

Full Length Research Paper

Construct validity of competency dimensions in a leadership assessment and development centre

Zovuyo Nako and Antoni Barnard*

Department of Industrial and Organisational Psychology, University of South Africa (UNISA), P. O. Box 392, 0003, South Africa.

Accepted 3 August, 2012

The issue of construct validity has become a contentious issue in the study of assessment centres around the world. The purpose of the study was to investigate the construct validity of assessment centre dimension ratings through correlation and factor analysis. The sample consisted of 138 individuals who participated in a two-day assessment centre for selection as partners/directors in an auditing firm. Twenty-one dimensions were measured using six different exercises. Both correlation and factor analysis results showed no evidence of discriminant validity amongst dimensions measured in the simulation and interview exercises. Convergent validity among some dimensions was found only in one of the simulation exercises. Implications for assessment centre design and research have been discussed.

Key words: Construct validity, assessment centre, convergent validity, discriminant validity, multi-trait multi-method approach, competency based assessment.

INTRODUCTION

Fair psychological assessment practices significantly contribute to the employment, retention and development of talent in organisations. In addressing the issue of fairness, South African labour law requires psychological assessment in the workplace to be based on job-related competencies (Foxcroft and Roodt, 2001). As a result, the use of competency-based-assessment (CBA) as a strategy to ensure fair assessment practices in the workplace has risen (Eurich et al., 2009; Potgieter and Van der Merwe, 2002). With competencies as its foundation (Potgieter and Van der Merwe, 2002; Thornton and Rupp, 2006) assessment centres (ACs) are regarded as a fair method of workplace assessment for selection and development purposes (Amen, 2010; Eurich et al., 2009; Schlebusch and Roodt, 2008; Thornton and Krause, 2009). ACs are generally believed to improve the productivity and effectiveness of organisations (Bartram, 2004). Due to their reported

multicultural appropriateness (Schlebusch and Roodt, 2008) ACs are ideal for selection, demonstrating less adverse impact compared to cognitive ability tests and to interviews (Dean et al., 2008; Harel et al., 2003; Thornton and Rupp, 2006). AC utility research has however yielded varying results generally indicating good predictive validity yet inconsistent support for construct validity.

Construct validity being fundamental to fair assessment (Taylor and Boeyens, 1991), highlights the importance of continuous AC research as it has implications for improved AC design and application.

This study intended to extend construct validity research in the South African context by studying the construct validity of competency dimensions measured in a leadership assessment and development centre (LADC) that was designed for selection of partners/directors in an auditing firm.

THEORETICAL BACKGROUND

Owing to their ability to predict future job performance (Gaugler et al., 1987; Jansen and Stoop, 2001; Jones

*Corresponding author. E-mail: barnaha@unisa.ac.za. Tel: 0823752696. Fax: 0866652696.

and Born; 2008; Lance, 2008), ACs have been in use in many public and private sector organisations over the world (Eurich et al., 2009; Harel et al., 2003). An AC is a standardised approach of evaluating behaviour observed by a number of trained assessors (Amen, 2010; Joiner, 2000; Thornton and Rupp, 2006). In an AC, job-related behavioural dimensions or competencies (for example, interpersonal skills, leadership skills, problem-solving) are measured through a combination of assessment methods (here referred to as AC exercises) such as psychometric tests, interviews and simulation exercises (Amen, 2010; Joiner, 2000). Thus, repetitive information about participants' distinct work-related attributes is obtained (Haaland and Christiansen, 2002; Joiner, 2000) implying the simultaneous measurement of various job-related competencies, which are otherwise difficult to measure (Thornton and Rupp, 2006). Based on the target job, simulation exercises used usually resemble real-life work situations and may include role plays, presentation exercises and/or case studies (Amen, 2010; Joiner, 2000). Given the use of multiple dimensions and exercises observed by more than one assessor, the approach has been termed the multi-method multi-trait (MTMM) approach (Campbell and Fiske, 1959).

Predictive validity evidence for ACs abound (Gaugler et al., 1987; Harel et al., 2003; Jansen and Stoop, 2001; Jones and Born, 2008; Thornton and Rupp, 2006). In particular Harel et al. (2003) reported on 17 reviews of meta-analytical studies, in which the AC method was ranked with the highest validity when job performance ratings are used as criteria (Greyling et al., 2003).

Despite the predictive validity evidence of ACs, a lack of construct validity in ACs has been highlighted in numerous studies (Arthur et al., 2000; Kleinmann and Koller, 1997; Shore et al., 1990; Thornton and Rupp, 2006). Construct validation is a process of investigating whether the test measures the constructs that it claims to measure and includes evidence of both convergent and discriminant validity (Haaland and Christiansen, 2002). Several approaches have been used to study the construct validity of AC ratings. Shore et al. (1990) supports Campbell and Fiske's (1959) MTMM approach to determine convergent and divergent validity of multiple dimensions across multiple exercises. MTMM findings that support convergent validity of ACs have been made, yet mostly show a lack of discriminant validity (Greyling et al., 2003; Jones and Born, 2008; Lance, 2008).

Within the MTMM approach, some researchers ascribe the lack of discriminant validity in ACs to method variance (Arthur et al., 2008; Greyling et al., 2003; Kleinmann and Koller, 1997), implying that although ACs were designed to rate individuals on competency dimensions measured across various exercises, the focus seems to have shifted to the rating of an individual's competence within a particular exercise instead (Lance, 2008; Meriac et al., 2008).

Factor analysis has also been used to study AC

construct validity to establish whether the underlying factor structure of competency ratings strongly represent the AC dimensions or exercises (Arthur and Woehr, 2003; Damitz et al., 2003; Fleenor, 1996; Greyling et al., 2003; Kaufmann et al., 1993; Shore et al., 1990). A dimension rating/factor consists of ratings for one dimension in different exercises whereas an exercise rating/factor consists of the ratings of more than one dimension within one exercise. Different to the study by Fleenor (1996), some factor analytical studies have shown adequate support for construct validity evidence of an AC (Damitz et al., 2003; Shore et al., 1990).

In an AC, competency dimensions should represent stable behavioural categories within exercises and be consistent across exercises (Lance, 2008; Lievens and Conway, 2001). Greyling et al. (2003) agree that ACs are designed with the expectation of agreement on same dimension ratings across exercises and differentiation between dimension ratings within a particular exercise. Construct validity is therefore evidenced in strong correlations between same dimension ratings across exercises (that is, convergent validity) and weak correlations between different dimensions within a specific exercise (that is, divergent validity). Similarly they note that factor analysis provides evidence of construct validity if similar dimensions' ratings cluster together rather than exercise ratings clustering together.

METHODS

Type of study and sample

In this study assessment centre data were gathered over four years and analysed from a quantitative-relational approach, with $n=138$. The correlation study explored the relationships between scores on various dimensions within and across different exercises in an AC conducted to identify partners/directors in an audit firm. Business unit managers, nominated by their senior executives for potential selection into partner or director position, participated in the LADC. The sample ranged from 26 to 52 years, including 49 female and 89 male participants. In terms of race the sample included 23 Asian, 30 Black, 10 Coloured and 75 White participants.

AC design

The AC methodology encompassed the assessment of 21 competency dimensions through three simulation exercises, two interviews and two psychometric tests namely, the cognitive process profile (CPP) and the occupational personality profile (OPQ32i). Specifically the AC included a simulated presentation exercise (SPS), a one-on-one executive interview (EXEC), an e-based in-basket simulation called the applied leader skills assessment (ALS), a competency-based panel interview (CBI), the two psychometric assessments (PSYCHE) and a written report called practice office automation (PON).

The competency dimensions resulted from a job profiling process within the company during which essential competencies required for success as an executive manager in auditing, aligned with the company's shared values, were identified. Simulations were developed in-house through a qualitative focus-group and critical incident methodology to reflect typical daily tasks of a partner in an

Table 1. Discriminant validity: SPS and EXEC competency dimensions intercorrelations.

SPS	Written communication	Oral communication	Personal impact
Written communication	1.00		
Oral communication	0.477**	1.00	
Personal impact	0.441**	0.763**	1.00
EXEC			
Written communication	1.00		
Oral communication	0.491**	1.00	
Personal impact	0.748**	0.748**	1.00

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

Table 2. Discriminant validity: Correlations between dimensions in ALS.

	Client orientation	Applied judgment and insight	Business acumen	Knowledge-up-to-datedness	Practice leadership	Winning business	Product and service oriented
Client orientation	1.00						
Applied judgment and insight	0.666	1.00					
Business acumen	0.657	0.570	1.00				
Knowledge-up-to-datedness	0.363	0.445	0.437	1.00			
Practice leadership	0.446	0.463	0.509	0.215	1.00		
Winning business	0.609	0.533	0.647	0.339	0.426	1.00	
Product and service oriented	0.385	0.355	0.447	0.401	0.373	0.485	1.00

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

auditing firm.

Behavioural anchors were assigned to each competency dimension in each simulation and rated on a 5-point Likert scale. Norm scores obtained on related competency dimensions in the PSYCH were adjusted to the 5-point Likert scale applied in the simulations.

Assessor panel and training

According to Joiner (2000) the reliability of an AC relates to assessor objectivity and therefore predominantly depends on using multiple assessors, standardised assessor training and standardisation of the conditions under which the assessments are applied. In the current AC assessors included two partners/directors, a psychologist, and three business unit managers. To enhance reliability all assessors attended a four-hour training workshop based on the assessor training model of Saunders (2000). Previous reliability and construct validity studies for the OPQ32i (SHL, 2000; 2001) as well as the CPP (Prinsloo, 2000) have reported good results.

In an attempt to counter halo effects, assessors rotated to observe different participants in different exercises. The across exercise rating approach (AER) is proposed for improved construct validity by Lance (2008), Lievens and Conway (2001) and Meriac et al. (2008). Here the AER was combined with the within exercise rating method (WER) to derive competency ratings as assessors consolidated their observations and ratings per dimension after each exercise and again per dimension once the AC was completed. Participants received written and oral feedback within two weeks after participating in the LADC.

Statistical procedure

Pearson product-moment correlation coefficients were calculated to obtain correlations between dimensions within LADC exercises and between same dimensions across exercises. AC ratings were then subjected to the principal component analysis (PCA), followed by a varimax rotation using SPSS, to cluster dimensions together per exercise and in the LADC as a whole.

RESULTS

Pearson product-moment correlation between dimensions within exercise

Pearson correlations in Table 1 between the three dimensions measured in the SPS and in the EXCEC were all significant at the 0.01 level of error. All the correlations between the dimensions assessed in the ALS (Table 2) were statistically significant at the 0.01 level, with a moderate correlation between practice leadership and knowledge-up-to-datedness ($r = 0.215$).

Only three correlations between dimensions measured in the CBI were statistically non-significant (Table 3) whereas a number of insignificant correlations were found in the PSYCH dimensions. In the CBI most correlations ranged from medium to large effects, with the highest correlation occurring between teamwork and

Table 3: Correlations between dimensions measured in the CBI.

	Interpersonal skills	Achievement drive	Team work	Respect for diversity	Negotiation	Resilience	Flexibility	Leadership	General applied judgement
Interpersonal skills	1.00								
Achievement drive	.360**	1.00							
Team work	.628**	.453**	1.00						
Respect for diversity	.306**	.271*	.374**	1.00					
Negotiation	.186*	.478**	.427**	.03	1.00				
Resilience	.381**	.478**	.425**	.222*	.359**	1.00			
Flexibility	.461**	.540**	.450**	.242*	.456**	.380**	1.00		
Leadership	.358**	.566**	.547**	.351**	.513**	.433**	.554**	1.00	
General applied judgement	.358**	.508**	.473**	.17	.479**	.421**	-0.05	.515**	1.00

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

interpersonal skills ($r = 0.628$). With the exception of statistically significant correlations between leadership and negotiation ($r = 0.501$), team work and interpersonal skills ($r = 0.366$), leadership and interpersonal skills ($r = 0.354$) and leadership and team work ($r = 0.432$), altogether 69% of the correlations between the PSYCH dimensions were statistically non-significant, indicating evidence of discriminant validity amongst the 69% of dimensions measured in the PSYCH.

43% (89) of the 209 correlations between the 21 PON dimensions were statistically non-significant, indicating evidence of discriminant validity amongst 43% of dimensions measured in the PON. Other inter-correlations range from moderately significant (30% of the correlations) to strong (28%), with the largest correlation between general judgment and applied judgment and insight ($r = 0.540$). Approximately 58% of the correlations in the PON therefore did not show evidence of discriminant validity.

Pearson product-moment correlation between same dimensions across exercises

Pearson product-moment correlation coefficients were used to calculate correlations between similar dimensions across exercises (that is, convergent validity). The results shown in Table 4 revealed mostly non-significant correlations (70% of the correlations) with some moderate correlations (5 of the 43 correlations, that is, 12%). Statistically significant correlations were found in 8 of the 43 inter-correlations (that is, 19%) with the correlation between the CBI and PSYCH for the leadership dimension that is the strongest ($r = 0.515$). Other statistically significant correlations across the CBI and PSYCH exercises are evident for interpersonal skills ($r = 0.482$), achievement drive ($r = 0.482$), teamwork ($r =$

0.389), respect for diversity ($r = 0.396$), negotiation ($r = 0.469$) and resilience ($r = 0.419$). A significant correlation also occurred between the EXEC and SPS for the written communication dimension ($r = 0.319$). These correlations are supportive of convergent validity across these two exercises.

Factor analysis

AC ratings were then subjected to PCA, followed by a varimax rotation to cluster competency dimensions together. The suitability of data for factor analysis was assessed prior to performing the PCA and the inspection of the correlation matrix revealed the presence of many coefficients of 0.5 and above. The PCA results reveal the presence of 17 factors, which explained 76% of the variance. All SPS ratings were loaded on the third factor, whilst the EXEC ratings were all loaded on the seventh factor. ALS ratings were loaded on the second factor, with the exception of product and client service, which was loaded on the sixteenth factor. Ratings of the CBI were loaded on the first factor, with the exception of respect for diversity, which was loaded on the twelfth factor. The PSYCH ratings were grouped on six factors. Teamwork and leadership were loaded on the tenth factor; flexibility and interpersonal skills were loaded on the third factor, while negotiation was loaded on the eleventh factor and respect for diversity on the twelfth factor. General judgment and achievement drive were loaded on the ninth factor, while resilience was loaded on the fifteenth factor. PON ratings were also loaded on ten different factors. General judgment, applied judgment, written communication and oral communication were loaded on the fourth factor. Teamwork, flexibility, interpersonal skills and leadership were loaded on the

Table 4: Correlations between dimensions across exercises (convergent validity).

Dimension		Exercise Method	
		SPS	EXEC
Written Communication	EXEC	.319**	
	PON	.12	.10
Oral Communication	EXEC	.16	
	PON	-0.12	.03
Personal Impact	EXEC	.14	
	PON	.13	.09
Interpersonal Skills	CBI		PSYCH
	PSYCH	.482**	
Achievement Drive	PON	.09	-0.03
	PSYCH	.482**	
Team Work	PON	.05	.04
	PSYCH	.389**	
Respect for diversity	PON	0.00	.203*
	PSYCH	.396**	
Negotiation	PON	.240*	.19
	PSYCH	.469**	
Resilience	PON	.17	.282*
	PSYCH	.419**	
Flexibility	PON	.09	-0.07
	PSYCH	.266*	
Leadership	PON	.14	.03
	PSYCH	.515**	
General Judgement	PON	.15	.11
	PSYCH	.234*	
Client Orientation	PON	.19	.12
	ALS	.07	
Applied Judgement & Insight	ALS	-0.04	
Business Acumen	ALS	-0.06	
Knowledge up to datedness	ALS	.13	
Practice Leadership	ALS	-0.03	
Winning Business	ALS	.06	
Product Service Knowledge	ALS	-0.12	

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

fifth factor. Winning business, business acumen, client orientation and negotiation were loaded on the sixth factor, while resilience was loaded on the eighth factor. Knowledge currency was loaded on the ninth factor and product and service on the tenth factor. Personal integrity and creativity and innovation were loaded on the fourteenth factor, while respect for diversity was loaded on the fifteenth factor. Personal impact was loaded on the thirteenth factor and practice leadership on the sixteenth factor.

To investigate whether the number of dimensions measured in the LADC can be meaningfully collapsed into a smaller number of dimensions, a composite score on every dimension across exercises was calculated for every individual.

The inter-correlation matrix of the composite dimension ratings were subjected to PCA, followed by a varimax rotation. PCA revealed the presence of a dual factor pattern structure (Table 5). Similar to the study by Greyling et al. (2003), the resulting factor structure pattern in this study indicate that the LADC competency dimensions are not necessarily indicative of completely separate concepts and only two factors may be required to explain the variations in assessor ratings.

DISCUSSION

The findings of the present study showed evidence of discriminant validity only for some dimensions measured

Table 5. Factor matrix of composite scores on 21 dimensions.

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

in the PSYCH and some dimensions measured in the PON. Due to validation studies conducted a priori in the development of psychometric tests, evidence of discriminant validity with the PSYCH intercorrelations was not surprising.

The small number of competencies in the SPS and EXEC exercises combined with their logical underlying similarity (e.g. personal impact and oral communication) was furthermore not efficient in distinguishing behaviour. However, despite employing the AER approach no discriminant validity could be established amongst the competency dimensions within the AC simulation and interview exercises namely SPS, EXEC, ALS and CBI. The pattern of factor analytic results for the SPS, EXEC, ALS and CBI further confirmed poor discriminant validity as all competency dimensions in each of these exercises clustered around a single factor with the exception of ALS dimensions loading on two factors. Very little evidence of convergent validity was apparent in the correlations of the same dimensions across different exercises as only eight of the 43 inter-correlations were found to be statistically significant, of which seven included the correlation of PSYCHE competency dimensions with similar dimensions in the CBI and PON. At best assessors were able to differentiate between competency dimensions measured in the PON, where the factor pattern structure elicited 10 factors from the 21 dimensions included in the PON.

The results of the study hold implications for improving construct validity of the LADC through improved AC design. Construct validity of ACs depends primarily on the accuracy of ratings and thus assessors' ability to process complex cognitive information inherent to the AC process (Kolk et al., 2004; Melchers et al., 2010; Moses, 2008; Thornton and Krause, 2009). In the LADC the assessors' cognitive load was affected by the large number of competency dimensions. Assessors struggled to differentiate between the small number of competency dimensions (three) within the SPS and the EXEC as they struggled with differentiating the seven competencies in the ALS and the nine in the CBI. It is therefore suggested that distinguishing the number of competencies within a particular exercise are affected by the total number of competencies in an AC. Collapsing competencies per exercise may thus not be adequate, but less competencies within the AC as a whole may affect construct validity in a positive way.

The small number of competencies in the SPS and EXEC exercises combined with their logical underlying similarity (for example, *personal impact* and *oral communication*) were furthermore not efficient in distinguishing behaviour. Kolk et al. (2004) recommends three to five competencies per exercise. Competencies assessed in a particular exercise should then however be more distinct to ensure that the AC remains competency based rather than become exercise focussed involving

redundant and inefficient assessment of similar competencies. In this regard Lievens (2008) makes valuable recommendations for exercise design when developing an AC. Inter-correlations between dimensions within exercise as well as the PCA of dimensions within each exercise, further suggest possible overlap in assessors' understanding of the meanings of the various LADC competency definitions. There is supporting evidence in literature of improved construct validity findings when dimensions were grouped into conceptually related and distinct dimensions (Arthur and Woehr, 2003; Damitz et al., 2003; Lievens and Conway, 2001; Shore et al., 1990).

In agreement with Kolk et al. (2004) it is worthwhile to note that AC dimensions are never completely orthogonal and therefore less discriminant validity may naturally be expected. In the ALS simulation for example it is expected that similar behaviour would relate to the *winning business* and *business acumen* competency dimensions. Clustering AC constructs *a priori* into distinct categories that are conceptually logical such as the triadic cluster of Kolk et al. (2004) or the great eight criterion dimensions of Bartram (2005) may address the issues of assessors' halo effect as well as the orthogonal nature of behavioural competencies and as such address AC construct validity (Jones and Born, 2008).

The transparency of dimensions to candidates in the present study is unknown. Kleinmann and Koller (1997) and Kleinmann et al. (2011) found that candidates' ability to identify evaluation criteria in an AC improves their criterion related performance in the AC, and by implication the overall validity of the AC. Consequently, to improve discriminant validity in the LADC it is recommended that dimensions are made transparent to all participants.

In conclusion, the findings of the present study mostly support the AC construct validity dilemma in that poor discriminate and construct validity was evident in the simulation and interview exercises included in the LADC under study. The findings concur with that of Arthur et al. (2008) as well as the Greyling et al. (2003) study in that assessor ratings seem to cluster according to exercise factors rather than dimension factors (that is, competencies). The results may seem to align with Lance's (2008) consequent dismissal of applying a dimension or competency-based approach in favour of an exercise based AC approach. However for similar reasons as proposed by Howard (2008), Melchers and Konig (2008), Arthur et al. (2008), Rupp et al. (2008), Lievens (2008), Moses (2008) and Jones and Born (2008) we do not recommend the redesign of ACs based on an exercise focussed orientation. Such an approach would inefficiently forfeit the value derived from the CBA approach to assessing in multi-cultural and global employment contexts. An exercise based approach would assess *being competent* in performing a particular task rather than the human qualities or the

competency/ies necessary to function effectively in a managerial work context.

Whilst AC validity research highlights potential problems and provides invaluable guidance with regard to AC design, practice and policy, varying results show that every newly designed AC's utility still needs to be established or approximated at least. Due to the fact that ACs are not equivalent representations of one another (Moses, 2008), but unique to each assessment context and purpose, AC construct definitions should be validated *a priori* the operationalisation of a new AC. As such the value of this study lies in the recommendations for the redesign of the LADC specifically, for AC design in general and it provides additional data in the AC debate and for further intended meta-analyses in the field.

REFERENCES

- Amen U (2010). Assessment centre as an effective tool to select the potential candidate for future management needs of an organisation. *Interdis. J. Contemp. Res. Bus.* 2(2):134-143.
- analysis of assessment centre validity. *J. Appl. Psychol.* 72(3):493-51.
- Arthur W, Day EA, Woehr DJ (2008). Mend it, don't end it; an alternative view of assessment centre construct-related validity evidence. *Ind. Organ. Psychol.* 1(1):105-111.
- Arthur W, Woehr DJ (2003). The construct-related validity of assessment centre ratings: A review and meta-analysis of methodological factors. *J. Manag.* 29(2):231-258.
- Arthur W, Woehr DJ, Maldegen R (2000). Convergent and discrimination validity of assessment centre dimensions: A conceptual and empirical re-examination of the assessment centre construct-related validity paradox. *J. Manag.* 26(4):813-835
- Bartram D (2004). Assessment in organisations. *Appl. Psychol. Int. Rev.* 53(2):237-259.
- Bartram D (2005). The great eight competencies: A criterion-centric approach to validation. *J. Appl. Psychol.* 90(6):1185-1203.
- Campbell DT, Fiske DW (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.* 56(2):81-105.
- Damitz M, Manzey D, Kleinmann M, Severin K (2003). Assessment center for pilot selection: construct and criterion validity and the impact of assessor type. *Appl. Psychol. Int. Rev.* 52(2):193-212.
- Dean MA, Roth PL, Bobko P (2008). Ethnic and gender subgroup differences in assessment centre ratings: A meta analysis. *J. Appl. Psychol.* 93(3):685-691.
- Eurich TL, Krause DE, Cigularov K, Thornton GC (2009). Assessment Centers: Current Practices in the United States. *J. Bus. Psychol.* 24(4):387-407.
- Fleener JW (1996). Constructs and developmental assessment centres: Further troubling empirical findings. *J. Bus. Psychol.* 10(3):319-334.
- Foxcroft C, Roodt G (2001). An introduction of psychological assessment in the South African context. Cape Town: Oxford University Press.
- Gaugler BB, Rosenthal DB, Thornton GC, Bentson C (1987). Meta-Greyling L, Visser D, Fourie L (2003). Construct validity of competency dimensions in a team leader assessment centre. *SA J. Ind. Psychol.* 29(2):10-19.
- Haaland S, Christiansen ND (2002). Implications of Trait-Activation Theory for evaluating the construct validity of assessment centre ratings. *Pers. Psychol.* 55(1):137-163.
- Harel GH, Arditi-Vogel A, Janz T (2003). Comparing the validity and utility of behavior descriptions interview versus assessment centre ratings. *J. Manage. Psychol.* 18(2):94-104.
- Howard A (2008). Commentary: Making assessment centres work the way they are supposed to. *J. Ind. Organ. Psychol.* 1(1):98-104.
- Jansen WP, Stoop BAM (2001). The dynamics of assessment centre

- validity: results of a 7-year study. *J. Appl. Psychol.* 86(4):741-753.
- Joiner DA (2000). Guidelines for ethical consideration for assessment centre operations: International Task Force on Assessment Centre Guidelines. *Public Pers. Manag.* 29(3):315-331.
- Jones RG, Born M (2008). Assessor constructs in use as the missing component in validation of assessment center dimensions: a critique and directions for research. *Int. J. Select. Assess.* 16(3):229-238.
- Kleinmann M, Ingold PV, Lievens F, Jansen A, Melchers KG, Konig CJ (2011). A different look at why selection procedures work: the role of candidates' ability to identify criteria. *Organis. Psychol. Rev.* 1(2):128-146.
- Kleinmann M, Koller O (1997). Construct validity of assessment centres: appropriate use of confirmatory factor analysis and suitable construction principles. *J. Soc. Beh. Person.* 12(5):65-84.
- Kolk NJ, Born M, Van Der Flier H (2004). A triadic approach to the construct validity of the assessment centre. *Eur. J. Psychol. Assess.* 20(3):146-156.
- Lance CE (2008). Why assessment centres do not work the way they are supposed to. *Ind. Organis. Psychol.* 1(1):84-97.
- Lievens F (2008). What does exercise-based assessment really mean? *Ind. Organis. Psychol.* 1(1):112-115.
- Lievens F, Conway JM (2001). Dimension and exercise variance in assessment centre scores: A large-scale evaluation of multitrait-multimethod studies. *J. Appl. Psychol.* 86(6):1202-1222.
- Melchers KG, Kleinmann M, Prinz MA (2010). Do assessors have too much on their plates? The effects of simultaneously rating multiple assessment centre candidates on rating quality. *Int. J. Select. Assess.* 1(1):125-127.
- Melchers KG, Konig CJ (2008). It is not yet time to dismiss dimensions in assessment centres. *Ind. Organis. Psychol.* 1(1):125-127.
- Meriac JP, Hoffman BJ, Woehr DJ, Fleisher MS (2008). Further evidence of validity of assessment centre dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *J. Appl. Psychol.* 93(9):1042-1052.
- Moses J (2008). Assessment centres work, but for different reasons. *Ind. Organis. Psychol.* 1(1):134-136.
- Potgieter TE, Van der Merwe RP (2002). Assessment in the workplace. A competency-based approach. *SA J. Ind. Psychol.* 28(1):60-66.
- Prinsloo M (2000). Cognitive process manual. Unpublished manuscript. Johannesburg: Magellan Consulting.
- Rupp DE, Thornton GC, Gibbons AM (2008). The construct validity of the assessment centre method and usefulness of dimension as focal constructs. *Ind. Organis. Psychol.* 1(1):116-120.
- Saunders E (2000). Assessing human competence: Practical guidelines for South African managers. Johannesburg: Knowledge Resources.
- Schlebusch S, Roodt G (2008). Assessment centres: Unlocking potential for growth. Randburg: Knowres Publishing.
- SHL (2000). Reliability study: occupational personality questionnaire 32i. Study no R030. [Internet: <http://www.shl.research.co.za>; downloaded on 2012-07-30.]
- SHL (2001). Reliability study: occupational personality questionnaire 32i. Study no R036. [Internet: <http://www.shl.research.co.za>; downloaded on 2012-07-30.]
- Shore TH, Thornton GC, Shore L (1990). Construct validity of two categories of assessment centre dimension ratings. *Pers. Psychol.* 43(1):101-116.
- Taylor TR, Boeyens JCA (1991). The comparability of scores of blacks and whites on the South African Personality Questionnaire: An exploratory study. *SA J. Psychol.* 21(1):1-11.
- Thornton GC, Krause DE (2009). Selection versus development assessment centers: an international survey of design, execution and evaluation. *Int. J. Hum. Res. Manage.* 20(2):478-498.
- Thornton GC, Rupp DE (2006). *Assessment Centres in Human Resource Manage. Strategies for prediction, diagnosis and development.* New Jersey: Lawrence Erlbaum Associates, Publishers.