

AN EVALUATION OF THE 360° PROJECT MANAGEMENT COMPETENCY ASSESSMENT QUESTIONNAIRE

D THERON

Human Resources, SASOL Technology

G ROODT

*Department of Human Resources Management
Rand Afrikaans University*

ABSTRACT

The primary purpose of this study was to evaluate a 360° project management competency questionnaire relevant to a chemical engineering environment. The competency questionnaire was developed using the input of the employees who took part in the appraisal. The secondary purpose of this study was to determine if significant differences existed between the multi-rater competency evaluations of different rater groups. Eighty technically qualified employees within a technology development environment were each evaluated by a number of raters, including themselves, their managers, customers and peers. In the case of both the Importance and the Performance Scales, single factors were extracted with internal reliabilities of 0,943 and 0,941 respectively. No significant differences were obtained on paired *t*-tests between the various rater groups. These findings and their implications are further discussed.

OPSOMMING

Die primêre doel van hierdie studie was om 'n 360° projekbestuurbevoegdheidsvraelys binne 'n chemiese ingenieurswese-omgewing te evalueer. Die bevoegdheidsvraelys is ontwikkel deur gebruik te maak van die insette van die werknemers wat deelgeneem het aan die beoordeling. Die sekondêre doel van die studie was om te bepaal of daar beduidende verskille bestaan tussen die multibeoordelaarbevoegdheidsevalueringe van verskillende beoordeelaarsgroepe. Tagtig tegniese gekwalifiseerde werknemers binne 'n tegnologie-ontwikkelingsafdeling is elk beoordeel deur 'n aantal beoordeelaars wat die werknemers self, hul bestuurders, kliënte en kollegas ingesluit het. In die geval van beide die Belangrikheid- en die Prestasieskaal is een faktor onttrek met interne betroubaarheid van onderskeidelik 0,943 en 0,941. Geen beduidende verskille op gepaarde *t*-toetse kon tussen die onderskeie beoordeelaarsgroepe gevind word nie. Hierdie bevindinge en die implikasies word verder bespreek.

Multi-rater assessment tools (360° degree assessment tools) have become increasingly popular as a personal development tool and as an appraisal/compensation tool. The multi-rater process is now being used in many major organisations as an integral part of their management processes (Bracken, 1994). The purpose of this process is to provide accurate feedback, information on critical behaviours for success, and direction for individual development (Morical, 1999). A 360° system collects performance assessments from supervisors, peers, clients, subordinates and the ratees themselves (Wells, 1999). These assessments are compiled into a feedback report that provides the ratee with comprehensive feedback on his or her performance or competence. Multi-rater assessment thus fulfils the need for providing the individual with a more holistic and useful set of feedback criteria, which can greatly facilitate development (Theron & Roodt, 1999). Yammarino and Atwater (1997) indicated that the feedback received from a 360° assessment increases the accuracy of self-perceptions. It could also be predicted that by using the feedback obtained from the 360° assessment, perceptions of the various rater groups could be aligned by discussing the differences in ratings. This process can thus facilitate the clarification of expectations that could lead to aligned frames of references and, consequently, aligned ratings.

In research and in practice, there is still a lot to be learned about 360° assessment and how to ensure that such a system is successfully developed and implemented. Comprehensive research is currently being conducted on this topic. In this article, the most important research findings on 360° assessment to date are analysed.

According to Morical (1999) the success of 360° systems is a result of (1) proper integration of HR (human resources) systems and business strategy, (2) valid content, (3) accurate responses, (4) clear and specific results, (5) career development and planning support and (6) measurable improvement.

Proper integration of HR systems

Integrated HR systems are most successful. These would include a business-driven competency model, recruitment criteria, succession planning, development planning, training, follow-up and progress monitoring. Any multi-rater system needs to be integrated with other HR systems to ensure alignment of HR activities and face validity. It has to be perceived to add value.

Valid content

The 360° assessment should be integrated with business strategy and should thus measure competencies that truly affect business effectiveness. The validity of what is measured by the 360° performance or competency questionnaire, greatly influences the acceptance and thus the success of the results obtained. The issue of validity will be examined in more detail.

Accurate responses

To obtain accurate responses and avoid raters from simply completing the questionnaire for the sake of getting it done, it is important to find the correct balance between a short questionnaire, and a comprehensive and descriptive one. Shorter questionnaires reduce respondent fatigue and enhance participation. A too comprehensive questionnaire can seriously affect the validity and reliability of the results of a multi-rater assessment.

Clear and specific results

The results of the 360° assessment must be clearly communicated in some form of graphic presentation. It must present the intended meaning and should not be vague and difficult to interpret. Written comments can greatly contribute to the clarity of graphic results.

Career development and planning support

A system that provides specific development recommendations can greatly facilitate individual performance development.

Measurable performance improvement

The purpose of any 360° assessment is to improve business results. The dimensions or competencies measured by the sys-

tem should have a direct link with business performance. Competency identification tools, such as the repertory grid, derive competencies from examples of high performance.

Considering the above-mentioned success factors of implementing a 360° assessment system, it is crucial to ensure that a valid questionnaire is designed. Three main factors impact on the validity of a questionnaire, namely design, implementation and evaluation factors.

Design factors

Edwards and Ewen (1996) were of the opinion that user perception of the fairness of the system is important. One way to ensure that the system is perceived to be fair, is to allow users of the system to participate in its design.

The process used to identify the competencies to be measured is of the utmost importance, because the contents of the questionnaire should accurately describe high performance behaviour. The two most commonly used methods of identifying competencies are the repertory grid and behavioural event (critical incident) interviews (Boam & Sparrow, 1992). Both these techniques use examples of excellent performance to identify the competencies needed to ensure top performance. The identified competencies thus relate to what a person needs to know and do to achieve a high level of performance. Bracken (1994) added that the competencies covered in the questionnaire should (1) relate to the ratee, (2) cover topics that the ratee can control, and (3) relate to behaviours that the ratee can act on and that can easily be translated into a development plan.

Once the questionnaire has been designed, the implementation that follows should also be done with care.

Implementation factors

Edwards and Ewen (1996) stated that users of the system should be properly trained in using the system. Proper user training can, to a large extent, contribute to the validity of the results obtained in the assessment. It contributes to uniform understanding of the dimensions measured by the system and uniform understanding of the scale used.

A second very important factor is the user-friendliness of the system. Bracken (1994) indicated that if a questionnaire is too long or too difficult to complete, the data obtained from such an assessment is of no or little value.

Edwards and Ewen (1996, p. 20) summarised this by stating that to successfully develop and implement the system, the following features should be included:

- (1) the process design should be done by those who will use the system;
- (2) a valid process should be used for identifying the competencies;
- (3) a valid method should be used for choosing raters;
- (4) anonymity of raters must be ensured;
- (5) all participants should be trained in using the system; and
- (6) the process should be assessed by the users for fairness, accuracy and validity.

Evaluation factors

The Employment Equity Act (EEA) (No. 55 of 1998) requires all psychological tests and other similar assessments to be valid and reliable, fair, as well as not biased against any employee or any specific group of employees. In order to comply with the EEA, 360° assessments should also be evaluated and validated. After the first round of assessments, it is recommended that a sample be gathered from the group of people that was subjected to the system to assess the success of the assessment effort (Nowack & Hartley, 1999). They can assess the overall assessment effort by focusing on different dimensions. Evaluation dimensions may include:

- the validity of the items (assessing content and face validity);

- the process followed in data collection (procedural reliability); and
- the way in which the data has been analysed (appropriate, valid and reliable analysis procedures).

Improving on the system after every assessment by answering and addressing the above-mentioned issues, can have a major effect on the continuously improved quality of results obtained from each assessment.

Once a questionnaire has been validated, it is no guarantee that the raters won't differ in their assessment of the same person. Cheung (1999) mentioned that differences in multi-rater assessments are becoming a central concern in performance appraisal research. Multi-rater research indicates conflicting results with regard to variability in multi-rater assessment. Some studies found similarities between multi-raters, while other studies found differences between the assessment of multi-raters (*cf.* Theron & Roodt, 1999, 2000). These authors presented various reasons for variability in assessment scores.

Previous research mainly focused on the correlation between manager assessments and self-assessments. One seemingly common result is that employees assess themselves higher than other assessors (Bradley, 1978; Snyder, Stephan & Rosenfield, 1976).

When comparing self-assessments with supervisor assessments, Williams and Seiler (1973) found a high average self-supervisor correlation of 0,60. Pym and Auld (1965) found a self-supervisor correlation of 0,56 across three independent studies. Baruch (1996) reported that studies conducted in the United Kingdom and in Israel found a congruence of 0,73 and 0,81 respectively, when comparing self-performance appraisal results with direct-manager appraisals. These appraisals were used for individual employee development. These studies all found relatively high correlations.

However, these research findings are inconclusive. Klimoski and London (1974) reported an average self-supervisor correlation of 0,05 and Ferris, Yates, Gilmore and Rowland (1985) reported a self-supervisor correlation of 0,02. These correlations are very low. Research shows conflicting results regarding the extent of self-supervisor agreement in performance assessment.

Examining variability in peer and self-ratings, Harris and Schaubroeck (1988) found that most studies provide inconsistent findings. They suggest two reasons that might explain the inconsistency in ratings. The first reason relates to sampling error. Most of the studies are based on small sample sizes ($N = 71-80$). A second reason for the inconsistency lies in the nature of the job. It seems that blue-collar and service jobs have a higher consistency, based on the fact that these jobs are relatively routine and performance is well defined, compared to managerial and professional jobs with low consistency as these jobs cannot be easily defined.

Heneman (1980) and Bracken (1994) indicated that the reasons for discrepancies between manager, self, and peer ratings are not known, and that a theoretical foundation still needs to be developed to support such research. Yammarino and Atwater (1997) did, however, identify a couple of factors that may explain differences in ratings. These include (1) biodata, (2) individual characteristics, (3) job-relevant experiences, (4) cognitive processes, and (5) context/situation. *Biodata* refers to biographical factors such as gender, age, educational level, level in the organisation, etc. It relates to stereotyping as a result of any biographical differences between raters and the ratee. According to Yammarino and Atwater (1997) *individual characteristics*, including interpersonal orientation, locus of control, intelligence, analytical ability, levels of self-esteem and many others, can influence ratings. People tend to assess others against their own standards. *Job-relevant experiences* can directly or indirectly influence assessment data. Past successes or failures on the job tend to predict future ratings. People's *cognitive*

processes, i.e. how they gather, process, store, retrieve and use information, including their attitudes, beliefs and frames of references, can influence the way they rate themselves and others. Again, raters tend to use their own preference as a standard against which they rate others. Lastly, *contextual factors* can also play a role. Contextual factors such as familiarity with the respondent, job pressures and political processes can contribute to the differences in ratings.

Theron and Roodt (1999) did a study on marketing employees, testing the differences between the rater groups (managers, peers, clients and self-assessments). In that study, the results of the first implementation of a newly developed questionnaire were statistically examined. The results yielded statistically significant differences between all the rater groups. Theron and Roodt (2000) also found different factor structures on ratings for the different rater groups, which implies that they have used different mental models as a frame of reference for interpreting the questionnaire.

This study has now been repeated, with the exception that within this study, the questionnaire was developed three years ago and the results of the third application of this questionnaire were used for the analyses.

The lack of a coherent theory on multi-rater assessment is still very evident and clearly indicates that further research is crucial. One has to take into consideration that this study builds on previous research (Theron & Roodt 1999, 2000), and that knowledge and experience of a system might have an impact on the differences between rater groups. Consequently, the following research question was formulated.

“Are there significant differences between multi-rater competency assessments by different rater groups?”

The aim of this study is, firstly, to evaluate the metric properties of the 360° project management competency questionnaire and, secondly, to determine if significant differences exist between the competency assessments of different rater groups, i.e. managers, peers, customers and self-assessments.

The dramatic flattening of organisational structures poses certain challenges in terms of performance appraisal methods. It is clear that the manager is the best possible evaluator in a hierarchical structure with strict reporting systems and controls (Jones & Bearley, 1996). However, where employees operate in teams within a flat structure, and where individual responsibility is toward team members, there is limited contact between managers and their subordinates. Flattening structures might have the effect that appraisals by team members (peers) and customers should be considered in addition to managerial assessments. It implies using a more objective means of performance measurement that is in line with organisational changes. Lawler (1967) states that the multi-rater approach to performance appraisal has not received a lot of attention, but appears to have the advantage of being a more objective measure of performance.

The following hypothesis is formulated and supporting literature is quoted:

H₁ Statistically significant differences exist between the different rater groups (i.e. managers, peers, customers and self).

The only relatively consistent finding in multi-rater research indicates that employees assess themselves consistently higher than other assessors (Bradley, 1978; Snyder et al., 1976; Theron & Roodt, 1999; Thornton, 1980).

Baruch (1996), Pym and Auld (1965) and Williams and Seiler (1973) found relatively high correlations between self-performance appraisals and direct-manager appraisals. The sample, however, consisted of employees from government departments such as the navy, where a strict hierarchical structure exists.

In contrast, Ferris et al., (1985) and Klimoski and London (1974) found very low self-supervisor correlations. These studies, however, did not provide any indication of the moderating effect of organisational structure.

Theron and Roodt (1999) reported that significant differences existed between the ratings of managers and customers. They deduced that, due to a higher level of contact between the customer and the incumbent, compared with a lower level of contact between the manager and the incumbent, manager and customer ratings will differ significantly. Their research results supported this notion. The above differences in ratings could be a result of flat organisational structures.

Lawler (1967) indicated that manager ratings are used twice as frequently as peer ratings. However, he described peer ratings as being far more objective than manager ratings. Springer (1953) found manager ratings to be more conservative than those of peers.

Furthermore, Barclay and Harland (1995) concluded that peer raters were perceived as fair if they were educated and experienced. Although a lot of research has been done on fairness and accuracy of multi-rater assessments, limited research has been done on differences in assessments of multi-raters. Also, little evidence is available on the effect of organisational structure.

Theron and Roodt (1999) found no significant differences between customer ratings and peer ratings. They concluded that in a company with a flat organisational structure, customer and peer ratings would be in agreement, provided both parties know the incumbent and his/her job and interacts with him/her on a frequent basis.

METHOD

Participants

The population consists of technically qualified people in an engineering environment (N=100) employed by a petrochemical technology development company. These employees function in a flat organisational structure and are operating in self-directed cross-functional matrix teams. The sampling frame included 100 tertiary qualified technical employees, of which 80 employees responded, which established a response rate of 80% (Table 1).

TABLE 1
BIOGRAPHICAL DATA OF RESPONDENTS (n = 80)

GENDER	FREQUENCY	PERCENTAGE
Male	59	74%
Female	21	26%
HOME LANGUAGE	FREQUENCY	PERCENTAGE
White	73	91%
Asian	2	2,5%
Black	5	6,5%

Seventy-four percent of respondents were male, while 26% were female. With regard to the representation of different race groups in the sample, whites represented 91%, Asians 2,5% and blacks 6,5%.

Measuring instrument

The Project Management Competency Questionnaire (PMQCQ) that was used for this study is a custom-designed assessment questionnaire. The purpose of the questionnaire is to serve as a developmental tool that can aid in identifying competency developmental areas. The dimensions measured by the questionnaire include technical competence, problem-solving skills, personal drive, innovation, personal impact, interpersonal skills, team skills and leadership skills.

Each item in the questionnaire states a behavioural output that specifies high performance behaviour related to the dimension (competency). Raters are expected to evaluate both the level of importance of a competency and the level of performance of a ratee in terms of the specified output. Responses were recorded on a 6-point scale of which the first number and the last number are defined. A value of one in terms of importance refers to no importance and a value of six refers to critical importance. A value of one in terms of performance refers to extremely poor performance and value of six refers to excellent performance.

The questionnaire was developed using the repertory grid and behavioural event (critical incident) interviews as basis (Boam & Sparrow, 1992). Using the information gathered with the behavioural event interviews and the repertory grids, a draft questionnaire was compiled. This questionnaire was tested with a sample of technical professionals to identify ambiguities, omissions or related problems. Questions were added, deleted and reworded, based on the respondents' feedback. With the input of the sample, the layout of the PMCQ, as well as the process of assessment were finalised. The first implementation took place in 1998. At that stage the questionnaire consisted of 31 items. After the first assessment, a sample of technical professionals revisited the contents of the questionnaire. Questions were again added, deleted and reworded. The questionnaire was finally reduced to 24 items. A second assessment took place in 1999, after which this process was repeated. This time no items were added or omitted, but certain items were reworded. The data of the third assessment that took place in 2000 were used in this study. Reliability coefficients of the questionnaire are reported under the heading 'Results'.

Research procedure

Group training sessions were held to educate the users in the understanding of the content of the system and the process of choosing raters. Raters were chosen on a basis of a high level of interaction with the employee being assessed. They had to choose at least one manager, one peer and one customer to evaluate them, as well as evaluating themselves. The questionnaires were computerised and distributed via e-mail. On opening the questionnaire on e-mail, a message box appeared with a message stating that the evaluation was for development purposes and that the raters' honest feedback would be appreciated.

An attempt was made in this study to limit moderator influences. This was done through the following (Jacobs, 1989; Klimoski & London, 1974; Lawler, 1967; Steel & Ovalle, 1984):

- designing a job-specific questionnaire relevant to the incumbents, but also developed with the input of the incumbents;
- allowing each incumbent, with his or her line manager, to compile a list of raters with whom they interact frequently to ensure valid feedback;
- using real incumbents in real technical jobs, evaluated by actual managers, peers and partners;
- making it clear to all incumbents and their raters that the evaluation would be used to identify development areas, and not to determine compensation; and
- defining evaluative criteria for each dimension measured, to prevent raters from generalising, and thus limit halo effects.

RESULTS

All respondents without a complete set of assessors (including self-assessment, manager assessment, peer assessment and customer assessment) were excluded from the sample. A factor analysis procedure as suggested by Schepers (1992, pp. 140-144) was conducted for both the importance and performance ratings. All fully and properly completed questionnaires that were returned were analysed statistically. The item scores on the twenty-four items of both the importance and the performance

scales of the PMCQ were inter-correlated separately and subjected to principal factor analyses and rotated to simple structures by means of the varimax rotation. Three factors were postulated for each scale, based on eigenvalues larger than unity (Kaiser, 1961). Three simplified factor scores (SFS) were calculated for each of the scales, inter-correlated and rotated to simple structures by means of the oblique rotation. From these analyses, one factor was extracted for the importance scale, including all twenty-four items for all the different rater groups. For the performance scale, one factor was extracted including all twenty-four items (refer to Tables 2- 5).

TABLE 2
EIGENVALUES LARGER THAN UNITY, IMPORTANCE SCALE

Factor	Eigenvalue	Cumulative variance (%)
1	2,45	81,79
2	0,336	92,90
3	0,213	100,00

TABLE 3
SIMPLIFIED FACTOR SCORE (SFS) LOADINGS, IMPORTANCE SCALE

S F S Loadings	
Factor 3	0,891
Factor 1	0,882
Factor 2	0,784

TABLE 4
EIGENVALUES LARGER THAN UNITY, PERFORMANCE SCALE

Factor	Eigenvalue	Cumulative variance (%)
1	2,30	76,71
2	0,429	91,01
3	0,270	100,00

TABLE 5
SIMPLIFIED FACTOR SCORE (SFS) LOADINGS, IMPORTANCE SCALE

S F S Loadings	
Factor 3	0,862
Factor 1	0,846
Factor 2	0,714

An iterative item analysis was conducted on the importance scale that retained all 24 items. A Chronbach alpha of 0,943 was obtained after no iteration. An iterative item analysis was also conducted on the performance scale that yielded a Chronbach alpha of 0,941 with item 16 omitted after one iteration. (Refer to Tables 6 & 7.)

TABLE 6
ITEM ANALYSIS, IMPORTANCE SCALE

Items	Item mean	Standard deviation	Item reliability	Item total correlations
1	4,952	0,735	0,419	0,570
2	4,645	0,828	0,477	0,576
3	4,569	0,811	0,462	0,570
4	4,788	0,756	0,486	0,643
5	4,944	0,640	0,433	0,677
6	4,771	0,658	0,427	0,649
7	4,742	0,675	0,405	0,599
8	4,446	0,782	0,514	0,657
9	4,771	0,703	0,389	0,554
10	4,610	0,739	0,510	0,690
11	4,591	0,748	0,468	0,627
12	4,649	0,738	0,574	0,777
13	4,381	0,805	0,583	0,724
14	4,652	0,729	0,551	0,756
15	4,738	0,735	0,562	0,764
16	4,303	0,900	0,531	0,655
17	4,571	0,744	0,542	0,729
18	4,290	0,753	0,521	0,692
19	4,580	0,826	0,564	0,683
20	4,688	0,750	0,480	0,640
21	4,641	0,754	0,539	0,715
22	4,610	0,856	0,554	0,647
23	4,426	0,811	0,561	0,692
24	4,446	0,835	0,528	0,639

Chronbach alpha 0,943

TABLE 7
ITEM ANALYSIS, PERFORMANCE SCALE

Items	Item mean	Standard deviation	Item reliability	Item total correlations
1	4,595	0,814	0,557	0,684
2	4,255	0,812	0,530	0,653
3	4,301	0,775	0,525	0,677
4	4,392	0,831	0,568	0,684
5	4,502	0,803	0,551	0,686
6	4,476	0,732	0,457	0,624
7	4,671	0,793	0,472	0,596
8	4,286	0,796	0,499	0,627
9	4,690	0,902	0,611	0,678
10	4,180	0,833	0,592	0,711
11	4,069	0,825	0,511	0,619
12	4,459	0,953	0,635	0,666
13	4,028	0,810	0,517	0,638
14	4,446	0,841	0,565	0,672
15	4,491	0,795	0,509	0,640
*16	4,524	0,778	0,277	0,355
17	4,390	0,825	0,508	0,616
18	4,065	0,739	0,486	0,657
19	4,201	0,823	0,556	0,675
20	4,498	0,809	0,522	0,645
21	4,366	0,900	0,624	0,693
22	4,260	0,808	0,550	0,680
23	4,316	0,825	0,570	0,690
24	4,160	0,731	0,470	0,643

Chronbach A° alpha 0,941

*16 was omitted in the item analysis

TABLE 8
PAIRED SAMPLES t-TEST: IMPORTANCE RATING

Rater groups	Mean of difference scores	Standard Deviation of difference scores	Standard error mean	t	df	Sig.(p) 2-tailed
Manager vs. self	2,51	16,68	1,86	1,35	79	0,182
Manager vs. customer	1,73	15,56	1,74	0,99	79	0,321
Manager vs. peer	2,70	18,02	2,01	1,34	79	0,184
Self vs. peer	-4E-02	15,13	1,47	-0,02	105	0,980
Customer vs. peer	0,52	17,70	1,84	0,28	92	0,779
Self vs. customer	0,53	17,07	1,77	0,30	92	0,767

TABLE 9
PAIRED SAMPLES t-TEST: PERFORMANCE RATING

Rater groups	Mean of difference scores	Standard Deviation of difference scores	Standard error mean	t	df	Sig.(p) 2-tailed
Manager vs. self	2,58	18,43	2,06	1,255	79	0,213
Manager vs. customer	-0,20	17,55	1,96	-0,102	79	0,919
Manager vs. peer	0,60	17,88	2,00	0,300	79	0,765
Self vs. peer	-2,29	16,27	1,77	-1,632	105	0,106
Customer vs. peer	-3,00	18,96	1,96	-1,526	92	0,130
Self vs. customer	-0,12	18,95	1,96	-0,060	92	0,952

The stated hypothesis proposed statistically significant differences between the different rater groups (managers, peers, clients and self-assessments). Paired *t*-test results indicate that no statistically significant differences were identified between any of the rater groups on both the importance rating and the performance ratings.

DISCUSSION

In this study an attempt was made to evaluate the 360° PMCQ and to identify statistically significant differences between the competency assessments of multi-raters functioning within a flat organisational structure. The PMCQ yielded high internal consistencies within the various raters groups, both on actual performance and the level of importance of expected performance. This questionnaire can possibly be used by other researchers in similar professional technical environments as a reliable instrument for assessing project management competencies. It can also be concluded that when the purpose of assessment is stated as development, relatively higher reliability and validity coefficients of ratings are obtained (Fahr, Cannella & Bedeian, 1991).

In this study, in order to improve the reliability of the questionnaire, an attempt was made to reduce rater errors and improve rater agreement. One method applied was to clearly define the rating dimensions in order to establish a common frame of reference in the rating instructions. This frame of reference reduced the ambiguity in the rating situation. It decreased the likelihood for raters to resort to improper responses (rater leniency or halo) as a means of compensating for lack of task clarity.

All the rater groups, with the exception of customers, were involved in the development of the questionnaire. They agreed that the questions that were finally included in the questionnaire were of critical importance for every individual functioning in a professional-technical environment. As a result it was expected that the rater groups would have agreement on the importance of the items.

Paired *t*-tests for dependent samples were conducted on the importance and performance scales to establish possible differences between the different rater groups (Table 9).

It was hypothesised that no statistically significant differences would exist between the different rater groups. Using paired *t*-tests, it was determined that no significant differences exist between any of the rater groups. Theron and Roodt (1999) did a similar study in the company, using a marketing competency questionnaire. It was found that significant differences existed between the different rater groups when referring to the performance rating. Both questionnaires were designed in exactly the same way. The only difference was that the results of the first assessment were used with the marketing employees. In this particular study, the results of the third assessment were used. Yammarino and Atwater (1997) concluded that the feedback received from a 360° assessment increases the accuracy of self-perceptions. It can be inferred that, due to the feedback received from the two previous assessments, the raters had ample time to align their own perceptions with the expectations and perceptions of the other rater groups.

Theron and Roodt (1999) highlighted quite a few uncertainties surrounding 360° assessment. They indicated that Thornton (1980, p. 269) was of the opinion that self-appraisals should be used carefully. Consistent with previous research, it was indicated that self-assessments are inflated (Bradley, 1978; Snyder, Stephan & Rosenfield, 1976; Thornton, 1980; Roodt & Theron, 1999). When managers are using self-appraisals of their subordinates, they should be aware of the fact that their own appraisal could differ vastly from their subordinate's self-appraisal. Springer (1953) also found manager ratings to be more conservative than those of peers. The validity of manager assessments as the only form of assessment in a flat organisation structure can be questioned considering the fact that it is significantly more conservative than peer, self and customer ratings. It does seem, however, that repeated feedback can eliminate this problem not only between managers and self-assessments, but also between all the different rater groups. After a first 360° assessment, the focus should be on clarifying the reasons for the differences in ratings. If this process is done thoroughly, significant differences between raters could be eliminated in a second or third assessment. Alignment of ratings can be established, because respondents know what is expected of them and can perform accordingly. The different rater groups also get feedback on the perceptions and standards used by other raters. This kind of feedback leads to the alignment of standards applied when an employee is being assessed. This provides ratings from the different rater groups with no significant differences between these ratings.

Possible limitations of this study that could be considered, are the relatively small sample size and the fact that this study is limited to technically qualified employees in the petrochemical industry. Further research is still needed to study the differences in rating scores between raters, but this research indicates a new factor that contributes to variability in rating scores, namely the number of times the questionnaire had been previously applied. It is not clear from previous research whether the same questionnaire, applied in second or third ratings, would lead to differences in those ratings. It seems that feedback can play a major role in eliminating assessment differences. Further research can compare ratings for development purposes and where the assessment is used for promotional purposes or remuneration increases.

REFERENCES

- Barclay, J.H. & Harland, L.K. (1995). Peer performance appraisals. *Group and Organizational Management*, 20(1), 39-60.
- Baruch, Y. (1996). Self-performance appraisal vs. direct-manager appraisal: a case of congruence. *Journal of Managerial Psychology*, 11(6), 50.
- Boam, R. & Sparrow, P. (1992). *Designing and achieving competency: A competency-based approach to developing people and organisations*. London: McGraw-Hill.
- Bracken, D.W. (1994). Straight talk about multirater feedback. *Training and Development*, September, 44-51.
- Bradley, G.W. (1978). Self-serving biases in the attribution process: a re-examination of the fact or fiction question. *Journal of Personality and Social Psychology*, 36, 56-71.
- Cheung, G.W. (1999). Multifaceted conceptions of self-other ratings disagreement. *Personnel Psychology*, 52(1), 1-36.
- Edwards, M.R. & Ewen, A.J. (1996). *360° Feedback*. New York: American Management Association.
- 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.
- Ferris, G.R., Yates, V.L., Gilmore, D.C. & Rowland, K.M. (1985). The influence of subordinate age on performance ratings and causal attributions. *Personnel Psychology*, 38, 545-557.
- Harris, M.H. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.
- Heneman, H.G. (1980). Self-assessment: a critical analysis. *Personnel Psychology*, 33, 297-300.
- Jacobs, R. (1989). Getting the measure of management competence. *Personnel Management*, June, 32-37.
- Jones, E.J. & Bearley, W.L. (1996). *360° feedback: strategies, tactics and techniques for developing leaders*. Amherst, MA: MRD.
- Kaiser, H.F. (1961). A note on Guttman's lower bound for the number of common factors. *British Journal of Statistical Psychology*, 14 (1), 1.
- Klimosky, R.J. & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology*, 59, 445-451.
- Lawler, E.E. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, 51(5), 369-381.
- Morical, K.E. (1999). A product review: 360° assessments. *Training and Development*, 53(4), 43-47.
- Nowack, K.M. & Hartley, J. (1999). How to evaluate your 360° feedback efforts. *Training and Development*, 53(4), 48-53.
- Pym, D.L.A. & Auld, H.D. (1965). The self-rating as a measure of employee satisfactoriness. *Journal of Occupational Psychology*, 39, 103-113.
- Schepers, J.M. (1992). *Toetskonstruksie: Teorie en praktyk*. Johannesburg: RAU- Drukkers.
- Snyder, M.L., Stephan, W.G. & Rosenfield, D. (1976). Egotism and attribution. *Journal of Personality and Social Psychology*, 33, 435-441.
- Springer, D. (1953). Ratings of candidates for promotion by co-workers and supervisors. *Journal of Applied Psychology*, 37, 347-351.
- 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.
- Theron, D. & Roodt, G. (2000). Mental models as moderating variable in 360-degree competency assessments. *Journal of Industrial Psychology*, 26(2), 14-19.
- Thornton, C.T. III (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, 33, 263-271.
- Wells, S.J. (1999). A new road: travelling beyond 360-degree evaluation. *HR Magazine*, 44 (9), 82-84.
- Williams, W.E. & Seiler, D.A. (1973). Relationship between measures of effort and performance. *Journal of Applied Psychology*, 57, 49-54.
- Yammarino, F.J. & Atwater, L.E. (1997). Do managers see themselves as others see them? Implications of self-other rating agreement for human resources management. *Organizational Dynamics*, 25(4), 35-36.