.
- Pallant, J. (2010). *SPSS Survival Manual. A step by step guide to data analysis using SPSS* (4th ed). Berkshire, England: McGraw-Hill
- Pell, G., Homer, M. S., & Roberts, T. E. (2008). Assessor training: its effect on criteria-based assessment in a medical context. *International Journal of Research & Method in Education*, 31(2), 143-154
- Salkind, N. J. (2009). *Exploring research* (pp. 243-251). Saddle River, NJ: Pearson Education Inc.
- Schlebusch, S. (2008). Before the Centre. In S. Schlebusch & G. Roodt (Eds.), *Assessment Centers: Unlocking potential for growth* (pp. 176-196). Randburg, South Africa: Knowres Publishing (Pty) Ltd.
- 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.
- SPSS Inc. (2008). *SPSS 16.0 for Windows*. Chicago, IL.

- Struwig, F. W., & Stead, G. B. (2001). *Planning, designing and reporting research*. Cape Town, South Africa: Pearson Education South Africa.
- Sulsky, L. M., & Kline, T. J. B. (2007). Understanding frame-of-reference training success: a social learning theory perspective. *International Journal of Training and Development*, *11*(2), 121-131.
- Thornton, G. C. III, Murphy, K. R., Everest, T. M., & Hoffman, C. C. (2000). Higher cost, lower validity and higher utility: comparing the utilities of two tests that differ in validity, costs, and selectivity. *International Journal of Selection and Assessment*, *8*, 61-75.
- Thornton, G. C. III., & Mueller-Hanson, R. (2004). *Developing organisational simulations: A guide for practitioners and students*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Thornton, G. C., & Rupp, D. E. (2005). *Assessment Centres in human resource management: Strategies for prediction, diagnosis and development*. Mahwah, NJ: Lawrence Erlbaum.

CHAPTER 3

CONCLUSIONS, LIMITATIONS AND RECOMMENDATIONS

In this chapter, conclusions regarding the study are given in accordance with the general and specific objectives. The limitations of this research are discussed, followed by recommendations for the organisation and future research.

3.1 CONCLUSIONS

The main focus of this study was to evaluate a frame-of-reference (FOR) training programme for assessors of assessment centres (ACs). A FOR training programme for assessors was compiled, which aimed at improving the overall evaluation and assessing skills of postgraduate Industrial Psychology students. The results revealed that the FOR training programme did in fact have an effect on the experimental group's participants' level of comfort with specifically the one-on-one simulation and the group discussion simulation. For the control group, however, no significant differences could be reported for the post-test. This indicates that no improvement was recorded for these participants between the pre- and post-tests. This further supports the statement that the FOR training programme had a significant effect on the common understanding and mutual frame-of-reference of the experimental group.

The first objective of this study was to conceptualise assessment centres and the assessors in an assessment centre according to previous literature. Hoffman, Melchers, Blair, Kleinmann and Ladd (2011) point out that during an AC, a candidate is exposed to a variety of exercises, which resemble typical work-like simulations as can be expected in the specific position being assessed, while being observed and evaluated by a group of assessors. These work-like simulations are specially designed to activate certain behaviours in a candidate; the reason being that these behaviours can be observed and evaluated in order for the group of assessors to make a recommendation based on the AC process (Goodstone & Lopez, 2001; Schlebusch, 2008; Hoffman et al., 2011; International Task Force on Assessment Centre Guidelines, 2010). To be able to observe these behaviours, these assessors need to be trained (Goodstone & Lopez, 2001; Lievens, 2009; Schlebusch, 2008). The training of assessors should thus focus on improving their ability to observe, record, classify and evaluate (ORCE) manifested behaviours of candidates during any assessment centre simulation (Schlebusch, 2008). A technique that can be implemented in such an instance is that of frame-of-reference (FOR)

training, of which the objective is to create and establish a common understanding among the assessors of the competencies required in the AC as well as how to identify certain competencies (Jackson, Atkins, Fetcher & Stillman, 2005; Lievens, 2002; Lievens, Tett & Schleicher, 2009; Schleicher, Day, Mayes & Riggio, 2002).

The second objective of this study aimed at researching the content and methodology involved in creating a FOR training programme. An already existing training programme for assessors of assessment centres was used as an example and adapted to incorporate FOR principles specific to the required context. A helping skills programme by Du Preez and Jorgensen (2012) was also incorporated to equip the assessors with the necessary facilitation skills such as listening, objective attending and paraphrasing, as stipulated by an international survey of assessment centre practices (2008). This resulted in a 3-day training programme with 5 separate workshops. Workshop 1 entailed training, exercises and discussion on basic facilitation skills, as indicated above. Workshop 2 consisted of background information on ACs, characteristics of ACs, reasons for using ACs, objectives of ACs and simulations used in AC practices. In this workshop the competencies utilised in this study and certain behavioural indicators of these competencies as well as the scoring sheet were also discussed and explained in detail. The first two workshops were based on previous FOR research indicating certain characteristics of FOR training (Bernarding, Buckley, Tyler & Wiese, 2000; Melchers et al., 2011; Schlebusch, 2008; Sulsky & Kline, 2007).

The second day of training consisted of workshop 3, which entailed the opportunity to practically evaluate and rate 'make believe' candidates in two separate AC's on the specified competencies for the one-on-one simulation. The third day of training afforded the participants an opportunity to rate once again 'make believe' candidates in two separate AC's on the pre-determined competencies for the presentation as well as group discussion simulation. This was also designed according to previous literature on FOR principles (Bernarding et al., 2000; Melchers et al., 2011; Schlebusch, 2008; Sulsky & Kline, 2007).

In regard to the third objective of this study, namely to evaluate a frame-of-reference training for assessors of assessment centres, the following results were revealed. It was found that the experimental group had improved in their familiarity during the one-on-one simulation and the group discussion. This was also found in a study done by Melchers et al. (2011) where

they found that FOR training indicated a stronger improvement in some instances than in other.

For the one-on-one simulation it appeared that the experimental group had a common understanding of the competencies being measured, more specifically Candidate 2. This indicates that during the post-test the experimental group had a more definite definition as well as understanding of the competencies being measured during the one-on-one simulation.

For the presentation simulation no significant differences could be reported for the post-test. The assumption can thus be made that the training programme did not show much of an improvement for this specific simulation. A possible reason for the lack of evidence in this instance could be that the least amount of time was spent on this simulation, since the participants claimed to feel comfortable with the simulation seeing that they had previous personal experience with presenting. That the amount of training can significantly affect the accuracy and reliability of the ratings, has been found by various previous researchers (Goodstone & Lopez, 2001; Hoffman et al., 2011; International Task Force in Assessment Centre Guidelines, 2009, Lievens, 2002; Schlebusch, 2008). Another explanation can be that the simulation used in the training of ACs were subject specific and the subject was not Industrial Psychology. This means that the 'make believe' candidates had difficulty truly exhibiting the competencies required for the presentation simulation which in turn made it difficult for the participants to rate these behaviours, ultimately concluding in insufficient exposure to and practice of ratings for the competencies required for the presentation simulation.

In the group discussion simulation there seemed to be a difference in the ratings for the experimental group during the post-test for two of the three candidates, which could be a result of a more mutual frame-of-reference regarding the competencies being measured in the simulation. An interesting finding is that the competencies evaluated in the one-on-one simulation correlate with the competencies measured in the group discussion. This could indicate that the experimental group had a more mutual definition and understanding of the various competencies. Melchers et al. (2010) also mention that the group discussion is one of the most difficult simulations to evaluate within an AC since it consists of multiple candidates that have to be observed simultaneously. It is thus a very good indicator of success

for the FOR training programme that the experimental group indicated the most significant results for the group discussion simulation.

The results further indicated that although the experimental group showed significant differences between the pre- and post-test, indicating a significant influence, there were only limited differences between the ratings of the experimental group and the control group. This could indicate that either the participants had a natural inclination to rating behaviour or they had previous knowledge of the subject matter.

Seeing that the participants were Industrial Psychology postgraduate students, it could possibly be that they had a natural instinct for rating behaviour as they were in the field of behavioural sciences. As previously mentioned, they were also Honours students, and the competencies being evaluated in the pre- and post-test were Honours students' competencies. Therefore they could have had personal experience with and knowledge of the competencies being measured. That psychology students have good rating ability has been found by previous research (Jackson et al, 2005).

However, overall there was a definite indication that the experimental group had a more mutual understanding of some of the competencies being measured in the specific AC. This in turn proves that the related FOR training programme for assessors of assessment centres does improve the assessing skills of graduate Industrial Psychology students.

3.2 LIMITATIONS OF THIS RESEARCH

With regard to the limitations of the present study, the following can be outlined:

Firstly the length of the training should be extended for the participants to gain more practical experience to be able to rate efficiently and accurately on all nine competencies. According to the Guidelines for assessment and development centres in South Africa (2007), training for assessors could vary between four and five days. Although this is the guideline, it could be extended to 5-7 days. Although the current programme consisted of three days of intensive training, the cognitive load was immense. It can be concluded from the feedback received from participants that the cognitive demands during the three days of training were excessive; thus they could not always gain the correct skills and experience from the process.

Secondly, the participants were postgraduate students in Industrial Psychology and, although it does not form part of their course work, they had previous knowledge, though limited, of assessment centres. This could have an effect on the experiment seeing that the pre-test results were already predominantly positive and it was challenging to prove significant differences in the post-test results. The previously mentioned phenomenon is supported by the results gained from the presentation and group discussion's pre- and post-test results. During their course work, the population had extensive exposure to evaluating classmates in presentations and even during group work. This information only surfaced during the post-test and with further research it was discovered that they had an instinctual reaction to evaluating and rating behaviour, which could be expected, seeing that they were students in a behavioural field.

Thirdly, the training occurred approximately two to three weeks before the university's official Master's selection for the Industrial Psychology programme would take place. The majority of the population group applied for this programme and were informed that during the Master's selection they would be exposed to an assessment centre. During the training it could be observed that some of the participants were aiming to prepare themselves for the Master's selection and focussed on learning from the training regarding the assessment centre process. It could be observed by the researcher that the population saw themselves in the role of assessor rather than in the role of assessee and assessed themselves against the competencies. This significantly influenced the entire training process, seeing that these candidates did not reach the goal of the training programme, which was to become a better assessor – they were constantly evaluating themselves.

Lastly, the sample size ($N=22$) could be seen as a possible limitation of this study. Seeing that, with such a small sample size, any single irregular rating could influence the interpretation of the results. However, the entire population available was utilised and a larger sample size was not possible.

3.3 RECOMMENDATIONS

Despite these limitations, the research findings have important implications for organisations concerned, as well as for future research.

3.3.1 Recommendations for the organisation

The HPCSA (2010) expects industrial psychologists and psychometrists employed in organisations to be able to assist in diverse assessment procedures in various organisations; therefore it is critical for any HR staff member (human resource practitioners, industrial psychologists and psychometrists) to be efficiently trained to be able to act as an assessor in various assessment situations. In an International Survey of Assessment Centre Practices (2008) done across Europe, Oceania, Africa, Asia and United States, it was shown that 61% of assessors in assessment centres are HR staff, with line managers making up 53%. Although the Guidelines for assessment and development centres in South Africa (2007) stipulate that best practice would entail four to five days of training, studies indicate that only 23% of organisations provide 1.5 – 2 days of training, while 14% provide 0.5 – 1 days of training and 9% provide 1 – 1.5 days of training. This calculates to 46% of training for assessors being less than 2 days, with 40% of organisations providing more than 2 days of training for their assessors.

Since from the above-mentioned the most frequently used assessors in AC's are HR staff and sufficient training is not provided, it is recommended that all Honours and Master's students undergo FOR training before graduating and being exposed to the labour market. According to the International Task Force on Assessment Centre Guidelines (2010), any assessor should attend a refresher course within six (6) months from their original training; thus a further recommendation is that a shortened version of the FOR training is also developed and administered as a total product together with the longer version of the FOR training.

3.3.2. Recommendations for future research

For future research on a FOR training programme, a possible suggestion would be to use a different population group that has not had any prior behavioural science training. The reason for this recommendation is that during this research a definite limitation was that the population group had previous knowledge and experience with regard to assessment practices. As seen in the International Survey of Assessment Centre Practices (2008), although HR members of staff are mostly being used as assessors, line managers comprise 53% of assessors and members of staff with expertise make up 27% of assessors being used. A larger population would also be recommended, as for this research there were only 22 participants. The sample size can have an effect on the statistics, seeing that with a small

sample size one participant can drastically influence the descriptive statistics; this is eliminated with a bigger sample size. Therefore by making use of a different population group will enhance the research by making the results more generalizable as well as more relevant to the corporate labour market and assessment practices.

Another recommendation for this research would be to extend the FOR training implementation to a four- (4) day training programme rather than the current three (3) days of training. This will result in reducing the cognitive load for first-time assessors being trained which in turn could result in the trainees gaining more from the training since they are not so cognitively drained at the end of the training day. By extending the training programme it could also result in dedicating a day to focusing on the one-on-one simulation, presentation and group discussion each, as well as an entire day on technical background and competencies being measured in the specific AC. This could have a significant effect on the results obtained.

Another recommendation with regard to the training programme would be to utilise independent and totally objective role-players in the pre- and post-test as well as in the training programme. The reason is that in the present study Industrial Psychology Master's students were utilised. This made it difficult to give examples of appropriate behaviour for low ratings, seeing that they were competent at all the desired competencies (seeing that it was Honours student competencies). Some of the participants also knew the role-players and therefore could have made emotive ratings even after the training as it was more personal or because they knew they were Master's students and therefore, when in doubt, gave a rating based on the expectation of a Master's student, rather than on the actual performance.

Thirdly, future research should limit the candidates being assessed in the group discussion simulation to only two candidates. This could be done by still implementing a group discussion with three candidates, but adding a role-player into the three candidates that will then expect of the candidates to only observe two candidates during the simulation. During this research it was observed that having to observe three candidates during the group discussion simulation could have been an over-extension of their ability as it is best practice to have an assessor only observe two candidates at a time (International task force on Assessment Centre Guidelines, 2010). Another recommendation for future research is to also look into the design of the AC simulations being used. Together with a better design and an

effective training programme for assessors the construct validity of an AC could be significantly improved

Fourthly the longitudinal effect of a FOR training programme for assessors of assessment centres can be investigated. Two consequences can be derived from this recommendation. Firstly, Schlebusch (2008) states that a trainee assessor should, after observing an assessment centre and attending lecture room training, observe at least two assessment centres as trainee assessor being supervised by an experienced assessor. Only once the supervising assessor, AC administrator and members of the assessor team agree that the trainee assessor is competent, can the trainee assessor be seen as an assessor (Schlebusch, 2008). During this research the candidates only had exposure to the pre-test before undergoing lecture room training where they were afforded the opportunity to observe two other AC's. However, their only opportunity to truly exhibit their newly obtained skills was with the post-test, which can be seen as their first AC observed as trainee assessors. Since the results obtained from the post-test are being used in this research, the fact that according to various researchers this process, should still form part of the trainee assessor's training, it is possible that there could be more significant results had this study been carried out longitudinally and over a period of time. This in turn re-enforces the recommendation that an alternative population as well as the environment for the training programme should be considered.

REFERENCES

- An International Survey of Assessment Centre Practices. (2008). *The Global Research Questionnaire*. Surrey: Assessment & Development Consultants Ltd.
- Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.
- Bernarding, H. J., Buckley, M. R., Tyler, C. L., & Wiese, D. S. (2000). A reconsideration of strategies for rater training. *Research in personnel and human resources management, 18*, 221-274.
- Du Preez, J. & Jorgensen, L.I. (2012). The evaluation of a helping skills training programme for intern-psychometrists. *Journal of Psychology in Africa*, in press.
- Goodstone, M. S., & Lopez, F. E. (2001). The frame of reference approach as a solution to an assessment centre dilemma. *Consulting Psychology Journal: Practice and Research, 53*(2), 96-107.
- Health Professions Council of South Africa. (2010). Health Professions Act 1974 (Act 56 of 1974). Retrieved June 8, 2011, from <http://www.hpcs.co.za>.
- Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R.T. (2011). Exercises and dimensions are the currency of assessment centres. *Personnel Psychology, 64*, 351-395.
- 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 behaviours and ability traits. *Public Personnel Management, 34*(1), 17-30.
- International Task Force on Assessment Centre Guidelines. (2010). *Guidelines and ethical considerations for assessment centre operations*. San Francisco, CA: Development Dimensions International.
- Krause, D. E., & Gebert, D. (2003). A comparison of assessment centre practices in organisations in German speaking regions and the United States. *International Journal of Selection and Assessment, 11*, 297-312.
- Lievens, F. (2002). An examination of the accuracy of slogans related to assessment centres. *Personnel Review, 31*, 86-102.
- Lievens, F. (2009). Assessment centres: a tale about dimensions, exercises and dancing bears. *European Journal of Work and Organisational Psychology, 18*(1), 102-121.

- Lievens, F., Tett, R. P., & Schleicher, D. J. (2009). Assessment centres at the crossroads: toward a reconceptualization of assessment centre exercises. *Research in Personnel and Human Resources Management, 28*, 99-152.
- Melchers, K. G., Lienhardt, N., Von Aarburg, M., & Kleinmann, M. (2011). Is more structure really better? A comparison of frame-of-reference training and descriptively anchored rating scales to improve interviewers' rating quality. *Personnel Psychology, 64*, 53-87.
- Schlebusch, S. (2008). Before the Centre. In S. Schlebusch & G. Roodt (Eds.), *Assessment Centers: Unlocking potential for growth* (pp. 176-196). Randburg, South Africa: Knowres Publishing (Pty) Ltd.
- 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.
- South African Guidelines (2007, March). *Guidelines for assessment and development in South Africa*. Assessment Centre Study Group. Stellenbosch: Assessment Centre Study Group Conference.
- Sulsky, L. M., & Kline, T. J. B. (2007). Understanding frame-of-reference training success: a social learning theory perspective. *International Journal of Training and Development, 11*(2), 121-131.
- Thornton, G. C. III, Murphy, K. R., Everest, T. M., & Hoffman, C. C. (2000). Higher cost, lower validity and higher utility: comparing the utilities of two tests that differ in validity, costs, and selectivity. *International Journal of Selection and Assessment, 8*, 61-75.