

MENTAL MODELS AS MODERATING VARIABLE IN 360 DEGREE COMPETENCY ASSESSMENTS

D. THERON
G. ROODT

Department of Human Resource Management
Rand Afrikaans University

ABSTRACT

The purpose of this study is to determine whether rater-groups have different mental models regarding the construct that is being assessed. Sixty-eight marketing employees, working in the petro-chemical industry, were evaluated by themselves, their managers, customers and peers. The performance dimensions included in the Marketing Competency Questionnaire (MCQ) were generated on a consensus basis between representatives of the different rater-groups in an attempt to improve the inter-rater reliabilities of the questionnaire. Item scores were inter-correlated and simplified factor scores were compiled for each of the rater-groups. These simplified factor scores were again inter-correlated and single factors with high internal consistencies for each rater-group were obtained. An inter-battery factor analysis between these factors yielded a single factor with different factor loads from the various rater-groups' factors. The implications for these findings are further discussed.

OPSOMMING

Die doel van hierdie studie is om vas te stel of beoordelingsgroepe verskillende denkbeelde het met betrekking tot die konstruk wat beoordeel word. Agt-en-sestig bemarkingspersone wat in die petro-chemiese bedryf werksaam is, is deur hulleself, hulle bestuurders, kliënte en medewerkers geëvalueer. Die prestasiedimensies wat in die Bemarkingsbeoogdeheidsvraelys (MCQ) ingesluit is, is op 'n konsensusbasis tussen verteenwoordigers van die verskillende beoordelingsgroepe gegenereer in 'n poging om die tussen-beoordelaarbetroubaarheid van die vraelys te verhoog. Itemtellings is geïnterkorreleer en vereenvoudigde faktortellings is vir elke beoordelingsgroep bereken. Hierdie vereenvoudigde faktortellings is weer geïnterkorreleer en enkelfaktore met hoë interne betroubaarheid vir elke groep is verkry. 'n Inter-batteryfaktorontleding tussen hierdie faktore het 'n enkele faktor met verskillende faktorloadings vir faktore van die verskillende beoordelingsgroepe opgelewer. Die implikasies van hierdie bevindinge word verder bespreek.

At present, with the emergence of flatter organisational structures, Goodale (1992) suggested that 360° performance assessment (multi-rater assessment) is a product of participative management. Jones and Bearley (1996) defined 360° assessment as the multi-rater assessment of an individual where all the assessment data is fed back to the particular individual. In rapidly changing conditions, multi-rater assessment thus fulfils the need for providing the individual with a more holistic and useful set of feedback criteria. According to Edwards (1998) 360° assessments compared to single source assessments is more fair, accurate, credible and motivational. He also pointed out that while only 5-25% of people in large organisations were satisfied with traditional performance assessment, this improves to nearly 80% with 360° assessments. Beside these benefits, Edwards (1998) stated that 360° feedback also improves the following aspects of learning: motivation; diagnostics; employability; self-knowledge; self-development; the avoidance of derailment and the acceleration of learning. Corporate culture is also improved through the following aspects: aligning values; opening up communication; continuous learning; learning "how" and "what" counts; customer service; team-building and productivity (cf. Tomow & London, 1998).

However, performance assessments are not without any complexities. Keeping and Levy (1998) identified a number of issues of concern in the appraisal field. These are:

- This area is fraught with inconsistencies in terms of the measurement of appraisal attitudes.
- Measurement has seemingly been conducted without a theoretical basis.
- Many researchers have developed idiosyncratic measures of what appear to be the same construct, thus flooding the field with a multitude of scales.

- On a lighter note, but not less serious, the *jingle* fallacy refers to the belief that because two different measures are given the same label, they are in fact the same construct.
- On the other hand, the *jangle* fallacy refers to the belief that because two measures are given two different labels, they are in fact distinct constructs.
- Lastly, the widening gap between practitioners' practical needs and scientists' research focus.

To further add to the complexities of performance assessment there are a number of factors that contribute to variability in assessments [cf. Theron & Roodt (1999) for a more detailed discussion]:

- **Different working relationships.** According to Fund and Dobroth (1987) a job-incumbent can differ in his/her own assessments with his/her peers, managers, customers and subordinates due to the fact that he/she has different working relationships with them. Jones and Bearly (1996) suggested that flatter organisational structures result in changing management and subordinate relationships, allowing people to work more independent.
- **Ego-centric bias** is based on the premise that self-ratings are biased, while other raters share a set of common perceptions (Harris & Schaubroeck, 1988). These forms of bias can be ascribed to either defensiveness or to positive attribution.
- **Differences in organisational level.** Raters at different levels of responsibility in the organisation might weigh performance dimensions differently, based on the degree of closeness of the work relationship with the ratee (Harris & Schaubroeck, 1988).
- **Rater-leniency** is according to Thornton (1980) the stability of certain raters to rate individuals higher in comparison with other raters. Holzbach (1978) argued on a similar line by stating that leniency errors occur when ratings from different rating sources on the same ratee-group are significantly different. Holzbach (1978) and Klimosky and London (1974) found self-ratings to be more lenient than either supervisor or peer-ratings.
- **Halo effect.** The halo error is based on the premise that raters expect an employee who is performing well on one

* The authors wish to thank Prof. J. M. Schepers regarding suggestions made with the statistical analyses.

Requests for reprints should be addressed to D. Theron, Department of Human Resource Management, P.O. Box 524, Auckland Park, 2006.

performance dimension, to be just as proficient on other performance dimensions (Holzbach, 1978).

- **Purpose of ratings.** Cleveland and Murphy (1992, p. 138) asserted that *"an appraisal system that is used for one purpose may not (under similar circumstances) yield the same outcome when the appraisal system is used for a different purpose."* Harris, Smith and Champagne (1995) found administrative-based ratings to be more lenient than research-based ratings. Fahr, Cannella and Bedeian (1991) also found that ratings conducted for feedback and development purposes are less prone to leniency bias, against those that are used for reward and promotion purposes that are more prone to bias.
- **Differences in rating criteria.** An alternative view to explain inconsistencies between self and other ratings, holds that disagreement stems from a tendency of different types of raters to base their ratings upon different aspects of job performance (Klimoski & London, 1974; Lawler, 1967; Steel & Ovalle, 1984). Different rater-groups are thus using different types of criteria for ratings.

Beside these above-mentioned factors, there is also a possibility that raters may have different mental models (sets of perceptions) on what a role-incumbent should ideally be doing and what should therefore be assessed. This can be simply illustrated in the following way:

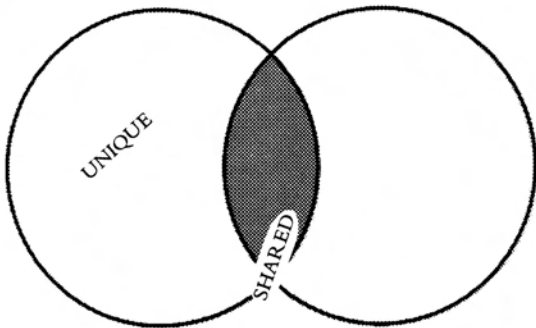


Figure 1
Unique and shared knowledge

It can be inferred from Figure 1 that there is some congruence (the shaded area) between the two sets of perceptions, but that there is also some unique knowledge that is not shared by the two parties. The same principle can be applied when there are more parties involved. In this case the picture becomes slightly more complex.

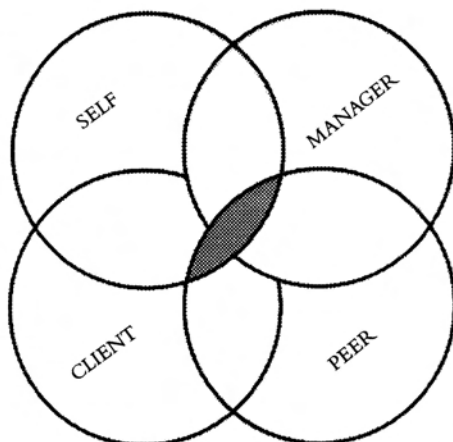


Figure 2
Unique and shared knowledge of different rater-groups

In this case there are more parties involved and each party also has its own unique knowledge. The larger the shaded area, the more agreement there is between rater-groups on the dimensions to be evaluated with a corresponding increase in inter-rater reliabilities. In the case of Figure 2 inter-rater reliabilities would most probably be low. This problem is normally addressed by getting agreement on what the rating-dimensions are, in other words increasing the shared knowledge. In this instance, Borman (1979) referred to *"frame of reference training"*. People working in different role-sets may however hold their own mental models on what they deem *"ideal"* for a particular position or role-set. Williams and Hummert (1990) were of the opinion that these performance standards are culturally defined and that people acquire the personal constructs in a process of social interaction behaviours in the organisation. In the case of this particular study and in this organisation, rater-groups may have different mental models on what they believe the *"ideal"* salesman is.

Researchers have, according to Matt (1998), failed to agree on a common definition for mental models as is illustrated by the definitions of the following three authors: Norman conceptualised mental models as the image that emerges from interactions with the system. Young suggested that mental models are metaphors used to guide interactions with and interpretations of a system. Holyoak claimed that mental models are *"psychological representations"* of a system's expected behaviours. Aronson (1997) referred to Kant who was of the opinion that all cognitive processes are model-driven. Kant's central thesis maintained that it was the mind's application of models to preconscious sense data that organises them into a cognitive unity, thereby raising sense data to one's consciousness. Here, mental models serve as rules to organise sense data. Mental model theories propose according to Langston, Kramer and Glenberg (1998) that mental models, like real situations, are played out in a medium analogous to a Euclidean space so that distance has functional consequences. Matt (1998) was of the opinion that mental models can best be studied by using hierarchical cluster analysis or multi-dimensional scaling as methods to assess their underlying structure. Mental models are for the purposes of this study, defined as unitary, spatial models where distance has functional consequence, and which are used to give meaning and understanding to complex systems or phenomena.

Against the background of the above-mentioned discussion, the following research question for this study can be formulated:

"Are the mental models of rater-groups that are used for assessing marketing competencies similar?"

It is postulated in this particular study, that although rater-groups have reached consensus on the rating dimensions, there are still differences with regard to the mental models that rater-groups hold about *"ideal"* marketing competencies.

METHOD

The research method followed in this study will be discussed under the following headings:

Participants

The population and sampling frame of this study consists of all the marketing, sales and distribution personnel (N = 128) employed by six globally operating petro-chemical business units operating in a larger organisation. All these business units have flat organisational structures and are operating with self-directed teams. Sixty-eight people responded, which resulted in a response rate of about 52%. For a more detailed description of the sample refer to Table 1.

Fifty-four percent of the respondents were male, while 46% were female. With regard to the representation of different language groups in the sample, English represented 28%, Afrikaans 60%, South Sotho four percent, Venda one percent

and the remaining categories seven percent. The ages of the respondents ranged from 21–60, the average being 37 years.

Table 1
Biographical Data of Respondents (n = 68)

GENDER	FREQUENCY	PERCENTAGE
Male	37	54
Female	31	46

HOME LANGUAGE	FREQUENCY	PERCENTAGE
Afrikaans	41	60
English	19	28
Other	4	6
South Sotho	3	4
Venda	1	1

AGE MINIMUM	MAXIMUM	MEAN	STANDARD DEVIATION
21	60	36,66	8,413

Measuring instruments

A custom designed assessment questionnaire, the Marketing Competency Questionnaire (MCQ), was used in this study. The purpose of the questionnaire was to serve as a developmental tool in identifying competency development areas. The dimensions assessed by the MCQ were agreed upon by representatives of different rater-groups and they were: technical competence; team skills; problem-solving skills; interpersonal skills; independence; customer orientation; personal drive; planning and organising. The repertory grid and critical incident interviews were used as a basis for the development of the questionnaire.

Each item in the questionnaire states a behavioural output that specifies high performance behaviour related to a particular dimension. Raters were expected to evaluate both the importance and the actual level of performance of the ratee in terms of the specified output. Responses were recorded on a 6-point scale of which only the first and last numbers were defined. A value of one refers to "no importance" and a value of six to "critical importance".

Research procedure

The MCQ was computerised and transferred to discs. Each ratee received his own evaluation disc at a group session and was given a choice of various raters. They had to choose raters with whom they had a high level of interaction and besides evaluating themselves they had to include at least one manager, one peer, and one customer to evaluate them. After completion, the ratee had to forward it to the different raters of his choice. On opening the questionnaire a message box appeared stating that the evaluation is for development purposes and that the rater's honest feedback would be appreciated.

Moderating influences in this study were limited by giving attention to the following:

- designing a job-specific questionnaire relevant to all the incumbents, but also with their inputs;
- allowing each incumbent, together with his line-manager to compile a list of raters with whom they interacted frequently to maximise valid feedback;
- using real incumbents in real marketing jobs, evaluated by actual managers, peers and customers;
- by making it clear to all the rater-groups that the assessment results would only be used for development purposes; and
- by defining the evaluative criteria for each dimension measured to prevent raters from generalising and thus limiting halo effects.

RESULTS

All respondents without a complete set of assessors (self, peer, manager, and customer assessments) were excluded from the

sample for analysis purposes. Only questionnaires that were completed fully were used in the statistical analyses.

A procedure for factor analysis as suggested by Schepers (1992) was applied in analysing the data of both the performance and the importance ratings. The scores of the thirty items of the MCQ were inter-correlated and subjected to a principal factor analysis followed by a varimax rotation. The first factor analyses identified a number of factors which differed for the various rater-groups; the importance ratings varied between 5–7 factors and the performance ratings between 6–8 factors (these results are displayed in Table 3–10). Simplified factor scores were compiled for each rater-group and these were again intercorrelated and subjected to a principal factor analysis and an oblique rotation that yielded single factors with high internal consistencies. All thirty items were included in these scales, but in a different order.

Subsequently, an iterative item analysis procedure was conducted for each of the scales (managers, peers, self and customer), using the NP-50 program of the NIPR, and high internal consistencies (Cronbach Alphas) were obtained. The resulting scales included all thirty of the original items on the importance and performance ratings (Refer to Table 2).

Table 2
Internal Consistencies (Cronbach Alpha) of the Scales

	Importance Scale	Performance Scale
Manager assessment	0,943	0,941
Peer assessment	0,956	0,963
Self assessment	0,950	0,925
Customer assessment	0,954	0,954

The scales yielded Alpha coefficients ranging between 0,925 and 0,963. This implies a fairly high degree of consistency among the raters of a particular rater-group.

The second part of the data analyses focuses on the results of the first-order factor analysis. In this section, one can ascertain which items were loading on a particular factor for the different rater-groups. Table 3 and 4 refer to the importance and the performance assessments from a manager's perspective, respectively.

Table 3
First-order factor analysis (PFA) – Manager importance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Item 3	Item 1	Item 14	Item 4	Item 5	Item 19
Item 9	Item 2	Item 21	Item 7	Item 6	Item 28
Item 10	Item 18	Item 24	Item 20	Item 8	
Item 11	Item 23	Item 25	Item 22		
Item 12	Item 26				
Item 13	Item 27				
Item 15					
Item 16					
Item 17					
Item 29					
Item 30					

Table 4
First-order factor analysis (PFA) – manager performance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Item 5	Item 1	Item 9	Item 21	Item 10	Item 16	Item 4
Item 6	Item 2	Item 13	Item 22	Item 11	Item 20	Item 7
Item 8	Item 3	Item 29	Item 24	Item 17		
Item 14	Item 12			Item 25		
Item 15	Item 19			Item 30		
Item 18	Item 26					
Item 23	Item 27					
	Item 28					

It is clear from the two tables that they differ with regard to the number of factors that were extracted (six and seven respectively). It further shows that different items are loading on similar numbered factors of the two scales.

Tables 5 and 6 refer to the importance and performance assessments from a peer's perspective, respectively.

Table 5

First-order factor analysis (PFA) – peer importance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Item 2	Item 8	Item 10	Item 1	Item 11	Item 4	Item 6
Item 14	Item 20	Item 13	Item 3	Item 29	Item 7	Item 12
Item 18	Item 21	Item 15	Item 5	Item 30	Item 17	
Item 23	Item 22	Item 16	Item 9			
Item 25	Item 24	Item 19				
Item 26	Item 28					
Item 27						

Table 6

First-order factor analysis (PFA) – peer performance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Item 1	Item 3	Item 10	Item 15	Item 7	Item 4
Item 2	Item 19	Item 11	Item 16	Item 14	
Item 5	Item 20	Item 12	Item 27	Item 17	
Item 8	Item 21	Item 13			
Item 9	Item 22				
Item 18	Item 28				
Item 23	Item 29				
Item 24	Item 30				
Item 25					
Item 26					

One can infer from the two tables above that they differ with regard to the number of factors that were extracted (seven and six respectively) and that different items are loading on similar numbered factors of the two scales. It is also clear that there are differences between the factor structures of peers and managers.

Tables 7 and 8 refer to the importance and performance assessments from the ratee's own perspective, respectively.

Table 7

First-order factor analysis (PFA) – self importance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Item 9	Item 7	Item 2	Item 21	Item 1
Item 10	Item 8	Item 5	Item 22	Item 3
Item 11	Item 15	Item 6	Item 23	Item 5
Item 12	Item 16	Item 14	Item 27	
Item 13	Item 17	Item 18		
Item 20	Item 19	Item 26		
Item 25	Item 24			
Item 29	Item 28			
Item 30				

Table 8

First-order factor analysis (PFA) – self performance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Item 1	Item 7	Item 3	Item 15	Item 2	Item 29	Item 6	Item 20
Item 8	Item 9	Item 18	Item 16	Item 4	Item 30	Item 25	Item 21
Item 12	Item 10	Item 23	Item 24	Item 5			Item 22
Item 14	Item 11	Item 27		Item 17			
Item 26	Item 13			Item 28			
	Item 19						

From these tables one can infer that there is a difference with regard to the number of factors that were extracted (five and eight respectively) and also that different items are loading on the first five factors of the two scales. It is also clear that the factor structure of these two scales differ from the peers' and the managers' scales.

Table 9 and 10 refer to the importance and performance assessments from a customer's perspective, respectively.

Table 9

First-order factor analysis (PFA) – customer importance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7
Item 1	Item 7	Item 5	Item 15	Item 14	Item 4	Item 3
Item 2	Item 8	Item 6	Item 17	Item 29		
Item 16	Item 10	Item 9	Item 25	Item 30		
Item 18	Item 13	Item 11	Item 28			
Item 21	Item 19	Item 12				
Item 22	Item 20					
Item 23						
Item 24						
Item 26						
Item 27						

Table 10

First-order factor analysis (PFA) – customer performance ratings

Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Item 3	Item 5	Item 20	Item 1	Item 4	Item 16
Item 7	Item 6	Item 21	Item 2	Item 24	
Item 8	Item 14	Item 22	Item 25		
Item 9	Item 18	Item 29	Item 26		
Item 10	Item 23	Item 30			
Item 11	Item 27				
Item 12					
Item 13					
Item 15					
Item 17					
Item 19					
Item 28					

The two tables above differ with regard to the number of factors that were extracted (seven and six respectively) and also with regard to the items that are loading on the first six factors. Furthermore, they also differ with regard to the factor structures of the other scales for managers, peers and the self.

In the third section of the data analyses, two inter-battery factor analyses were jointly conducted for the importance and the performance rating scales of the different rater-groups, respectively. Table 11 and 12 display the scale loads of the different rater-groups for the importance and performance assessments, respectively.

Table 11

Inter-battery factor analysis on second-order factor importance ratings

Rater-groups	Scale loads
Manager	0,446
Peer	0,753
Self	0,649
Customer	0,492

Table 11 clearly shows that the rater-groups' importance scales are loading on a single factor, but in varying degrees, ranging from 0,446 for the managers to 0,753 for the peers. The peer's

rating-scale clearly has the highest congruence with the inter-battery factor.

Table 12
Inter-battery factor analysis on second-order performance ratings

Rater-groups	Scale loads
Manager	0,545
Peer	0,811
Self	0,517
Customer	0,606

Table 12 clearly shows that the rater-groups' performance scales are loading on a single factor, but in varying degrees, ranging from 0,517 for the self to 0,811 for the peers. In this instance the peer's rating-scale also has the highest congruence with the inter-battery factor.

DISCUSSION

If the Alpha coefficients of the different scales (importance and performance) for the various rater-groups are superficially viewed, one may conclude that these scales are sufficiently reliable (based on internal consistency coefficients) and valid [based on the factorial validity; a form of construct validity (*cf.* Allen & Yen, 1979)]. A closer scrutiny of the factor structures of these scales, however, leads one to believe that they are measuring different dimensions, although these scales were compiled on a consensus basis.

In a related study on the same sample by Theron and Roodt (1999) it was found that the scores of some of the rater-groups were statistically significantly different with regard to the performance and the importance ratings. Power tests further indicated that these differences were practically significant. In this particular study the focus was on possible causes for those differences and on the different mental models raters are holding about what they perceive as the "ideal" marketing competencies.

The results of the first-order factor analyses (displayed in Tables 3-10) indicate that the factor structures differ for the different rater-groups with regard to the importance and the performance assessments. Furthermore, there was also a large variation in the items that were loading on particular factors for a specific rater-group. These findings give some indication that the various rater-groups have different views on the dimensions of marketing competencies.

These above-mentioned conclusions were further supported by the results of the two inter-battery factor analyses (Tables 11 and 12) which clearly show that the scales of the different rater-groups have varying loadings on the "common" inter-battery factor. If one refers back to Figure 2, one can conclude that the rater-groups share varying degrees of congruence with the common mental model (the shaded area). In both instances, the rating-scores of the peers have the highest loadings on the inter-battery factor. These findings suggest that rater-groups hold different mental models on marketing competencies. Therefore, mental models of rater-groups should be considered as a moderating variable when comparing assessment results of different rater-groups.

On a more practical note, these findings suggest that the mental models of rater-groups could be practically different from each other and that one should take note of this fact when interpreting 360° assessment results. The fairly high internal consistencies indicate a high degree of congruence within each rater-group, but the factor-structures of the different scales (importance and performance) suggest that each rater-group has a fairly unique view on the competency dimensions. By enlarging the area of shared knowledge, the differences between rater-groups could be decreased or eliminated and the overall internal-consistency coefficients could be further improved.

If Figure 1 and 2 are revisited, one can conclude that the results of this study indicate that the 'unique' areas of each rater-group have not sufficiently been drawn into the area of common knowledge or into a shared frame of reference. Although there is possibly a large pool of shared ideas, resulting in the drafting of a consensus scale for assessing marketing competencies, more can be done to include the 'unique' areas of the different rater-groups. In this regard, the use of the repertory grid technique (Epting, 1984; Stewart, Stewart & Fonda, 1981) as a method to identify the mental models of particular rater-groups rather than only the shared areas of knowledge, can potentially be of great help.

A possible limitation of this study, is the small sample size, despite the fact that the whole population of marketing personnel in this firm was included in the initial survey. Studies conducted on a wider base (including more organisations) may provide an avenue to address this issue, but poses the danger of including many other moderating factors related to differences in organisational structures and processes.

Suggestions for future research are to explore the real nature and structure of the mental models of the various rater-groups, as well as differences in this respect, by making use of hierarchical factor analysis and multi-dimensional scaling as was suggested by Matt (1998).

REFERENCES

- Allen, M.J. & Yen, W.M. (1979). *Introduction to Measurement Theory*. Monterey, CA.: Brooks/Cole Publishing Company.
- Aronson, J.L. (1997). Mental models and deduction. *American Behavioral Scientist*, 40(6), 782-798.
- Borman, W.C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, 64, 410-421.
- Cleveland, J. & Murphy, K.R. (1992). Analyzing performance appraisal as goal-directed behavior. In Ferris, G. & Rowland, K.R. (Eds.) *Research in Personnel and Human Resources Management*. Greenwich, CT: JAI Press. (pp. 121-185).
- Dalton, M.A. (1998). Using 360-degree feedback successfully. *Leadership in Action*, 18(1), 2-11.
- Edwards, M.R. (1998). *The future of 360° technologies: Power tools to accelerate workplace learning*. Paper presented at the ASTD International Conference, 3 June, San Francisco, CA.
- Epting, F.R. (1984). *Personal Construct Counseling and Psychotherapy*. New York: John Wiley & Sons.
- Fahr, J., Cannella, A.A. & Bedeian, A.G. (1991). Peer ratings the impact of purpose on rating quality and user acceptance. *Group and Organization Studies*, 16(4), 367-386.
- Funder, D.C. & Dobroth, K.M. (1987). Differences between traits: properties associated with interjudge agreement. *Journal of Personality and Social Psychology*, 52, 409-418.
- Goodale, J.G. (1992). Improve employee performance reviews. *Hydrocarbon Processing*, November, 231-232, 244.
- Harris, M.H. & Schaubroeck, J. (1988). A meta-analysis of supervisor, self-peer and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.
- Harris, M.H., Smith, D.E. & Champagne, D. (1995). A field study of performance appraisal purpose: Research- versus Administrative-based ratings. *Personnel Psychology*, 48, 151-160.
- Holzbach, R.L. (1978). Rater bias in performance ratings: superior, self- and peer ratings. *Journal of Applied Psychology*, 63(5), 579-588.
- Jones, E.J. & Bearly, W.L. (1996). *360° feedback: strategies, tactics and techniques for developing leaders*. Amherst, MA.: MRD.
- Keeping, L.M. & Levy, P.E. (1998). *Performance Appraisal Attitudes: What are we really measuring?* Paper presented at the XIII Annual Conference of the Society for Industrial and Organisational Psychology, April, Dallas, TX.

- Klimosky, R.J. & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology*, 59, 445-451
- Langston, W., Kramer, D.C. & Glenberg, A.M. (1998). The representation of space in mental models derived from text. *Memory and Cognition*, May. (no page numbers).
- Lawler, E.E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, 51(5), 369-381.
- Matt, L.J. (1998). *Using multi-dimensional scaling and cluster analysis to assess Mental Models*. Paper presented at the XIII Annual Conference of the Society of Industrial and Organisational Psychology, April, Dallas, TX.
- Shepers, J.M. (1992). *Toetskonstruksie: Teorie en Praktyk*. Johannesburg: RAU-drukkers.
- Stewart, V., Stewart, A., & Fonda, N. (1981). *Business Applications of Repertory Grid*. London: McGraw-Hill Book Company.
- Steele, R.P. & Ovalle, N.K. (1984). Self-appraisal based on supervisory feedback. *Personnel Psychology*, 37, 667-685.
- Theron, D., & Roodt, G. (1999). Variability in multi-rater competency assessments. *Journal of Industrial Psychology*, 25(2), 21-27.
- Thornton, C.T. III (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, 33, 263-271.
- Tornow, W.W. & London, M. (1998). 360-Degree feedback: The leadership challenges. *Leadership in Action*, 18(1), 1,12,13.
- Williams, S.L. & Hummert, M.L. (1990). Evaluating performance appraisal instrument dimensions using construct analysis. *The Journal of Business Communication*, 27(2), 117-134.