

VALIDATION OF THE HIGH PERFORMANCE LEADERSHIP COMPETENCIES AS MEASURED BY AN ASSESSMENT CENTRE IN-BASKET

H H SPANGENBERG

Centre of Leadership Studies (Southern Africa)

Graduate School of Business

University of Stellenbosch

C C THERON

Department of Industrial Psychology

University of Stellenbosch

ABSTRACT

The purpose of this study was to validate Schroder's High Performance Leadership Competencies (HPLCs), measured by a specially designed In-basket, against multiple criteria. These consisted of six measures of managerial success, representing managerial advancement and salary progress criteria, and a newly developed comprehensive measure of work unit performance, the Performance Index. An environmental dynamism and complexity questionnaire served as moderator variable. Results indicated disappointing predictive validity quotients for the HPLCs as measured by an In-basket, in contrast to satisfactory predictive and construct validity obtained in previous studies by means of a full assessment centre. The implications of the findings are discussed and suggestions are made for improving the validity of the In-basket.

OPSOMMING

Die doel van hierdie studie was die validering van Schroder se Hoëvlak Leierskapsbevoegdheids, gemeet deur 'n spesiaal ontwerpte Posmandjie, teen veelvoudige kriteria. Dit behels ses metings van bestuursukses wat bestuursbevorderings- en salarisvorderingskriteria insluit, sowel as 'n nuutontwikkelde, omvattende meting van werkeenheidsprestasie, die Prestasie-indeks. 'n Vraelys wat die dinamika en kompleksiteit van die omgewing meet, het as moderator veranderlike gedien. Resultate dui op teleurstellende geldigheidskwosiente vir die Hoëvlak Leierskapsbevoegdheids soos gemeet deur 'n posmandjie, in teenstelling met bevredigende voorspellings- en konstruktiewe geldigheid wat in vorige studies deur middel van 'n volle takeersentrum verkry is. Die bevindinge word bespreek en voorstelle word gemaak om die geldigheidskwosiente te verbeter.

Validation of the High Performance Leadership Competencies as measured by an assessment centre In-basket

Since 1997 the High Performance Leadership Competencies (HPLCs) (Schroder, 1989) have been used by various South African organisations for guidance with strategic planning and the assessment and development of executive and senior management. The HPLCs are typically measured by means of an assessment centre. The University of Stellenbosch Business School (USB) has recently initiated research on the leadership patterns of MBA students. Since the use of an assessment centre for research purposes is expensive and time consuming, it was decided to rather measure the HPLCs by means of an In-basket only, a key exercise of the assessment centre. Schroder was commissioned to develop such an exercise specifically for the USB. The purpose of the current study was to examine the validity of unit performance related inferences made from the HPLCs as measured by the USB In-basket.

The assessment centre and its validity

The essential feature of an assessment centre is the use of situational tests (simulations) to observe specific behaviours demonstrated by a participant (Thornton, 1992). Assessment Centre Guidelines specify the elements of an assessment centre as follows: Competencies based on job profiling and relevant organisational analyses. Exercises or simulations are designed to provide information for the competencies being assessed. Multiple exercises such as an In-basket, group and individual meetings, fact-finding exercises, interviews, and psychometric tests are used. Multiple observers are used to observe and evaluate each participant. A rigorous method for gathering and reporting data is used. Observers use a systematic procedure for recording specific behaviour observations and prepare a report on each exercise observed. Data from the various exercises are pooled and final ratings for a participant are obtained by means of a data

integration session or a validated statistical process (Assessment Centre Study Group, 1999). An assessment centre at managerial level may comprise an In-basket, a co-operative and/or competitive group exercise, interview simulation and fact-finding exercise

Since the introduction of the assessment centre more than forty years ago at AT&T in the United States of America, its use has spread widely. Assessment centres function in a wide variety of organisations including business, public service and semi-state organisations, education, and professions. Assessment centres are used at all organisational levels for the following purposes: Selection, placement, training and development, proficiency assessment, organisation development, and career and succession planning. Lately, the assessment centre has increasingly being used to lead change by defining and measuring the competencies and supporting behaviours required to meet future challenges (Howard, 1997). The criterion-related validity of assessment centres is well established. Meta-analytic studies provide evidence that observer ratings are predictors of a variety of managerial success criteria (Gaugler, Rosenthal, Thornton & Bentson, 1987; Hunter & Hunter, 1984; Schmitt, Gooding, Noe & Kirsh, 1984). The average validity coefficient is approximately .40. Validation criteria include, for example, career progress, overall performance ratings, dimensional performance ratings, potential ratings, wages, and training performance. In a South African study in which a middle management assessment centre was validated against Behaviourally Anchored Rating Scales (n=110), a multiple correlation of .37 (after shrinkage) was reported (Spangenberg, Esterhuyse, Visser, Briedenhann & Calitz, 1989). The consistent evidence of criterion-related validity is probably the reason for the vast expansion of assessment centres internationally, especially during the past 20 years.

In contrast, construct validity of the assessment centre analysed by means of multi-trait multi-method (MTMM) data has been low, indicating method (exercise) factors rather than stable personality characteristics (competencies) as determinants of assessment centre ratings (Sagie & Magnezy, 1997). Recent research indicates, however,

that construct validity of the assessment centre may be improved by careful construction. This includes, for instance, splitting the broad concept of leadership into individual and team leadership which represent separate constructs that require different skills; rating of behaviour only after obtaining sufficient data to do so; and providing comprehensive training to observers (Howard, 1997). Furthermore, better statistical analyses and more careful construction principles can improve the construct validity of assessment centres. Kleinmann and Koller (1997) researched two possible causes of low construct validity indices, namely the way of analysing MTMM data by means of confirmatory factor analysis, and the way assessment centres are constructed. Based on a paper by Marsh (1989), Kleinmann and Koller (1997) proposed that exercise effects should be modelled by way of correlated uniqueness to ensure that the uniqueness of the observable variable contain both error variance and variance explained by method effects. Applying this kind of analysis to MTMM data leads to more valid estimations regarding the convergent and discriminant validity of assessment centres. By reanalysing the data provided by Bycio, Alvares and Hahn (1987), Kleinmann and Koller (1997) found the impact of dimension factors to be substantially higher than initially reported by the authors. In their own study, Kleinmann and Koller (1997) brought about structural changes to their assessment centre such as improving observability of the behaviour dimensions, and limiting the number of behaviour dimensions to be observed. When they analysed the MTMM matrix by means of confirmatory factor analysis, they found evidence of substantial dimension effects, which were the highest for a model with three dimension factors. These factors explained more than 40% of the behaviour variance, implying adequate convergent validity of the assessment centre. In two studies related to observability (Kleinmann, 1993; Kleinmann, Kuptsch & Koller, 1996), it was found that transparency enhanced convergent construct validity. Kleinmann et al. (1996) contended that lack of transparency might be the cause of low discriminant validity and that this deficit might be rectified by informing candidates in an assessment centre about the dimensions being measured and the kind of behaviours that would be relevant for each exercise. It was subsequently found that with knowledge of behaviour requirements, candidates behaved more consistently and received more consistent ratings from observers on identical dimensions across different exercises. By combining the above research with results of other related findings (Harris, Becker & Smith, 1993; Shore, Shore & Thornton, 1992), it transpired that with increased transparency the candidates can more clearly show behaviours connected with dimensions (Salgado, 1999). In the same way observers can obtain clearer, more comprehensive behaviour information on which to base their ratings. Based on overall research results, Kleinmann and Koller (1997) concluded that the construct validity of assessment centres has been underestimated in the past.

The validity of the High Performance Leadership Competencies (HPLCs) developed by Schroder has been established by several studies (Chorvat, 1994; Cockerill, 1989; Cockerill, Schroder & Hunt, 1993; Schroder, 1989). To measure these competencies, Schroder designed an assessment centre operating in a dynamic environment. This assessment centre comprised four exercises, namely an In-basket, an information search interview, and collaborative and competitive group exercises. Behaviourally anchored rating scales were constructed to measure observed behaviour. In a criterion-related validity study involving 58 middle managers from a utility company, assessment centre ratings of all the competencies except for self-confidence and presentation were significantly related to the managers' work unit measures (Schroder, 1989). In another criterion-related study, Cockerill (1989) used structured behaviour observation to assess the eleven HPLCs. Each of 30 senior managers in a large global financial services organisation in the UK was observed over active periods for sixteen hours. The same competency rating scale was used as the one in the previous study. All the HPLCs correlated significantly with performance measures.

In a major study in the United Kingdom involving 150 senior managers from five organisations, the hypothesis was tested that

the HPLCs were generic (i.e. are valid across different industries) and are specifically predictive in dynamic environments (Cockerill, Schroder & Hunt, 1993). The performance measure was a 360° instrument based on Nicholson and Brenner's (1994) systems model of organisational performance. It assessed four dimensions of performance, namely outputs, climate, adaptability and resource inputs. Results indicated that all the competency factors except achievement orientation were significantly related to one or more dimensions of performance across these five different organisations. Importantly, this study demonstrated that the HPLCs were significantly related to performance only when units were operating in dynamic environments.

Construct validity of the HPLCs was examined in a study by Chorvat (1994) that involved 207 middle and senior British managers. A one-day Schroder-based assessment centre was used to measure the 11 HPLCs. This assessment centre was designed in such a way that multiple measures of the HPLCs were obtained with multiple methods, i.e. an MTMM design. This allowed confirmatory factor analysis with competing hierarchically nested models to be utilized, permitting rigorous examination, including statistical tests of construct validity and the presence of trait (dimension) and method (exercise) variance. High levels of construct validity were found for the eleven HPLCs. Based on the findings of the above criterion and construct validity studies, the HPLCs were slightly modified. The revised competencies are presented in Table 1.

TABLE 1
THE SCHRODER HIGH PERFORMANCE LEADERSHIP COMPETENCIES

THINKING CAPABILITIES

1. Informational Capability (IC)
The breadth of current and future information gathered and exchanged with regards to issues.
2. Conceptual Capability (CC)
Linking different kinds of information and ideas to form diagnostic and system-level concepts about a desired future.
3. Strategic Capability (SC)
Designing alternative routes to support learning about change and how to reach desired futures.

LEARNING CAPABILITIES

4. Developmental Capability (DC)
Providing stretching job opportunities and facilitating the generation of developmental feedback and competence development.
5. Interpersonal Learning (IL)
Sharing ideas in a non-evaluative setting to gain an understanding of the "other's" ideas from their viewpoint.
6. Cross-Boundary Learning (CBL)
Facilitating dialogue about shared ideas to form higher-level, explanatory team ideas about change.

INSPIRATIONAL CAPABILITIES

7. Purpose Building (PB)
Building commitment to shared purposes which are owned and used by members to initiate new thinking and ideas.
8. Confidence Building (CB)
Building a unit/organisation in which members value the reactions of others to their ideas, feel confident that they will succeed and celebrate the successes they achieve.

ACTION CAPABILITIES

9. Proactive Capability (PC)
Reduces organisational constraints and controls on members so they can take broader responsibility and use discretion in putting ideas about direction/change into action.
 10. Achievement Capability (AC)
Setting progressive measures of challenging objectives so that members can use performance feedback to learn and continuously improve performance.
-

The In-basket and its validity

There are, however, practical constraints with the use of assessment centres for research purposes, for instance time and cost. The cost of putting one candidate through a one-day assessment centre (which is the absolute minimum time required) is approximately R3 500. In contrast, application of an In-basket takes approximately 3 hours to complete. Depending on facilities, a group of up to 50 candidates can be assessed during one session. The current rate for scoring an In-basket and writing a report on the findings is approximately R450 per individual.

A survey in the USA indicated that 81% of assessment centres used an In-basket (Thornton, 1992). This finding corroborates the conception that in the South African context all managerial assessment centres include an In-basket exercise. The in-basket forms an integral part of the managerial assessment centre and is considered a major exercise in terms of complexity, number of competencies measured, and time allotted for administration and scoring.

Research on content, criterion-related, and construct validity of the In-basket up to 1989 was reviewed by Schippman, Prien & Katz (1990). They made three conclusions. Firstly, in spite of weaknesses inherent in the research and development methodologies applied during the 1960's, it would seem that early conclusions about the usefulness of In-basket measures of performance remained to some degree still valid. The authors contended that the technique's high face validity, combined with the compelling need to develop alternative assessment methods, might have made research seem less important. Furthermore, the In-basket never really developed an identity apart from the assessment centre. Therefore, supporting research directed specifically at the In-basket was neglected.

Secondly, evidence from various studies indicated that In-baskets could be reliably scored, although reliability values obtained were modest. It seems that that evidence of criterion-related validity was at best marginal and generally higher in settings where the In-basket was specifically constructed for a specific target position. However, In-baskets developed for a particular position were not very common and that generic products were being used more frequently. Based on a limited sample of four studies, Schippmann et al. (1990) found evidence of construct validity where the In-basket was designed to measure a specific theoretical construct. They concluded that the evidence of these studies was encouraging but not convincing in terms of either hypothesised constructs or work performance criterion constructs.

Thirdly, the most significant finding was that research and reporting in this area were very fragmented. Studies and reports that did include empirical evaluations often suffered from methodological and conceptual shortcomings that limited their usefulness.

More conclusive results about construct validity of the In-basket were obtained in a recent study by Rolland (1999) in which dimension ratings from an In-basket were examined. Two separate samples were used. The existence of ten dimensions based on four underlying postulated clusters were tested through exploratory and confirmatory factor analyses (sample 1: N=441; sample 2: N=355). Results indicated that the 10 dimensions assessed were not independent, and were not clustered according to the underlying hypothesised constructs. Rolland's results were fairly similar to those of Brannick, Michaels & Baker (1989) who found only some evidence of convergent validity and very little discriminant validity for the in-basket. Rolland concluded that results from his study, combined with previous research "question the validity of inferences about the 'different' managerial traits derived from In-basket scores" (1999, p. 254).

The aim of this study is to investigate the validity of Schroder's High Performance Leadership Competencies measured by a

specially designed In-basket. The underlying assumption is that a carefully constructed In-basket (measuring independent high level leadership competencies), scored by thoroughly trained assessors, will significantly predict a range of criteria, including work unit performance as well as managerial success criteria reflecting career advancement and salary progress.

In view of the finding that HPLCs predict performance significantly only in dynamic environments (Cockerill et al., 1993), it was decided to include a measure of environmental dynamism and complexity as a moderator variable.

METHOD

Measuring Instruments

Predictor measures:

In-Basket.

The Centre of Leadership Studies (Southern Africa) was in a fortunate position to have available an In-basket that was developed specifically for the Graduate School of Business of the University of Stellenbosch by Schroder for the purpose of measuring the Schroder's High Performance Leadership Competencies described above. The In-basket reflects a complex, dynamic setting in the somewhat futuristic organic food industry.

A unique feature of the application of this In-basket was the extensive programme for training and certification of assessors. In contrast to Assessment Centre Guidelines that prescribe approximately one day of training for in-basket scoring (Assessment Centre Study Group, 1999), a period of eight working days spread over two weeks was used for initial assessor training. Schroder's requirement for certification of assessors was .85 reliability, which means that an assessor must consistently reach .85 agreement with him as an expert assessor. Quality control arrangements were made to ensure reliability of scoring. Accuracy of ratings was monitored for a period of more than six months. Each quality check was accompanied by recommendations by professor Schroder. Assessors were certified as competent assessors only after this training period. At a later stage, due to natural attrition of assessors, additional assessors were trained by Schroder.

This In-basket measures all of the 10 HPLCs presented above. Some of the raters omitted to provide ratings on the interactive competencies of interpersonal learning and cross-boundary learning. In such cases a median rating of '2' was allotted.

Criterion measures:

Criterion measures consisted of a number of managerial success indicators as well as a 360° assessment instrument that measures work unit performance (Performance Index, PI).

1. Managerial success indicators. In a comprehensive review of the validity of personnel selection methods, Schmidt and Hunter (1998) reported that assessment centres predict managerial success criteria such as rate of promotion and advancement, in addition to work performance. In this study, criterion measures were selected to cover two broad areas, namely managerial advancement and salary progress.
 - 1.1 Managerial advancement criteria
 - 1.1.1 Managerial Advancement Quotient (MAQ). (Hall, 1976).

$$MAQ = \frac{5(5 - \text{level}^*) \times 100}{\text{Age}}$$
 - 1.1.2 Managerial Success Index (MSI). (Luthans, Rosenkrantz & Hennessey, 1985).

$$MSI = \frac{5(5 - \text{level}^*) \times 100}{\text{Organisation}^{\wedge} \text{tenure}}$$

1.1.3 Promotion Index (McCall & Segrist, 1980).

Promotion Index = 5 (5 - level*) x 100
 Total years of service in all organizations

* Level in MAQ, MSI, and Promotion Index were measured on a scale of 1-4.

^ Organisation refers to current organisation.

1.1.4 Number of promotions

Positions, following initial position, which represented a promotion

1.2 Salary progress

1.2.1 Current salary

1.2.2 Salary increment

Current salary divided by first work salary, directly after finishing technikon or university studies to qualify for a profession.

2. Performance Index (PI)

The Performance Index is a newly developed measure of work unit performance. It was developed for the purpose of diagnosing the health and effectiveness of organisational work units, as well as to serve as a validation criterion for research purposes. The questionnaire measures eight independent dimensions by means of 56 items, on a five-point scale. See Table 2. The development and psychometric properties of the PI are discussed elsewhere (Theron & Spangenberg 2002).

TABLE 2
BRIEF SUMMARIES OF PERFORMANCE INDEX DIMENSIONS

1. Production and efficiency include quantitative outputs such as meeting goals, quantity, quality and cost-effectiveness, and task performance.
2. Core people processes reflect organisational effectiveness criteria such as goals and work plans, communication, organisational interaction, conflict management, productive clashing of ideas, integrity and uniqueness of the individual or group, learning through feedback and rewarding performance.
3. Work unit climate is a global perception of the psychological environment of the unit, and gives an overall assessment of the integration, commitment and cohesion of the unit. It includes working atmosphere, teamwork, work group cohesion, agreement on core values and consensus regarding the vision, achievement-related attitudes and behaviours and commitment to the unit
4. Employee satisfaction, and centres around satisfaction with the task and work context, empowerment, and career progress, as well as with outcomes of leadership, e.g. trust in and respect for the leader and acceptance of the leader's influence.
5. Adaptability reflects the flexibility of the unit's management and administrative systems, core processes and structures, capability to develop new products/services and versatility of staff and technology. Overall, it reflects the capacity of the unit to react appropriately and expeditiously to change.
6. Capacity (wealth of resources) reflects the internal strength of the unit, including financial resources, profits and investment; physical assets and materials supply; and quality and diversity of staff.
7. Market share/scope/standing includes market share (if applicable), competitiveness and market-directed diversity of products or services, customer satisfaction, and reputation for adding value to the organisation.
8. Future growth serves as an overall index of projected future performance and includes profits and market share (if applicable), capital investment, staff levels and expansion of the unit.

Moderator variable:

Environmental Dynamism and Complexity Survey

The Environmental Dynamism and Complexity Survey measures four dimensions, i.e. dynamism, complexity, technical sophistication and munificence (Cockerill et al., 1993.) Since the Dynamism and Complexity Survey was a

cumbersome instrument to use, it was adapted for this study. Furthermore, it was expanded to include 31 items. Modifications included differentiating between internal and external change, improving the layout of the questionnaire in order to make it user-friendly, consistently using a five-point scale, and using behavioural descriptions on scale points 5, 3, and 1 to facilitate clarity. The model of the enlarged survey instrument is presented in Table 3. Brief descriptions of item clusters, e.g. rate of change, and items belonging to each cluster, are indicated for each of the four dimensions.

TABLE 3
ENVIRONMENTAL DYNAMISM AND COMPLEXITY MODEL (ADAPTED FROM COCKERILL, SCHRODER & HUNT, 1993)

Dynamism	Complexity	Technological Sophistication	Munificence
Degree of change ● due to external organizational pressure (1-7) ● due to internal organizational pressure 8-12)	Frequency of having to satisfy needs of stakeholders (15-18)	Technological sophistication in environment (27-28)	Opportunities for growth (25-26)
Overall rate of change due to ● external organizational environment (13) ● internal organizational environment (14)	Number of stakeholders (19)	Research and Development orientation in organization (30)	Control in environment: allowance for initiative (29)
Safety of environment (21-22)	Diversity of demands made by stakeholders (20)		Degree of stress and hostility in the organizational environment (31)
Predictability of environment (23-24)			

Sample

The total sample consisted of Modular English students of the 1998 intake, and all Modular and Part Time MBA students of the 1999 and 2000 intakes. Since one of the major criteria in the study, namely work unit performance, requires full time tenure as a manager of an organisational work unit, the Full Time group was excluded from the study. The In-basket sample is presented in Table 4.

TABLE 4
IN-BASKET SAMPLE

MBA Group	1998	1999	2000	Total
Modular English	68	70	52	190
Modular Afrikaans	-	58	42	100
Part Time	-	62	57	119
				409

Since only MBA students that manage work units qualified for the study of the validation of the HPLCs against work unit performance, the sample for this part of the study was much smaller. Out of a possible number of 115 eligible unit managers, 60 participated in the study. This figure represents a 52 % participation that can be considered as satisfactory. See Table 5 for the sample for validating the HPLCs against the Performance Index.

TABLE 5
SAMPLE FOR VALIDATING HPLCs AGAINST
THE PERFORMANCE INDEX

MBA Group	1998	1999	2000	Total
Modular English	6	7	15	28
Modular Afrikaans	-	12	11	23
Part Time	-	6	3	9
				60

Application of instruments

At the beginning of each year all entrant students completed an In-Basket, as well as a Biographical Questionnaire. The Environmental Dynamism and Complexity Questionnaire, as well as a short questionnaire containing sensitive career information relating to salary and career progress, was completed during the second semester of 2000.

Concurrently, students who manage work units or departments were asked to have a 360° performance evaluation done of their units, using the PI. A work unit or department was defined in terms of the following criteria: It could be large or small, with at least 3 subordinates; be nested in a public, private or not-for-profit organisation; and have its own goals and measure its own performance.

Fieldwork for the PI was a laborious effort that took longer than anticipated. The main reason for slow progress was the fact that MBA students, in the prime of their careers, had very busy schedules by combining heavy workloads with study commitments. This phase of the project was completed by the end of the second term of 2001. Questionnaires were perused for proper completion as well as for patterns of missing values. Furthermore, work unit managers for whom only one or two questionnaires were completed were excluded from the study.

STATISTICAL ANALYSIS AND RESULTS

Correlation between the HPLCs and dimensions of unit performance

The High Performance Leadership Competencies are viewed as relatively stable sets of managerial behaviour, which should produce significantly superior work unit performance through their effect on specific unit performance dimensions. Correlating the high performance leadership competencies with the separate dimensions of unit performance without a guiding nomological network that maps the HPLCs on the separate unit performance dimensions, therefore seems to constitute an open invitation to the opportunistic interpretation of significant correlations. Theron and Spangenberg (2002) developed a hypothesis on the internal structure of the Performance Index (PI) described above. Figure 1 extends on this, as yet untested hypothesis, by postulating specific paths between the HPLC and the separate unit performance dimensions.

Table 6 depicts correlations between the high performance leadership competencies and the eight unit performance dimensions assessed by the Performance Index. Only interpersonal learning shows a borderline significant ($p < 0,07$), although weak positive correlation (0,213) with the overall unit performance score on the Performance Index. When turning the attention to the individual unit performance dimensions, only a slightly more encouraging picture emerges. Information competence correlates moderately negative (-0,288) and statistically significantly ($p < 0,05$) only with the unit performance dimension of employee satisfaction. Both the fact that information competence correlates significantly with this specific performance dimension and the negative nature of the correlation constitute somewhat unexpected, and in that sense,

disconcerting results (see Figure 1). Interpersonal learning shows significant ($p < 0,05$) and moderately positive correlations with core people processes (0,286), climate (0,299) and employee satisfaction (0,294). All three these relationships are in accordance with the structural model depicting the manner in which the HPLCs are expected to impact on the unit performance dimensions. Developmental competence correlates low positive and statistically significantly ($p < 0,05$) with the unit performance dimensions of future growth (0,305). It was expected, however, that developmental competencies would have correlated with more immediate criteria such as production and efficiency, climate, and satisfaction. Proactive competence correlates low negative (-0,230) and statistically significant with the production and efficiency performance dimension. The remaining competencies, namely, conceptual competence, strategic competence, cross boundary learning, confidence building and achievement competence all correlate low and insignificantly ($p > 0,05$) with all eight unit performance dimensions. But for the hypothesized relationships between interpersonal learning and core people processes, climate, and employee satisfaction, the study fails to corroborate all other hypothesized linkages between the high performance leadership competencies and the dimensions of unit performance (see Figure 1).

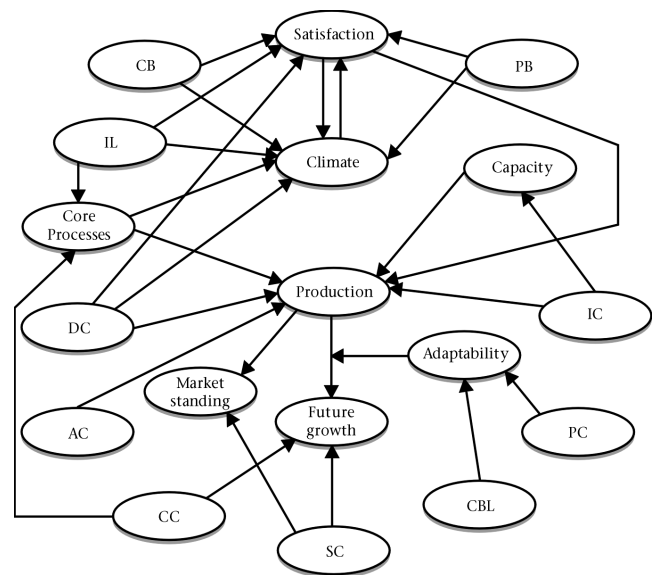


Figure 1: Hypothesized linkages between HPLC and the dimensions of unit performance

It is disappointing to note from Table 6 that the overall competency assessment obtained from the In-Basket correlates low and statistically insignificant ($p > 0,05$) with all eight unit performance dimensions and with the overall unit performance score.

Correlation between the HPLCs and various objective, individual managerial success criteria

Table 7 portrays the correlations between the individual high performance leadership competencies and six objective criteria of managerial success. The overall competency assessment obtained from the In-basket correlates consistently low and statistically insignificant ($p > 0,05$) with all six individual indicators of managerial success. Information competence correlates low (0,265) and statistically significantly ($p < 0,05$) with the number of promotions. Interpersonal learning correlates moderately positive (0,404), and statistically significant ($p < 0,05$), with the salary increment index. Developmental competence correlates low positive (0,233) and statistically significant with number of promotions. The remaining competencies, namely conceptual competence, strategic competence, cross boundary learning, purpose building, confidence building, proactive competence and achievement competence all correlate low and insignificantly ($p > 0,05$) with all six managerial success criteria.

TABLE 6
CORRELATIONS BETWEEN HPLCs AND UNIT PERFORMANCE

	IC	CC	SC	IL	CBL	DC	PB	CB	PC	AC	Mean
Production & efficiency	-.124 .186 54	-.048 .364 54	-.080 .281 54	.051 .356 54	.029 .419 54	-.064 .323 54	.092 .253 54	-.051 .356 54	-.230* .047 54	-.054 .349 54	.002 .494 53
Core people processes	-.089 .260 54	-.022 .436 54	.096 .244 54	.286* .018 54	.030 .414 54	.011 .469 54	.192 .082 54	.110 .214 54	-.090 .259 54	.086 .267 54	.110 .217 53
Climate	-.15 .136 54	.003 .491 54	.124 .186 54	.299* .014 54	.020 .442 54	-.166 .116 54	.213 .061 54	.062 .328 54	-.147 .144 54	.149 .141 54	.122 .191 53
Satisfaction	-.288* .017 54	-.017 .451 54	.107 .220 54	.294* .015 54	-.012 .465 54	-.050 .358 54	.083 .276 54	-.013 .464 54	-.149 .142 54	.132 .171 54	.054 .349 53
Adaptability	-.116 .202 54	-.033 .407 54	-.007 .481 54	.136 .163 54	-.117 .201 54	.116 .203 54	.088 .263 54	.040 .386 54	-.078 .288 54	.049 .363 54	.030 .415 53
Capacity	-.048 .366 54	-.051 .357 54	.027 .423 54	.073 .300 54	-.036 .399 54	.094 .249 54	.122 .190 54	.129 .176 54	-.055 .347 54	.115 .203 54	.072 .305 53
Market share	-.035 .400 54	.015 .456 54	.021 .441 54	.063 .326 54	-.170 .109 54	-.104 .227 54	.021 .441 54	.053 .352 54	-.163 .120 54	.099 .238 54	.042 .382 53
Future growth	-.092 .253 54	.070 .307 54	-.120 .193 54	.124 .186 54	.004 .488 54	.305* .012 54	.083 .275 54	.034 .403 54	.003 .490 54	.166 .115 54	.123 .189 53
PI	-.143 .152 54	-.013 .462 54	.040 .388 54	.213 .061 54	-.033 .405 54	.008 .478 54	.139 .158 54	.061 .329 54	-.140 .156 54	.118 .197 54	.089 .263 53

* Correlation is significant at the 0.05 level (1-tailed).

** Correlation is significant at the 0.01 level (1-tailed).

TABLE 7
CORRELATIONS BETWEEN HPLCs AND OBJECTIVE, INDIVIDUAL INDICATORS OF MANAGERIAL SUCCESS.

	IC	CC	SC	IL	CBL	DC	PB	CB	PC	AC	Mean
No. of promotions	.265* .029 52	.151 .143 52	-.030 .415 52	.021 .441 52	.054 .353 52	.233* .048 52	-.091 .262 52	.184 .096 52	.164 .123 52	.208 .070 52	.173 .112 51
Current salary	.016 .457 52	-.024 .433 52	-.042 .384 52	-.011 .471 52	.079 .288 52	.064 .326 52	-.052 .357 52	.037 .398 52	.027 .424 52	.034 .404 52	.115 .211 51
Salary increase	.019 .446 52	.199 .079 52	-.111 .216 52	.404** .002 52	-.068 .315 52	.092 .258 52	-.114 .210 52	.041 .386 52	.114 .209 52	.078 .291 52	.024 .433 51
MAQ	.218 .063 51	-.034 .405 51	.109 .224 51	-.066 .322 51	-.072 .309 51	.043 .383 51	.148 .150 51	.029 .421 51	.046 .374 51	-.051 .362 51	.084 .281 50
MSI	-.196 .087 50	.042 .387 50	-.139 .168 50	.024 .435 50	.130 .184 50	.071 .312 50	-.037 .398 50	-.036 .403 50	.176 .111 50	.185 .099 50	.155 .144 49
PMI	.153 .145 50	-.049 .368 50	.052 .359 50	.089 .268 50	.167 .123 50	-.102 .240 50	.189 .094 50	.057 .346 50	-.177 .109 50	-.009 .474 50	.121 .204 49

* Correlation is significant at the 0.05 level (1-tailed).

** Correlation is significant at the 0.01 level (1-tailed).

The moderating role of environmental dynamism and complexity

The high performance leadership competencies as assessed by an assessment centre are reported to affect managerial performance especially in more dynamic and complex organizational environments (Cockerill et al., 1993). In addition, it could probably be argued that environmental dynamism and

complexity in and by itself will also affect managerial performance. It could therefore be argued that a measure of environmental dynamism and complexity should explain variance in a measure of managerial performance when added to a model already containing a composite competency main effect. Specifically, both the dynamism-complexity main effect and in the complexity-competency interaction effect should

explain unique variance in managerial performance not explained by the competency main effect.

The Environmental Dynamism and Complexity Survey (EDCS) comprise fourteen indicators of the dynamism and complexity of the environment in which a managerial unit operates (see Table 3). Item analysis was performed on all 31 items in anticipation of the formation of a dynamism-complexity index. Two items, one on environmental control and one on environmental stress, had to be reflected. The coefficient of internal consistency for the reduced EDCS was 0,8653. The EDCS items were next combined into an unweighted linear combination to serve as an index of environmental dynamism and complexity.

The overall unit performance score derived from the PI was subsequently regressed on a linear combination of the composite competency score, derived from the In-basket, the dynamism-complexity index and a competency and complexity interaction term. To be able to convincingly claim that the HPLC In-basket assessments explain variance in managerial performance, especially in a dynamic and complex environment, which in itself hinders effective management, the saturated model needs to fit the data well. This would require that the linear composite significantly explains variance in unit performance, that all partial regression coefficients are significant and that the signs of regression coefficients are in the expected direction. The overall competency assessment should be positively related to unit performance for any fixed level of environmental complexity. Furthermore, environmental complexity should be positively related to unit performance for any fixed competency level (given that low EDCS scores indicate dynamic, complex environments), but the rate at which unit performance increases as competency levels improve should accelerate as the environment becomes more dynamic and complex (a negative regression weight for the interaction term). Table 8, however, indicates that none of the three effects significantly ($p > 0,05$) explain unit performance variance not

explained by the other effects in the model. The directions of the signs of the three partial regression coefficients also do not correspond with expectations. The saturated model, furthermore, only explains variance in unit performance with borderline significance ($F=2,257; 3,47; p < 0,10$), and also only in rather modest proportions ($R^2=0,126$). The zero-order correlation of the competency main effect with overall unit performance is insignificant ($0,089; p > 0,05$). The addition of the dynamism-complexity main effect and the complexity-competence interaction effect to the basic model does significantly explain variance in unit performance not explained by competence ($F=6,328; 2,47$). However, when the saturated regression model is reduced by dropping the interaction term, Table 9 indicates that the overall competency score derived from the in-basket still does not significantly ($p > 0,05$) explain variance in unit performance even when controlling for environmental dynamism and complexity. The dynamism-complexity main effect does, however, significantly ($p < 0,05$) explain variance in unit performance when added to a model already containing a competence main effect. If, instead, the saturated model is reduced by dropping the complexity main effect and retaining the interaction term, Table 10 indicates that the overall competency score derived from the In-basket still does not significantly ($p > 0,05$) explain variance in unit performance when included in a model already containing the interaction term. The complexity-competence interaction effect significantly ($p < 0,05$) explains variance in unit performance when partialling out the competency main effect. This reduced model explains approximately 10% of the variance in unit performance. Table 11 reflects the zero-order correlations between the variables included in the reduced regression model. The zero-order correlation for the interaction term is low negative ($-.225$) and only of borderline significance ($p < 0,07$). When partialling out the effect of the interaction term from both the criterion and the competency main effect, the unique variance in the latter explains approximately 5% ($0,222^2$) of the unique variance in unit performance (see Table 10). When partialling out the effect of the interaction term from the

TABLE 8
REGRESSION OF UNIT PERFORMANCE ON COMPETENCE AND ENVIRONMENTAL DYNAMISM AND COMPLEXITY; SATURATED MODEL

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	1.361	3	.454	2.257	.094
Intercept	.751	1	.751	3.737	.059
MEAN	.193	1	.193	.961	.332
CSURVEY	.311	1	.311	1.545	.220
SURVEYIN	.228	1	.228	1.135	.292
Error	9.449	47	.201		
Total	624.221	51			
Corrected Total	10.810	50			

a R Squared = .126 (Adjusted R Squared = .070)

	Sum of Squares	df	Mean	F	Significance
Regression	1.361	3	.454	2.257	.094
Residual	9.449	47	.201		
Total	10.810	50			

	Unstandardized Coefficients		Standardized Coefficients	t	Significance	Correlations		
	B	Std. Error	Beta			Zero-order	Partial	Part
(Constant)	8.514	4.404		1.933	.059			
MEAN	-1.654	1.687	-.974	-.980	.332	.084	.142	-.134
CSURVEY	-1.822	1.466	-2.150	-1.243	.220	-.315	-.178	-.169
SURVEYIN	.600	.563	2.094	1.065	.292	-.225	.154	.145

TABLE 9
REGRESSION OF UNIT PERFORMANCE ON COMPETENCE AND ENVIRONMENTAL DYNAMISM AND COMPLEXITY; REDUCED MODEL

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	1.133	2	.567	2.810	.070
Intercept	6.046	1	6.046	29.990	.000
MEAN	6.001E-02	1	6.001E-02	.298	.588
CSURVEY	1.058	1	1.058	5.246	.026
Error	9.677	48	.202		
Total	624.221	51			
Corrected Total	10.810	50			

a R Squared = .105 (Adjusted R Squared = .068)

	Unstandardized Coefficients		Standardized Coefficients	t	Significance	Correlations		
	B	Std. Error	Beta			Zero-order	Partial	Part
(Constant)	3.883	.709		5.476	.000			
MEAN	.127	.232	.075	.546	.588	.084	.079	.075
CSURVEY	-.265	.116	-.313	-2.290	.026	-.315	-.314	-.313

TABLE 10

REGRESSION OF UNIT PERFORMANCE ON COMPETENCE AND ENVIRONMENTAL DYNAMISM AND COMPLEXITY; REDUCED MODEL WITH INTERACTION TERM

Source	Type III Sum of Squares	df	Mean Square	F	Significance
Corrected Model	1.051	2	.525	2.584	.086
Intercept	5.036	1	5.036	24.766	.000
MEAN	.506	1	.506	2.487	.121
SURVEYIN	.975	1	.975	4.796	.033
Error	9.760	48	.203		
Total	624.221	51			
Corrected Total	10.810	50			

a R Squared = .097 (Adjusted R Squared = .060)

	Unstandardized Coefficients		Standardized Coefficients	t	Significance	Correlations		
	B	Std. Error	Beta			Zero-order	Partial	Part
(Constant)	3.094	.622		4.977	.000			
MEAN	.418	.265	.246	1.577	.121	.084	.222	.216
SURVEYIN	-9.775E-02	.045	-.341	-2.190	.033	-.225	-.301	-.300

competency main effect only, the unique variance in the latter explains approximately 4% (0,216²) of the variance in unit performance (see Table 10). When the competency main effect is partialled out of both the criterion and the interaction effect, the unique variance in the latter explains approximately 9% (-0,301²) of the total criterion variance and approximately 9% (-0,300²) of the unique criterion variance.

The dynamism-complexity index correlates moderately negative (-0,345) and statistically significantly ($p < 0,05$) with overall unit performance (see Table 11), even when controlling for managerial competency (see Table 9). The composite competency rating is at any rate unrelated (-0,030) to environmental complexity. Given the fact that low scores on the EDCS indicate a dynamic, complex environment, the negative correlation is rather surprising.

DISCUSSION

The results reported here are somewhat disappointing. The high performance leadership competencies are viewed as relatively stable

sets of managerial behaviour, which should produce significantly superior work unit performance, especially in dynamic, complex organisational environments. Although limited support for the hypothesis was found on the level of individual competencies and specific unit performance dimensions, the study fails to convincingly corroborate the hypothesis on an aggregate level.

The expectation that the In-basket measures should be related to unit performance seems to have been reasonable given the fact that the In-basket

- was developed with care;
- for a specific target group;
- to elicit behaviours that reflect clearly defined high performance leadership competencies
- that were measured by means of behaviourally anchored rating scales which were developed to guide the scoring;
- using pre-selected assessors who were thoroughly trained and certified.

In addition, the psychometric properties of the Performance Index seem to be above suspicion (Theron & Spangenberg, 2002).

TABLE 11
INTER-CORRELATIONS BETWEEN REGRESSION TERMS

	Mean	PI_1	CSURVEY	SURVEYIN
Mean	1.000	.089	-.029	.475**
	.	.263	.421	.000
	53	53	51	51
PI_1	.089	1.000	-.294*	-.225
	.263	.	.013	.057
	53	64	58	51
CSURVEY	-.029	-.294*	1.000	.863**
	.421	.013	.	.000
	51	58	274	51
SURVEYIN	.475**	-.225	.863**	1.000
	.000	.057	.000	.
	51	51	51	51

** Correlation is significant at the 0.01 level (1-tailed).

* Correlation is significant at the 0.05 level (1-tailed).

Nonetheless it seems plausible that the failure to corroborate the hypothesised relationship between the high performance leadership competencies and unit performance, could be attributed to the inability of the In-basket to provide a finer distinction between competency levels in the intermediate range (i.e. in the current score interval 2-3) in conjunction with the relative homogeneity of the MBA student population. To the extent that the latent variable being assessed shows relatively little variance in the target population and to the extent that the In-basket is relatively insensitive to the little true score variance that exists, but is to a fixed extent susceptible to (random) measurement error, classical reliability of measurement must suffer. The problem with behaviourally anchored rating scales are that they in essence constitute single item scales. This would probably represent somewhat less of a problem if the behavioural anchors guiding the rating on each performance dimension would include expressions of different states of the underlying latent performance variable in close proximity to each other in the middle of the latent variable scale. However it is extremely difficult, if not practically impossible, to detect describable observed differences in the behavioural manifestations of small differences in the underlying latent performance variable. Classical measurement theory would, however, suggest that an increase in scale length (assuming parallel items are added) should increase true score variance and thereby enhance reliability (Lemke & Wiersma, 1976). Behaviour observation scales, mixed standard rating scales and summated checklists measure latent performance dimensions through multiple items. Therefore the use these rating scale formats, instead of the behaviourally anchored rating scale format used in the current instrument, might possibly assist in rendering finer distinctions on the latent variables being assessed and thus assist in the improvement of the reliability of the competency assessments. The behavioural observation scale format utilized in the scoring of in-basket protocols would, however, probably have to be adjusted slightly from the conventional format. The behavioural observation scale typically rates the frequency with which critical behavioural indicators of an underlying latent variable are displayed on a five or seven point scale. Since in-baskets elicit only a cross sectional sample of critical behavioural indicators, frequency ratings clearly would be inappropriate. The use of a simple dichotomy indicating whether the behavioural response was demonstrated or not, would be one possibility. Alternatively, a three-point response scale (No, Undecided, Yes) could be used.

The foregoing argument also seems to suggest that an item response theory approach (IRT) to item analysis and scale construction could possibly assist in countering the inability of

the In-basket to discriminate in the middle range of the latent trait (theta) scale. The nature of the response scale discussed earlier would have implications for the complexity level at which the analysis would have to be undertaken. IRT analysis provides a much more detailed picture of the measurement properties of test items than classical item analysis (Drasgow & Hulin, 1991). The ability of items to detect differences in the underlying latent trait at any specific point on the θ -axis typically would differ. The point on the underlying latent trait scale at which items are most discriminatory (or most informative) also normally would differ. Item analysis statistics derived from classical measurement theory are unable to reflect these differences in item functioning. These differences can, however, be captured by two- and three-parameter IRT models. In addition, IRT can estimate the amount of information any item provides about the latent trait at any value of the trait. Information is defined as the inverse of the standard error of the estimated ability θ^* for individuals with a fixed ability θ (Drasgow & Hulin, 1991). Likewise, the information about the latent trait provided by the scale at any fixed ability θ , can be determined. A target scale information function (Hulin, Drasgow & Parsons, 1983) that peaks in the interval on the θ -scale where maximum accuracy is desired (probably +1 to +2 on the θ -scale in the case of the In-basket), could consequently be approximated by selecting the majority of items with maximum information in that interval.

Overall, results of this study are in agreement with the sparse literature on In-basket validity described in the review of the literature. It confirms the variable validity of In-basket measures. In South Africa where the In-basket is often used as a single exercise to assess entry-level candidates or assessing candidates for promotion, the results of this study provide a clear warning signal to practitioners and researchers alike: The In-basket as a single exercise is not a viable alternative to the full assessment center. Furthermore, the intuitive belief in the overall validity and usefulness of the In-basket that lasted for three decades up to the end of the 1980s has been disproved by more recent research.

In the South African context a major intervention will be needed to ensure the effective and fair application of the In-basket as a single exercise. It should probably start with the sensitisation of practitioners to the scoring problem discussed above. The development of a scoring method that provides a wider range of ratings is needed. This would require experimenting with various In-basket scoring methods, using the same set of in-basket protocols.

Furthermore, a better understanding should be created of the construct validity of the In-basket and, indeed, of the full managerial assessment centre of which it forms an integral part. The literature review indicates signs of improving assessment centre construct validity through, for example, conceptualizing the underlying latent variables in terms of fewer and more clearly separated dimensions, and better understanding by participants of what is required of them in behavioural terms. Likewise, careful development and application of the In-basket, particularly with regard to scoring methodology, may improve the reliability and overall validity of the in-basket.

In order to reaffirm the utility of the In-basket, focussed and well-coordinated research is needed. In South Africa where assessment per se is viewed with circumspection, it is imperative that researchers, practitioners and consultants work together to improve the validity of this potentially very useful assessment instrument.

REFERENCES

- Assessment Centre Study Group. (1999). *Proposed guidelines and ethical considerations for assessment centres*. Adaptations to existing guidelines, presented by H.H. Spangenberg at the 19th Annual Conference of the Assessment Centre Study Group, Stellenbosch.

- Brannick, M.T., Michaels, C.E. & Baker, D.P. (1989). Construct validity of In-basket scores. *Journal of Applied Psychology, 74* (6), 957-963.
- Bycio, P., Alvares, K.M. & Hahn, J. (1987). Situational specificity in assessment centre ratings: A confirmatory factor analysis. *Journal of Applied Psychology, 72*, 463-474.
- Chorvat, V.P. (1994). *Toward the construct validity of assessment centre leadership dimensions: A multitrait-multimethod investigation using confirmatory factor analysis*. Unpublished doctoral dissertation, University of South Florida, Tampa, USA.
- Cockerill, A.P. (1989). *Managerial competence as a determinant of organisational performance*. Unpublished doctoral dissertation, University of London, England.
- Cockerill, A.P., Schroder, H.M. & Hunt, J.W. (1993). *Validation study into the high performance managerial competencies*. Unpublished report, London Business School, England.
- Drasgow, F. & Hulin, C.L. (1991). Item response theory. In M.D. Dunnette & L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2nd ed.) Volume 2. Palo Alto, California: Consulting Psychologists Press, Inc.
- Gaugler, B.B., Rosenthal, D.B., Thornton, G. C. III. & Bentson, C. (1987). Meta-analysis of assessment centre validity (Monograph). *Journal of Applied Psychology, 72*, 493-511.
- Hall, J. (1976). To achieve or not: The manager's choice. *California Management Review, 18*, 5-18.
- Harris, M.M., Becker, A.S. & Smith, D.E. (1993). Does the assessment centre scoring method affect cross-situational consistency of ratings? *Journal of Applied Psychology, 78*, 675-678.
- Howard, A. (1997). A Reassessment of assessment centers: Challenges for the 21st Century. *Journal of Social Behaviour and Personality, 12* (5), 13-53.
- Hulin, C.L., Drasgow, F. & Parsons, C.K. (1983). *Item response theory: An application to psychological measurement*. Homewood, Ill: Jones-Irwin.
- Hunter, J.E. & Hunter, R.F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.
- Kleinmann, M. (1993). Are assessment centre rating dimensions transparent for participants? Consequences for criterion and construct validity. *Journal of Applied Psychology, 78*, 988-993.
- Kleinmann, M. & Koller, O. (1997). Construct validity of assessment centres: Appropriate use of confirmatory factor analysis and suitable construction principles. *Journal of Social Behaviour and Personality, 12* (5), 65-85.
- Kleinmann, M., Kuptsch, C. & Koller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An International Review, 45*, 67-84.
- Lemke, E. & Wiersma, W. (1976). *The principles of psychological measurement*. Boston: Houghton Mifflin Company.
- Luthans, F., Rosenkrantz, S.A. & Hennessey, H.W. (1985). What do successful managers really do? An observation study of managerial activities. *The Journal of Applied Behavioural Science, 21* (3), 255-270.
- Marsh, H.W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and some solutions. *Applied Psychological measurement, 13*, 335-361.
- McCall, M.W. & Segrist, C.A. (1980). *In pursuit of the manager's job: Building on Mintzberg* (Technical Report No. 14). Greensboro, NC: Center for Creative Leadership.
- Nicholson, N. & Brenner, S.O. (1994). Dimensions of perceived organisational performance: Tests of a model. *Applied Psychology: An International Review, 43* (1), 69-108.
- Rolland, J.P. (1999). Construct validity of In-basket dimensions. *European Review of Applied Psychology, 49* (3), 251-259.
- Sagie, A. & Magnezy, R. (1997). Assessor type, number of distinguishable dimension categories and construct validity. *Journal of Occupational and Organisational Psychology, 70* (1), 103-109.
- Salgado, J.F. (1998). Personnel selection methods (Chapter 1). In C.L. Cooper & I.T. Robertson (Eds.). *International Review of Industrial and Organisational Psychology*. New York: John Wiley.
- Schippmann, J.S., Prien, E.P. & Katz, J.A. (1990). Reliability and validity of In-basket performance measures. *Personnel Psychology, 43*, 837-859.
- 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.
- Schmitt, N., Gooding, R.Z., Noe, R.A. & Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-422.
- Schroder, H.M. (1989). *Managerial Competence: The Key to Excellence*. Kendall Hunt: Dubuque, Iowa.
- Shore, T.H., Shore, L.M. & Thornton, G.C.111. (1992). Construct validity of self- and peer evaluations of performance dimensions in an assessment centre. *Journal of Applied Psychology, 77*, 42-54.
- Spangenberg, H.H., Esterhuysen, J.J., Visser J.H., Briedenhann, J.E. & Calitz, C. (1989). Validating an assessment centre against BARS - an experience with performance related criteria. *Journal of Industrial Psychology, 15* (2), 1-10.
- Theron, C.C. & Spangenberg, H.H. (2002). *Developing a performance measurement questionnaire for assessing organisational work unit effectiveness*. Paper presented at the Global Conference on Business and Management, Paris.
- Thornton, G. C. 111 (1992). *Assessment Centres in human resources management*. Reading, Massachusetts: Addison Wesley Publishing Company.