

# CONSTRUCT VALIDITY OF COMPETENCY DIMENSIONS IN A TEAM LEADER ASSESSMENT CENTRE

LEE-ANN GREYLING

DELÉNE VISSER

LINDA FOURIE

*Programme in Industrial Psychology  
Department of Human Resource Management  
Rand Afrikaans University*

## ABSTRACT

The aim of the study was to examine the construct validity of an assessment centre. The sample included 138 individuals who participated in a one-day call centre team leader assessment centre. Nine competency dimensions were rated using six exercises. Correlations and a principle axis factor analysis were utilised to study the convergent and discriminant validity of the dimension ratings. The results showed that the ratings clustered according to exercises rather than to dimensions (traits), thereby indicating a substantial amount of method variance. A further factor analysis of the nine competency dimensions yielded two factors that were named interpersonal and problem solving. Implications for the design of assessment centres are discussed.

## OPSOMMING

Die doel van die studie was om die konstruegeldigheid van 'n takseersentrum te ondersoek. Die steekproef het uit 138 persone bestaan wat aan 'n eendagtakseersentrum vir oproepsentrumgroepleiers deelgeneem het. Ses oefeninge is gebruik om nege bevoegdheidsdimensies te meet. Korrelasies en 'n hoofasfaktorontleding is gebruik om die konvergente en diskriminante geldigheid van die dimensiebeoordelings te bestudeer. Die resultate het getoon dat die beoordelings volgens oefeninge eerder as dimensies (eienskappe) gegroepeer het. Hierdie bevinding dui op 'n hoë mate van metodevariansie. 'n Verdere faktorontleding van die nege bevoegdheidsdimensies het twee faktore, wat as interpersoonlik en probleemoplossend benoem is, opgelewer. Implikasies vir die ontwerp van takseersentrums word bespreek.

Organisations make use of a variety of procedures to assist in the recruitment and selection of employees. Schmidt and Hunter (1998) found that an important component of any selection procedure is the ability to predict job performance. Consequently, use of selection procedures with increased predictive validity lead to increased productivity and a return on investment on the cost of selection and placement (Schmidt & Hunter, 1998). Lievens and Conway (2001) acknowledged the importance of predictive selection procedures, but suggested that when selection procedures are linked to personal development, more attention needs to be placed on why selection procedures work and what underlying constructs they measure. By identifying underlying constructs of selection procedures, individual feedback becomes valid and therefore beneficial to the development of the individual. Common selection procedures include interviews, personality and ability testing, reference checks, curriculum vitae screening, single performance tests and assessment centres (Thornton, 1992).

Assessment centres in particular measure a set of performance related traits such as flexibility, problem solving or interpersonal sensitivity. The performance related traits are usually referred to as dimensions and the assessment centre objective is to use several exercises and several assessors in order to achieve the most comprehensive and clearest indication of a dimension (Kleinmann, Kuptsch & Koller, 1996; Robertson, Gratton & Sharpley, 1987; Woodruffe, 1998). The assessment centre procedure according to Lievens and Conway (2001) can therefore be regarded as a dimension-based model.

What makes an assessment centre unique is its multi-method multi-trait multi-rater approach (Theron & Roodt, 2001). Various selection exercises are used where multiple dimensions are observed and measured by more than one assessor (Campbell & Fiske, 1959; Robertson et al., 1987). The assessment centre exercises may consist of psychometric testing, interviews and work sample tests. Spenser and Spenser (1993) pointed out

that work sample tests simulate on-the-job behaviour. An example of a work sample test is a case study, where the participant is faced with a management problem such as employee time off, decisions about resource allocation and conflict among co-workers. A second example takes the form of a group exercise in which participants in a group are given one or more problems to solve that requires collaboration. Further work sample testing could include a role play. In a role play, a participant can be asked to play the role of a manager dealing with an irate customer or a poorly performing employee. Often a business production game may be used where a participant is given a role as a manager in a game requiring goal setting and efficient use of resources whilst under time constraints (Spenser & Spenser, 1993). Multiple trained assessors observe participants' behaviour during the exercises and judgments about observed behaviour are made (Shore, Shore & Thornton, 1992). Thornton (1992) stated that these judgments are then pooled in a meeting among the assessors or by a statistical integration process. In this discussion process or "wash up", comprehensive accounts of behaviour and ratings are gathered. As mentioned, each dimension is observed in more than one exercise; therefore a participant has several opportunities to demonstrate capability in the dimension being measured. As reported by Boulter, Dalziel and Hill (1996), a dimension that is measured in more than one exercise also assists with ensuring that a rounded view of each participant is obtained, as the assessors are also rotated resulting in each assessor observing each participant at least once.

Assessment centres are growing in popularity and are widely used in many small and large organisations for selection, placement, succession planning, development and training of managers (Gaugler, Rosenthal, Thornton & Bentson, 1987; Robertson et al., 1987; Spenser & Spenser, 1993). Jansen and De Jongh (1997) suggested a possible reason for the increase in the use of assessment centres with particular relevance to South Africa. This reason centres on the fact that techniques used in selection and promotion procedures should be objective and not discriminatory. Assessment centres are focused on performance related traits or dimensions and not specific skills, which may be

affected by past opportunities. Due to the increased use of assessment centres, pressures to regulate and modify the use of assessment centres have been growing. Firstly, people have questioned whether the benefits outweigh the costs, especially in comparison with less costly selection procedures (Thornton, 1992). The typical assessment centre will take up to six assessors and five participants away from their jobs. In addition there is a need for administrative back up and in many cases the hire of suitable venues. Secondly, many theoretical arguments have been put forward stating that assessment centres do not work and that the procedure itself should be modified (Promotional Assessment Skills Service, 2001). In addition, Thornton (1992) suggested that a further criticism has arisen due to the use of different types of procedures for observing, reporting and combining behavioural observations.

Due to the increase in the use and the consequent need for standardisation of the assessment centre procedure, researchers have sought to examine the validity of this procedure, with a large portion of research focusing on the predictive aspect. Predictive validity may be said to measure the degree to which a selection procedure correctly predicts the relevant criterion (Huysamen, 1996). When applying the above definition to assessment centre procedure, one may consider whether the procedure predicts actual performance on the job and also whether it predicts the potential to do the job. Schmidt and Hunter (1998) suggested that there are many instances in which participants who received high assessment centre ratings proved to be successful at their jobs, because assessment centres scores do appear to predict acquisition of related knowledge. Furthermore, Turnage and Muchinsky (1984) not only found assessment centres to be highly correlated with ratings of potential, but also with progress in management level, salary and performance ratings. Further support for the predictive validity of assessment centres was found in a meta-analytic study conducted by Gaugler et al. (1987). The researchers conducted a meta-analysis of 50 assessment centre studies and found the mean validity coefficient for ratings of management potential to be statistically significant ( $r = 0,53$ ). The validity coefficient for ratings of performance was also significant ( $r = 0,36$ ), which proved to be lower than the validity of an intelligence test, but higher than the most common procedure used to select, namely the unstructured interview ( $r = 0,25$ ). In more recent research, Jansen and Stoop (2001) also found evidence to support the predictive validity of assessment centres. Kleinmann et al. (1996) suggested that the reason for predictive evidence is that assessment centres allow for an accurate evaluation of a participant's ability, and it is this ability that will decide whether a person will be a competent manager. It can thus be surmised that research has found that assessment centres do have predictive validity support and that assessment centres may be considered as one of the most predictive procedures used to select and develop employees in industry today (Gaugler et al., 1987; Jansen & Stoop, 2001; Kleinmann et al., 1996; Spenser & Spenser, 1993; Turnage & Muchinsky, 1984).

In an assessment centre, assessors derive performance ratings for each participant. These ratings are based on a set of traits, which are referred to as dimensions and these dimensions result from an analysis of the relevant job (Robertson et al., 1987). Assessment centres that are designed to assess management strengths and weaknesses rely on dimensions, common to both test performance and job behaviour (Shore, Shore & Thornton, 1990). Furthermore, Lievens and Conway (2001) suggested that feedback and development plans derived from assessment centres that do not adequately measure the dimensions, could prove to be invalid and even detrimental to the participant. Although predictive validity is important, further research needs to examine why assessment centres are predictively valid and whether the dimensions are actually being measured, hence construct validity is explored (Anastasi & Urbina, 1997).

Campbell and Fiske (1959) suggested that a way to determine the validity of a construct is to employ multiple measurement instruments that assessed multiple traits. This method became known as the multi-trait multi-method matrix (MTMM). The MTMM occurs when a set of traits is measured by a number of methods. The results are presented in a correlation matrix called the MTMM matrix (Campbell & Fiske, 1959). When strong correlations occur between two methods measuring the same trait, convergent validity is demonstrated. Discriminant validity is confirmed by weak correlations between two different traits measured by the same method (Campbell & Fiske, 1959; Kerlinger & Lee, 2000). When relating the MTMM definition to the assessment centre procedure a dimension may be seen as the trait and the method may be referred to as the exercise in which the dimensions are measured. Thus one is able to get insight into what an assessment centre measures by looking at the relationships among several dimensions measured in several exercises (Thornton, 1992). Thornton (1992) declared that high correlations between the ratings for the same dimension observed in two or more exercises would indicate evidence for convergent validity. Discriminant validity in assessment centres may be found when different dimensions measured in the same exercise yield low correlations.

Although evidence for convergent validity has been found, discriminant validity for dimensions measured in the same exercise is lacking (Shore et al., 1990). Gaugler and Thornton (1989) suggested that method variance could be a reason for the lack of discriminant validity. According to Campbell and Fiske (1959) method variance is demonstrated by high correlations between different dimensions measured in the same exercise. Spector (1987) inferred that method variance was not a reason for lack of discriminant validity when he assessed the amount of method variance by comparing monomethod correlations (different dimensions measured in the same exercise) and heteromethod correlations (different dimensions measured across exercises). The results showed that the monomethod correlations did not significantly differ from the heteromethod correlations and thus little evidence for method variance was found. In contrast with Spector's (1987) study more recent research has found evidence for method variance where ratings across dimensions measured in the same exercise correlated higher than ratings of a single dimension measured across exercises (Bagozzi & Yi, 1990; Kleinmann, 1993; Schneider & Schmitt, 1992; Spector, Schneider, Vance & Hezlett, 2000). Although the assessment centre procedure was thought of as a dimension based model, method variance may have prevented consistent ratings of the same dimension measured in two or more exercises. It follows then that there has been little support for a dimension based model with greater focus being placed on the ability of the exercise to replicate job behaviour, resulting in an exercise based model of assessment centre ratings (Lievens & Conway, 2001).

Woodruffe (1998) implied that a possible reason for the occurrence of method variance was due to an overall halo effect because assessors do not distinguish between dimensions measured in the same exercise. Thornton (1992) pointed out that the method of observing and rating behaviour in an assessment centre influences the pattern of ratings. Support for a halo effect comes from processes that may affect assessors' ratings such as social desirability, personality factors, the actor-observer effect and even physical attractiveness (Facteau & Craig, 2001; Huysamen, 1996). Biases stemming from observing and rating behaviour then result in artificially raising correlations among dimension ratings measured in the same exercise, causing lack of discriminant validity (Bagozzi & Yi, 1990; Lindell & Whitney, 2001). When exploring the rating method utilised in assessment centres, two different rating techniques occur. In the *within dimension method*, final dimension ratings are made by a consensus discussion of the exercise data by all the assessors. In the *within exercise method*, ratings are made by assessors after each exercise and final ratings are then made by either a "wash-up" session or a statistical

computation of the dimension scores (Spector et al., 2000). According to Thornton (1992), the halo effect tendency may be caused by asking assessors to determine dimension ratings after a single exercise. In contrast, the within dimension rating method makes use of dimension ratings based on ratings across several exercises. The within dimension rating method appears to be more consistent and accurate than the within exercise rating method (Thornton, 1992). Gaugler and Thornton (1989) proposed limiting the number of dimensions and stated that method variance could possibly be reduced and convergent validity improved. By this, assessors who deal with a few dimensions may be able to make more accurate observations and subsequent ratings, than assessors who deal with a greater number of dimensions. A further statistical procedure to test for construct validity is a factor analysis of the MTMM intercorrelations (Campbell & Fiske, 1959). Thornton (1992) mentioned that dimension factors consist of the ratings on an individual dimension measured in two or more exercises and exercise factors consist of ratings for different dimensions measured within a single exercise. When looking at the dimension ratings, Shore et al. (1990) showed that the dimension ratings have generally yielded two to four factors. Spector et al. (2000) agreed and stated that these factors are usually clustered as problem solving and interpersonal dimensions.

The above discussion has yielded a number of issues regarding the construct validity of assessment centres. In summary, if assessment centres are to be linked to development, they must be seen as dimension based models of selection and development (Lievens & Conway, 2001). Whereas current research has found evidence for convergent validity, divergent validity appears to be lacking with method variance suggested as a possible reason (Campbell & Fiske, 1959; Thornton, 1992). The occurrence of method variance has led to the assessment centre procedure being considered as a model based on the ability of the exercises to replicate the job function, shifting focus away from the underlying dimensions needed for individual development. Reasons for method variance resulting in overall lack of construct validity of dimension ratings have been discussed. Firstly, method variance may account for lack of discriminant validity in that assessors' observations and subsequent ratings may not be consistent across exercises (Facteau & Craig, 2001; Huysamen, 1996). Secondly, Thornton (1992) suggested that the rating process itself might be a source of error. Thirdly, Gaugler and Thornton (1989) asserted that the number of dimensions might overload assessors and collapse the dimensions that they are measuring. Wilson and Walwanis (2000) believed that dimension definition and exercise design should also be considered. By dimension definition is meant that the dimension to be measured may not have been accurately defined and so the meaning may differ from exercise to exercise. Exercise design can also be seen as a possible cause of low construct validity, because the opportunity to demonstrate a dimension in an exercise may vary from exercise to exercise. Although discriminant validity of dimension ratings appears to be lacking, construct validity for broad categories of final dimension ratings has been found because the final dimension ratings can be clustered into broad groupings of dimensions such as problem solving and interpersonal dimensions.

Assessment centres are designed with the expectation that there will be more agreement across exercises per dimension than between different dimensions measured in the same exercise. In other words, one would expect low correlations between different dimensions measured in the same exercise and high correlations between exercises rating the same dimension. Accordingly, if construct validity did exist, one would also expect that the dimension ratings would be clustered according to the dimension factors and not the exercise factors (Thornton, 1992). It follows then that one would also expect similar dimensions to cluster together thereby providing construct validity evidence of the scores.

The present study therefore aimed to answer the following questions:

- Are individual dimensions highly correlated across exercises in the assessment centre procedure? That is, are the

dimensions measured consistently, irrespective of the measurement method?

- Do exercises in the assessment centre procedure permit adequate discrimination between the dimensions measured? Is there a lack of method variance such that weak correlations occur between different dimension ratings measured in the same exercise (Campbell & Fiske, 1959)?
- Can the dimensions be meaningfully clustered into a smaller number of dimensions? For example, can the dimensions be clustered into problem solving and interpersonal dimensions?

## METHOD

The present study explored the relations of performance on six separate assessment centre exercises. The exercises were diverse and included a case study, leaderless group discussion, role play, structured interview, verbal ability assessment and numerical interpretation assessment.

### Participants

Assessment ratings were obtained on 138 participants who took part in a one-day call centre team leader assessment centre. The assessment centres spanned an 11-month period. The assessment centre was utilised for the development and selection of call centre team leaders in a medical insurance organisation situated in Gauteng. Twenty-one percent of the participants attended for development purposes and 79 percent for selection purposes. The ages of the delegates ranged from 20 to 50 years ( $M = 27,5$  and  $SD = 4,59$ ). Their race and gender breakdown is summarised in Table 1.

**TABLE 1**  
**RACE AND GENDER BREAKDOWN OF THE ASSESSMENT**  
**CENTRE PARTICIPANTS (N=138)**

	Asian	African	Coloured	White	Total
Male	8	6	4	30	48
Female	14	19	23	34	90
Total	22	25	27	64	138

### Assessment centre methodology

Assessment centre methodology consisted of the competency dimensions, the exercises in which the dimensions were measured and the training of assessors.

#### Dimensions

The assessment centre was based on nine dimensions. Three insurance organisations took part in a job profiling exercise to establish the dimensions that were specific to call centres. Similarities were found between all three organisations and the outcome was a call centre framework. The call centre framework listed nine dimensions critical to all roles in a call centre environment. Behavioural anchors were then assigned to the dimensions. The behavioural anchors allowed for performance on a particular dimension to be rated according to a five-point scale. The dimensions that were identified to be crucial to the call centre team role are described below (Riley, Ric-Hansen & Rushmere, 2000).

1. *Analytical thinking and decision making.* This is defined as gathering relevant information and analysing issues while breaking them down into their component parts. Analytical thinking and decision making involves creating systematic and rational judgements based on relevant information. It is also about identifying cause and effect relationships to solve problems. At the highest level it is about making judgements, even when all the information is not available.

2. *Business and commercial awareness.* This dimension refers to being aware of the competitive context and market trends in

which organisations operate. It is about viewing issues in terms of efficiency and effectiveness in the framework of costs, profits, markets and value.

3. *Forward thinking.* Forward thinking is defined as looking ahead to anticipate, prioritise and plan. It differs from strategic thinking in that it is operational and not meant to be visionary. It is about the tactics required to deal with the immediate future.

4. *Influencing and persuading.* This dimension is about using appropriate communication styles and methods to influence, convince and impress others in a way that results in acceptance, agreement or behaviour change leading to win-win relationships. It is about winning the support of others.

5. *Motivating others.* This is defined as maximising the contribution of groups and individuals by inspiring real energy, enthusiasm and effort towards the organisation's values. It involves choosing to invest time and effort; fostering an open and supportive environment; developing an understanding of the vision; motivating people and gaining the commitment of groups and individuals to the challenges ahead.

6. *Customer focus.* This refers to putting the customer first and being eager to provide service that exceeds expectations. It is about working to meet and exceed customer needs while at the same time looking after the customer's interests.

7. *Developing others.* This dimension refers to providing practical support to enable employees to develop improved performance and build capability for the future. It is also about creating a culture in which people take responsibility for their own learning and career development. It is about building organisational capability now and for the future.

8. *Driving results through others.* This can be described as maximising performance outcomes by monitoring and managing the efforts of others to achieve deliberately stretching goals.

9. *Self control.* Self control is defined as maintaining effective work behaviour in the face of setbacks or pressure, remaining calm, stable and in control in the face of adversity.

#### Assessment centre exercises

Six assessment exercises were designed to measure the nine dimensions. The exercises are explained in detail below.

1. *Case study.* The participants assumed the role of a call centre team leader, faced with the rostering of agents, client demands, problems and tasks similar to those encountered on the job. Participants were briefed by their manager (played by an assessor) and were allotted with limited time to read through and respond to items that could appear in a team leader's inbox. The participants were then required

to report back to the call centre manager in a debriefing session on how they were going to tackle the day's activities.

2. *Leaderless group discussion.* In this exercise, participants worked as a group of peers to solve a problem that call centre team leaders might encounter. At the start of the meeting, each participant was given a description of the problem and was asked to write down their own independent solutions. This part of the leaderless group discussion was not observed or rated by the assessors. After some time had lapsed, the individual participants were asked to present a convincing argument in a group meeting that supported their individual findings. The participants were required, as a group, to reach consensus concerning the best solution. Although participants initially showed some ownership, the discussion tended to be a cooperative exercise aimed at finding the best solution. Each participant was given an individual rating.

3. *Role play.* In this exercise, participants assumed the role of a call centre team leader, with an assessor playing the role of a subordinate. The participants were required to deliver negative feedback while at the same time gaining commitment and preventing the subordinate from resigning.

4. *Structured interview.* Participants were put through a structured interview that focused on work history and past performance. The assessor was required to look for specific behavioural incidents that allowed judgements to be made.

The ability measures used were well established and well researched ability assessments developed by Saville and Holdsworth (SHL). Participants were asked to complete two assessments from SHL's Critical Reasoning Test Battery.

5. *Verbal reasoning ability test (VC1.1).* This is a test of verbal evaluation and measures the ability to understand and evaluate the logic of various kinds of arguments. It includes a variety of topics that are relevant to junior management grades.

6. *Numerical interpretation ability test (NC2.1).* This test measures the ability to interpret data and in doing so, tests the ability to make correct decisions for numerical data. Straightforward statistical information and other numerical data are presented. This test is deemed to be appropriate for any job that involves analysis or decision-making based on numerical facts. (The results from the above two tests each provided a separate rating on only one dimension, which was analytical thinking). The different assessment centre exercises were intended to measure different parts of the content domain of the job independently. Each dimension was measured in at least two different exercises as Table 2 illustrates. This was done in order to give participants the opportunity to demonstrate the dimension in more than one situation. It also ensured that the dimension ratings were as consistent and objective as possible, as numerous assessors rated the same participant.

**TABLE 2**  
**DIMENSIONS MEASURED IN ASSESSMENT EXERCISES**

	Structured Interview	Group Exercise	Role Play	Case Study	Verbal Ability VC1.1	Numerical Ability NC2.1
Analytical thinking				X	X	X
Business and commercial awareness		X		X		
Forward thinking	X			X		
Influencing and persuading		X	X	X		
Motivating others		X	X			
Customer focus	X			X		
Developing others	X		X			
Driving results through others	X		X			
Self control	X	X	X	X		

### Assessors

The assessors consisted of call centre managers and HR practitioners. They participated in a one-day training workshop, which focused on general assessment skills of observing, recording, categorising and evaluating observed behaviour. All assessors were also taught how to avoid rating errors. The trainer instructed the assessors to make behavioural descriptions of the participants' behaviour. Next, assessors were tasked with defining the dimensions and then classifying behaviours per dimension. For example, the dimension influencing and persuading was defined as the use of appropriate communication styles and methods (written, oral, face-to-face, remote, group or individual) to influence, convince and impress others in a way that results in acceptance, agreement or behaviour change leading to win-win relationships. It is about winning the support of others (Riley et al., 2000). Furthermore, the use of the chosen dimensions was explained. Focus was placed on why the dimensions were important for the role of a call centre team leader and how the dimensions were aligned to the organisation's values. The dimension influencing and persuading was deemed important because, in entrepreneurial cultures in which an owner-manager focus is valued and where participative non-hierarchical decision-making is encouraged, gaining the buy-in and commitment for ideas and actions is critical. Influencing capability in this environment is a key differentiator for success in getting ideas accepted (Riley et al., 2000). The final part of the training included the ratings of the dimensions according to the behaviour observed. The trainer then elicited a discussion of which behaviours were used to decide on an assigned rating, clarifying any discrepancies among ratings. Finally, the trainer provided the assessors with feedback pertaining to assessor ratings (Lievens, 2001).

### Procedure

The participants in the assessment centre were employees who attended the assessment centre for either selection or development purposes. In each exercise assessors were required to observe and record behaviour exhibited by the participants. Thereafter, the assessors were tasked with having to match the observed behaviours with the behavioural anchors associated with the dimensions being measured. The behavioural anchors categorised the dimension being measured on a five-point rating scale, which were: low behaviours, moderate-low behaviours, moderate behaviours, moderate-high behaviours and high behaviours. After every exercise the assessor decided into which of the five categories the observed behaviour fell. Final ratings were then made by the group of assessors based on the combined ratings of all dimensions at the end of each assessment centre, thus utilising the *within dimension* rating method (Thornton, 1992). An administrator was also present at the assessment centres to facilitate and co-ordinate the process. In order to reduce the possibility of assessor bias occurring, a multi-rater approach was utilised, because the assessors were rotated to ensure that each assessor observed a particular participant in only one exercise during the day.

An assessor gave feedback to all participants, in the presence of their managers. A full developmental report was produced and this then served as a basis for further development.

## RESULTS

Descriptive statistics for the dimensions using composite mean scores are presented in Table 3. The composite scores were obtained by adding the scores making up each dimension and dividing the total by the number of exercises involved. The composite scores may then be interpreted on a five-point scale. The mean ratings for all of the dimensions are

close to 3,00, which is the middle of the five-point scale. Performance on the dimensions spanned almost the entire range for all nine dimensions.

**TABLE 3**  
**DESCRIPTIVE STATISTICS FOR THE DIMENSIONS INCLUDING**  
**COMPOSITE MEANS SCORES (N=138)**

Dimensions	M	SD	Skewness	Kurtosis
Analytical thinking	2,40	1,00	0,49	-0,79
Business and commercial awareness	2,63	0,87	0,22	-0,35
Forward thinking	2,90	0,83	0,13	-0,36
Influencing and persuading	2,96	0,79	-0,01	-0,83
Motivating others	3,03	0,81	-0,09	-0,38
Customer focus	2,88	0,80	0,39	-0,33
Developing others	2,98	0,78	-0,02	-0,02
Driving results through others	2,84	0,81	-0,10	-0,34
Self control	3,41	0,73	-0,12	-0,49

The first part of the study questioned the consistency or convergence of dimension ratings across exercise methods. Nonparametric statistical tests were used to calculate correlations in the present study. This was due to the ranking scale used in the rating of participants. Kendall's tau was used to calculate intercorrelations between assessment exercises for every dimension separately. Table 4 presents the intercorrelations between the exercises used to measure the various dimensions.

The intercorrelations between the ability tests and the case study for the dimension *analytical thinking* were all statistically significant with the highest correlation occurring between the case study and the numerical interpretation test ( $r = 0,47$ ). A moderate correlation occurred between the case study and the group exercise for the *business and commercial awareness* dimension. The correlation ( $r = 0,05$ ) for the *forward thinking* dimension between the interview and case study was not statistically significant. Moderate intercorrelations occurred between the role play, group exercise and case study for the *influencing and persuading* dimension. A moderately high correlation occurred between the role play and group exercise for the *motivating others* dimension. The *customer focus* dimension was rated in the interview and case study. The correlation was not statistically significant. A correlation between the interview and role play for the *developing others* dimension also produced results that were not statistically significant. The *driving results* dimension was measured in the interview and the role play. Once again the correlation was not statistically significant. The *self control* dimension was measured in four exercises, namely the interview, role play, group exercise and case study. The lowest correlation occurred between the role play and interview ( $r = 0,08$ ), whereas a statistically significant correlation occurred between the case study and the role play ( $r = 0,35$ ).

The second research question sought to examine whether individual exercises permitted adequate discrimination between the various dimensions. Kendall's tau was used to calculate intercorrelations between dimensions in every exercise. Table 5 shows the intercorrelations between the five dimensions rated by the *interview method*. All the intercorrelations were statistically significant, except for the correlation between self control and driving results through others ( $r = 0,13$ ).

**TABLE 4**  
**INTERCORRELATIONS BETWEEN EXERCISE METHODS FOR THE VARIOUS DIMENSIONS (P VALUES IN BRACKETS)**

Dimension	Exercise method				
		Case study	VC1.1	NC1.1	
Analytical thinking	Case study	1,00			
	VC1.1	0,28 (0,000)	1,00		
	NC2.1	0,47 (0,000)	0,35 (0,000)	1,00	
Business and commercial awareness	Group exercise	1,00			
	Case study	0,24 (0,001)	1,00		
Forward thinking	Interview	1,00			
	Case Study	0,05 (0,474)	1,00		
Influencing and Persuading	Role play	1,00			
	Group exercise	0,22 (0,002)	1,00		
	Case study	0,26 (0,000)	0,23 (0,001)	1,00	
Motivating others	Role play	1,00			
	Group exercise	0,54 (0,000)	1,00		
Customer focus	Interview	1,00			
	Case study	0,14 (0,058)	1,00		
Developing others	Interview	1,00			
	Role play	0,13 (0,075)	1,00		
Driving results through others	Interview	1,00			
	Role play	0,08 (0,247)	1,00		
Self control	Interview	1,00			
	Role play	0,08 (0,258)	1,00		
	Group exercise	0,19 (0,011)	0,22 (0,003)	1,00	
	Case study	0,10 (0,165)	0,35 (0,000)	0,24 (0,001)	1,00

The intercorrelations between the five dimensions assessed through the *role play* were moderate to high, ranging between 0,37 and 0,71. These intercorrelations are presented in Table 6.

For the *group exercise*, the intercorrelations between the four dimensions ranged from not being statistically significant to moderately high as presented in Table 7. The correlation between motivating others and influencing and persuading was 0,58.

Statistically significant intercorrelations ranging between 0,32 and 0,59 occurred between the five dimensions measured in the *case study exercise*.

A principal axis factor analysis, followed by a varimax rotation of the factor axes, was performed on the ratings to cluster dimensions together or to indicate possible method variance. Six factors were obtained that explained 67,01 percent of the variance. The sixth factor was not adequately determined because it consisted of less than three substantial loadings. The rotated factor matrix is presented in Table 9. All the *case study ratings* and *psychometric ratings* were loaded on the first factor. All the *role play ratings* loaded on the second factor, whereas the *group exercise ratings* loaded on the third factor. The *interview ratings* were grouped into

two factors, with self control, forward thinking and customer focus loading on the fourth factor and developing others and driving results through others loading on the fifth factor.

The pattern of factor analytic results confirmed the results of the intercorrelations presented in Tables 4 to 8, namely that the ratings clustered according to exercises rather than to dimensions (traits).

The third research question asked whether the dimensions could be grouped into a smaller number of dimensions on the basis of the intercorrelations between them. Another factor analysis was performed on the composite scores of the nine dimensions to determine whether the dimensions supported an underlying two-factor structure. A principal axis factor analysis followed by direct oblimin rotation of the factor axes was performed on the composite scores for each dimension and the results of the factor pattern matrix are shown in Table 10. Two factors explained 64,06 percent of the variance. The first factor encompassed the *interpersonal dimensions* including developing others, driving results through others, motivating others, influencing and persuading and self control. The second factor was made up of the *problem solving dimensions* and consisted of forward thinking, customer focus, analytical thinking and business and commercial awareness. The correlation between the two factors was equal to 0,55.

**TABLE 5**  
**INTERCORRELATIONS BETWEEN DIMENSIONS FOR THE INTERVIEW EXERCISE (P VALUES IN BRACKETS)**

	Forward thinking	Customer focus	Self control	Driving results through others	Developing others
Forward thinking	1,00				
Customer focus	0,38 (0,000)	1,00			
Self control	0,39 (0,000)	0,23 (0,000)	1,00		
Driving results through others	0,19 (0,010)	0,33 (0,000)	0,13 (0,065)	1,00	
Developing others	0,26 (0,000)	0,36 (0,000)	0,23 (0,002)	0,43 (0,000)	1,00

**TABLE 6**  
**INTERCORRELATIONS BETWEEN DIMENSIONS FOR THE ROLE PLAY EXERCISE (P VALUES IN BRACKETS)**

	Influencing and persuading	Motivating others	Developing others	Driving results through others	Self control
Influencing and persuading	1,00				
Motivating others	0,63 (0,000)	1,00			
Developing others	0,59 (0,000)	0,49 (0,000)	1,00		
Driving results through others	0,56 (0,000)	0,41(0,000)	0,71(0,000)	1,00	
Self control	0,58 (0,000)	0,59 (0,000)	0,45 (0,000)	0,37 (0,000)	1,00

**TABLE 7**  
**INTERCORRELATIONS BETWEEN DIMENSIONS FOR THE GROUP EXERCISE (P VALUES IN BRACKETS)**

	Influencing and persuading	Motivating others	Business and commercial awareness	Self control awareness
Influencing and persuading	1,00			
Motivating others	0,58 (0,000)	1,00		
Business and commercial awareness	0,32 (0,000)	0,23 (0,002)	1,00	
Self control	0,34 (0,000)	0,38 (0,000)	0,23 (0,001)	1,00

**TABLE 8**  
**INTERCORRELATIONS BETWEEN DIMENSIONS FOR THE CASE STUDY EXERCISE (P VALUES IN BRACKETS)**

	Business and commercial awareness	Customer focus	Forward thinking	Influencing and persuading	Self control
Business and commercial awareness	1,00				
Customer focus	0,37 (0,000)	1,00			
Forward thinking	0,43 (0,000)	0,50 (0,000)	1,00		
Influencing and persuading	0,48 (0,000)	0,36 (0,000)	0,49 (0,000)	1,00	
Self control	0,34 (0,000)	0,32 (0,000)	0,43 (0,000)	0,59 (0,000)	1,00

## DISCUSSION

The results of the present study found no evidence for discriminant validity between the dimensions, and very little convergent validity was demonstrated when individual dimensions were rated by means of different exercises. As also found by Thornton (1992), the intercorrelations between dimensions measured in the same exercise were greater than the

intercorrelations per dimension across various exercises, leading to ratings being clustered according to exercise factors and not dimension factors. Only two moderately high correlations occurred for the same dimension across exercises. These were analytical thinking ( $r = 0,47$ ) and motivating others ( $r = 0,54$ ), which did provide a degree of convergent validity support for these two dimensions.

**TABLE 9**  
**ROTATED FACTOR MATRIX OF DIMENSION RATINGS ON THE ASSESSMENT CENTRE EXERCISES**

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	h <sup>2</sup>
Case study: analytical thinking	0,84	0,12	0,09	0,05	0,21	-0,02	0,35
Case study: influencing and persuading	0,74	0,11	0,16	0,16	0,01	0,16	0,45
Case study: business and commercial awareness	0,73	0,04	0,11	0,06	0,08	-0,01	0,59
Case study: forward thinking	0,73	0,12	0,15	0,08	0,12	-0,11	0,44
Case study: self control	0,60	0,18	0,20	0,15	-0,01	0,25	0,57
Case study: customer focus	0,55	0,06	0,12	0,08	0,21	-0,01	0,79
Psychometrics: N.C.2.1 AC Rating	0,53	0,12	-0,01	0,06	-0,06	0,04	0,65
Psychometrics: V.C.1.1 AC Rating	0,43	0,02	-0,05	-0,19	-0,04	0,30	0,80
Role play: driving results through others	0,13	0,86	0,03	0,08	0,03	-0,26	0,83
Role play: developing others	0,12	0,84	0,25	0,05	-0,06	-0,08	0,74
Role play: influencing and persuading	0,14	0,80	0,13	0,08	0,20	0,25	0,80
Role play: motivating others	0,22	0,66	0,18	-0,05	0,22	0,30	0,55
Role play: self control	0,26	0,62	0,14	0,06	0,09	0,51	0,27
Group exercise: influence and persuasion	0,19	0,13	0,83	-0,15	0,15	-0,02	0,37
Group exercise: motivating others	0,10	0,17	0,71	0,03	0,07	0,06	0,74
Group exercise: self control	0,12	0,13	0,48	0,31	0,07	0,07	0,56
Group exercise: business and commercial awareness	0,23	0,12	0,34	0,02	0,28	-0,09	0,38
Interview: self control	-0,01	0,04	0,01	0,76	0,09	0,04	0,59
Interview: forward thinking	0,07	-0,01	-0,04	0,50	0,30	-0,09	0,63
Interview: customer focus	0,20	-0,03	0,09	0,49	0,40	-0,03	0,51
Interview: developing others	0,08	0,18	0,11	0,30	0,65	0,04	0,32
Interview: driving results through others	0,07	0,05	0,16	0,18	0,61	0,04	0,30

**TABLE 10**  
**FACTOR PATTERN MATRIX OF COMPOSITE SCORES ON THE NINE DIMENSIONS (N=138)**

	Factor 1	Factor 2	h <sup>2</sup>
Developing others	0,92	-0,12	0,40
Driving results through others	0,73	-0,04	0,42
Motivating others	0,69	0,04	0,56
Influencing and persuading	0,66	0,27	0,71
Self control	0,52	0,28	0,51
Forward thinking	-0,02	0,76	0,53
Customer focus	0,03	0,71	0,73
Analytical thinking	-0,02	0,64	0,49
Business and commercial awareness	0,05	0,62	0,51

As noted earlier, if assessment centres are to be used for development, dimensions need to be seen as the building blocks of assessment centres (Robertson et al., 1987). Dimensions are distinct performance related traits that have been profiled to the job in question in order to ensure the best fit (Robertson et al., 1987). Assessors are trained to make judgments based on observed behaviour of each individual dimension. The results of the current study do not support this view. The results reveal that the distinctions between individual dimensions measured in the same exercise are unclear. Moreover, a rating of one dimension in an exercise is related to ratings of different dimensions measured in the same exercise. Added to this, when a participant is rated on an individual dimension in more than one exercise, the ratings are not consistent.

Thornton (1992) suggested that assessor rating procedures might have influenced the pattern of results produced. The *within dimension* rating method involves final dimension ratings being based on an accumulation of evidence across a set of exercises. The *within exercise* rating method occurs when dimension ratings are made by individual assessors after each exercise. It follows then that within exercise dimension ratings are subject to rating errors and biases of a particular assessor when judging another's performance. In contrast, the within dimension rating

method is usually based on pooled observations of all assessors and is, therefore less likely to be biased, thereby increasing construct validity (Shore et al., 1990). The rating method used in the present study focused on the final dimension ratings and thus the possibility of rating errors and assessor biases occurring was reduced. It therefore seems unlikely that the rating procedure could be the main reason for the results obtained. Assessor training may also account for inconsistent dimension ratings (Gaugler et al., 1987). Although the assessors were all subjected to assessor training, functional background and assessment experience could influence the ratings.

Gaugler et al. (1987) stated that the presence of professional psychologists as assessors could be seen as a quality regulation to assessment centre validity. In the present study, trained psychologists were involved in the development and maintenance of this particular assessment centre. Lievens (2001) mentioned that assessor training should include the norms, values and personal qualities that an organisation considers to be crucial to sustain competitive advantage. Training in this way improves the discriminant validity (Lievens 2001). A strength of the present study is that the target job is very specific, being a team leader in a call centre environment thereby increasing the accuracy of a match to the specific job at a specific organisation (Spector et al., 2000). It follows then, that all assessors were in-house employees, with the trainer being an in-house psychologist. All dimensions used in this specific assessment centre were developed with the values and business drivers of the organisation in mind. The relevance of each of the dimensions to the organisation's goals and values were explained to the assessors. Consequently, assessor training as a reason for lack of construct validity could perhaps fall outside the scope of the current research.

Kleinmann et al. (1996) reasoned that another possible cause of poor construct validity might be the non-transparency of the dimensions. In a transparent condition, participants are expected to behave more consistently due to their knowledge of the behaviour requirements. This means that the ratings from assessors should be more consistent on individual dimensions measured in two or more exercises, which should result in



higher convergent validity. It can also be deduced that participants try to show the particular behaviour pattern more clearly, which means that dimensions assessed in one exercise can be differentiated, providing higher discriminant validity. However a possible downfall to this transparency as cited by Kleinmann et al. (1996) is that by divulging the dimensions to the participants, the assessment centre becomes unrealistic given that it might be used for everyday selection. Assessment centres are after all supposed to simulate everyday job relevant experiences, and the behaviours required to perform are not transparent. Future research might consider whether or not transparency of the dimensions in the assessment centre would still lead to predictive validity.

Individual dimension ratings may be affected by the demand characteristics of a particular assessment centre exercise, which may cause fluctuations in participant behaviour during the assessment centre day. For example, participants may feel more at ease in a group situation than having to debrief an assessor in a one-on-one environment as in the case study exercise. This variability in behaviour would tend to lower estimates of convergent validity (Shore et al., 1990).

As mentioned by Wilson and Walwanis (2000), the opportunity a participant has to demonstrate a dimension may vary from exercise to exercise which points to dimension definition as a possible cause. In the current study, dimension definition can be shown in that the dimensions driving results through others and developing others correlated substantively in both the interview ( $r = 0,43$ ) and the role play ( $r = 0,71$ ). The motivating others dimension and the influencing and persuading dimension were highly correlated in both the role play ( $r = 0,63$ ) and the group exercise ( $r = 0,58$ ). The above dimensions all relate to the interpersonal cluster and perhaps, on a further examination of the definitions and behavioural anchors of these dimensions, it might be found that the above dimensions are closely related to each other. It thus makes no sense to keep them as separate dimensions. This would then lead to fewer dimensions being measured and as suggested by Gaugler and Thornton (1989), would increase the possibility of greater construct validity of individual dimension ratings. In exploring dimension definition, Lievens (2001) suggested that management expectations are exercise specific and thus account for dimension ratings being factored as per the exercise. For instance, when assessors encounter behaviour in a role play exercise, this behaviour is matched against expectations regarding managerial behaviours when dealing with subordinate problems. The behaviours being associated with management expectations may be categorised in more than one dimension. As a result, relatively high correlations between different dimension ratings measured in the same exercise occur. In line with this, Gaugler and Thornton (1989) mentioned that this halo effect could be reduced by limiting the number of dimensions in an exercise that an assessor has to rate, as assessors have a limited capacity to process complex information. The greater the judgment task the more prone it will be to cognitive bias.

Despite overall lack of individual dimension construct validity, the present study did provide support for the construct validity of two categories of the final dimension ratings. The principal axis factor analysis of the composite scores supported the two major clusters of dimensions. This finding is consistent with previous studies, which suggests that assessors typically do not utilise more than a few dimensions in arriving at overall assessment ratings (Shore et al., 1990). Thus it appears that interpersonal and problem solving dimensions may well be a natural distinction in the minds of assessors.

Since organisations use the final dimension ratings for decision making as well as for development purposes, perhaps construct validation of the final dimension ratings is the more valuable approach regardless of how the assessment centre is structured. The most prudent explanation for the present study's results is

based on assessor bias and perhaps limitations in human information processing capabilities. Prior studies suggested that assessors have difficulty in making meaningful judgments when required to differentiate between large numbers of dimensions. For example, Gaugler and Thornton (1989) found that assessors classified behaviours more accurately into a smaller number of dimensions than into a large number of dimensions. However, it was concluded that the number of dimensions rated did not affect the discriminant validity in *within exercise* dimension ratings.

A limitation of prior research should also be acknowledged. The majority of the assessment centre construct validity studies did not include ratings from other sources besides the assessment centre (Shore et al., 1990). Further research is needed on the discriminant validity of individual dimension ratings in two or more exercises. It is these ratings that are used to make personnel decisions and therefore these ratings should be compared with ratings in the same dimension obtained from panel interviews, personality testing and performance appraisals (Gaugler & Thornton, 1989). Another limitation is the sample used in the present study. The sample was small and, although the size ensured accuracy and job specificity, a larger sample would be required to add greater generalisability of the results. Gaugler and Thornton (1989) also mentioned that other aspects of task complexity, such as specificity and observability of the dimensions, the number of exercises used and the number of different types of decisions and recommendations that assessors are asked to make, should also be looked at when trying to increase construct validity.

In summary, the present study has a number of important implications. Firstly, it builds on previous research by suggesting that final dimension ratings can be valid measures of underlying constructs. Concerns about the lack of construct validity of dimension ratings in a single exercise need to be addressed, as it is these ratings that assist in development. However, Jansen and Stoop (2001) stated that low construct validity of dimensions did not influence the predictive validity of assessment centres and indicated that it did not matter whether the overall results indicated a dimension based model or an exercise based model of ratings. Secondly, the study also builds on previous research conducted by Shore et al. (1990), indicating that assessor observations and subsequent judgements lead to a few broad categories such as interpersonal style and problem solving style. Organisations should consider providing assessors with a small number of dimensions to be rated as a way to improve the reliability and validity of assessor judgments. Where there is a lack of dimension construct validity, future assessment centre developers may need to create exercises that generate sufficient behavioural evidence in an exercise to measure a particular dimension, or consider dropping these dimensions to reduce any unnecessary cognitive demands on assessors.

## REFERENCES

- Anastasi, A. & Urbina, S (1997). *Psychological testing* (7<sup>th</sup> ed.). Upper Saddle River, NJ: Prentice-Hall.
- Bagozzi, R.P. & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self reported affect perceptions at work. *Journal of Applied Psychology*, 75 (5), 547-560.
- Boulter, N., Dalziel, M. & Hill, J (1996). *People and competencies*. London: Kogan.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Facteau, J.D. & Craig, S.B. (2001). Are performance appraisal ratings from different rating sources comparable? *Journal of Applied Psychology*, 86 (2), 215- 227.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G.C. & Bentson, C. (1987). Meta analysis of assessment centre validity. *Journal of Applied Psychology*, 72 (3), 493-511.

- Gaugler, B. B. & Thornton, G. C. (1989). Number of assessment centre dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, 74 (4), 611-618.
- Huysamen, G.K. (1996). *Methodology for the social and behavioural sciences*. Halfway House: Thomson.
- Jansen, P. & De Jongh, F. (1997). *Assessment centres: A practical handbook*. Chichester: Wiley.
- Jansen, P.G.W. & Stoop, B.A.M. (2001) The dynamics of assessment centre validity: Results of a 7-year study. *Journal of Applied Psychology*, 86 (4), 741-753.
- Kerlinger, F.N. & Lee, H.B. (2000). *Foundations for behavioural research* (4<sup>th</sup> ed.). New York : Harcourt.
- Kleinmann, M. (1993). Are rating dimensions in assessment centres transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology*, 78 (6), 988-993.
- Kleinmann, M., Kuptsch, C. & Koller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An International Review*, 45 (1), 67-84.
- Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability and discriminant validity. *Journal of Applied Psychology*, 86 (2), 255-264.
- 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.
- Lindell, M.K. & Whitney, D.J. (2001). Accounting for common method variance in cross-sectional research designs. *Journal of Applied Psychology*, 86 (1), 114- 121.
- Promotional Assessment Skills Service. (2001). *Guidelines and ethical considerations for assessment centre operations*. [Http://www.pass-prep.com/overview/ethical.htm](http://www.pass-prep.com/overview/ethical.htm).
- Riley, M., Ric-Hansen, K. & Rushmere, K. (2000, May). *Draft competency framework for contact centres*. London: MR Partnership.
- Robertson, I., Gratton, L. & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. *Journal of Occupational Psychology*, 60, 187-195.
- Schmidt, F. L. & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124 (2), 262-274.
- Schneider, J.R. & Schmitt, N. (1992). An exercise design approach to understanding assessment centre dimension and exercise constructs. *Journal of Applied Psychology*, 77 (1), 32-41.
- Shore, T.H., Shore, L.M. & Thornton, G.C. (1990). Construct validation of two categories of assessment centre dimension ratings. *Personnel Psychology*, 43, 101-115.
- Shore, T.H., Shore, L.M. & Thornton, G.C. (1992). Construct validity of self- and peer evaluations of performance dimensions in an assessment centre. *Journal of Applied Psychology*, 77 (1), 42-54.
- Spector, P. E. (1987). Method variance as an artifact in a self-reported affect and perceptions at work: Myth or significant problem? *Journal of Applied Psychology*, 72 (4), 438-443.
- Spector, P.E., Schneider, J.R., Vance, C.A & Hezlett, S.A. (2000). The relation of cognitive ability and personality traits to assessment centre performance. *Journal of Applied Social Psychology*, 30 (7), 1474-1491.
- Spenser, L.M. & Spenser, S.M. (1993). *Competence at work*. New York: Wiley.
- Theron, D. & Roodt, G. (2001). An evaluation of the 360-degree project management competency assessment questionnaire. *Journal of Industrial Psychology*, 27 (2), 51-56.
- Thornton, G.C., (1992). *Assessment centres in human resource management*. Massachusetts: Addison-Wesley.
- Turnage, J.J. & Muchinsky, P.M. (1984). A comparison of the predictive validity of assessment centre evaluations versus traditional measures in forecasting supervisory job performance: Interpretative implications of criterion distortion for the assessment paradigm. *Journal of Applied Psychology*, 69 (4), 595-602.
- Wilson, J.T. & Walwanis, M.M. (2000). Assessment centres 2000: New challenges for new directions. *Paper presented at the 46<sup>th</sup> annual meeting of the Southeastern Psychological Association*. New Orleans: Los Angeles.
- Woodruffe, C. (1998). *Assessment Centres: Identifying and developing competence*. London: Institute of Personnel Management.