

PERFORMANCE APPRAISALS IN THE PUBLIC SECTOR: ARE THEY ACCURATE AND FAIR?

DJ RADEMAN
HD VOS

*Department of Industrial Psychology,
University of Stellenbosch*

ABSTRACT

Traditionally, performance appraisal has been a much maligned human resources function. As a result of criticism normally being more generalised, research was conducted regarding the extent to which differing perceptions play a role in terms of specific aspects of performance appraisal. Perceptual differences, essentially between, supervisors and subordinates in the Public Sector were compared. Of the eleven factors investigated, statistically significant differences were established with regard to fairness, ethics, accuracy, rating error, and administrative-aspects. Although the remaining factors did not produce significant perceptual differences, the results did not necessarily reflect positive sentiments. The results of the research did however contribute greater insight into performance appraisal as a critical human resources function.

OPSOMMING

Prestasiebeoordeling word tradisioneel as een van die mees gekritiseerde menslikehulpbron-bestuursfunksies beskou. Vanweë die feit dat hierdie kritiek dikwels baie veralgemeen word, is dit ten doel gestel om die mate waartoe individuele persepsies met betrekking tot sekere aspekte van prestasiebeoordeling verskil, na te vors. Meer in die besonder is die perseptuele verskille tussen toesighouers en ondergeskiktes in die openbare sektor met mekaar vergelyk. Elf afsonderlike faktore wat as relevant met betrekking tot prestasiebeoordeling geag was, is ondersoek en statisties beduidende verskille is met betrekking tot die faktore regverdigheid, etiek, akkuraatheid, beoordelingsfoute en administratiewe aspekte, gevind. Alhoewel statisties beduidende perseptuele verskille nie met betrekking tot die oorblywende faktore gevind is nie, reflekteer die resultate terselfdertyd nie werklik 'n positiewe ingesteldheid ten opsigte van prestasiebeoordeling nie. Die navorsingsresultate het nietemin tot meer insig oor die waarde en betekenis van prestasiebeoordeling as 'n kritieke menslikehulpbron-bestuursaanleentheid bygedra.

The term performance appraisal essentially describes the evaluation of people in the workplace with regard to their job performance and potential for further development. The more traditional one-to-one system of evaluation of subordinates by supervisors or managers tends to elicit mixed feelings toward the process to the extent that it has been termed the Achilles heel of human resources management (Cascio, 1987), while Folger and Lewis (1993), suggest that performance appraisals tend to create as much enthusiasm as paying taxes. Deming (1986) takes a more radical line by referring to performance appraisal as one of seven deadly sins afflicting managers. Folger and Lewis (1993) add that the process becomes a case of 'going through the motions' or a superficial exercise carried out merely because it has been prescribed as such by the organisation. Sadly this seems to be a general sentiment concerning appraisals.

Milkovich and Boudreau (1988) on the other hand, suggest that performance assessment represents one of the most important interactions which take place between supervisors and subordinates, to the extent that it can either boost or reduce the effects of other human resources management activities. Cascio (1987) expresses the same sentiments by saying that despite the intensive awareness of the difficulties involved with performance appraisal, surveys consistently show that managers of both small and large organisations are unwilling to abandon performance appraisal because they regard it as an important facet in assessing the abilities and skills of workers. In support of this statement and despite the conjecture surrounding the merits or de-merits of performance appraisal, it is interesting to note that up until 1977 between 74% and 89% of organisations in the USA had some form of formal appraisal system while in Great Britain the figure was set at around 82% (Murphy & Cleveland, 1991). Longenecker and Goff (1992) put this percentage for the USA at over 90% while Latham and Wexley (1994) put the figure at 94%. The obvious conclusion therefore, is that more and more organisations are moving towards some form of formal appraisal system.

Schuler (1981, p. 211) provides a more comprehensive definition of performance appraisal by describing it as, "a formal structured system of measuring and evaluating an employee's job-related behaviours and outcomes to discover how and why the employee is presently performing on the job and how the employee can perform more effectively in future so that the employee, the organisation and society all benefit". This definition also leads to the introduction of the participants of a performance appraisal system, namely the person (manager or supervisor) who carries out the measuring and evaluating of performance levels and who is often referred to as the rater or appraiser, and the person whose performance is being measured or evaluated (subordinate) is referred to as the ratee or appraisee. Under normal circumstances managers or supervisors are responsible for the management and evaluation of performance activities of the subordinates directly under their jurisdiction. It can be expected therefore, that the attitudes of each group towards appraisal will differ from the point of view of the judge and the judged and it is for this reason that the effectivity of the system can be influenced (Dipboye & Pontbriand, 1981). Some further negative implications of appraisal by a single rater can be as follows:

- Reliability of the process may be questionable.
- Appraisees are rated according to the singular perspective, opinion or judgement of one rater.
- Other important sources of information such as inputs from peers, other departments, customers etc., are generally not taken into account.
- With more and more emphasis being placed on fairness in the workplace, the compliance of one-to-one appraisal processes with new legislation could also be questionable.

Therefore, the increased use of multi-source appraisal processes or as London (1997) refers to it, the 360 degree approach, seems to be a more acceptable solution to the traditional criticisms and problems surrounding performance appraisal. The reason for this is that inputs come from all possible sources which in turn creates a broader base for observation through different perspectives, leading to more accurate and reliable evaluation.

Earlier researchers and authors tended to consider performance appraisal as being a facet of human resources management whereas more recently, performance appraisal is rapidly be-

coming an integral part of a wider, or more holistic concept referred to as *performance management*. Spangenberg (1994) proposes that a change to a more holistic approach is necessary in order to establish situational and organisational factors so that the system will work more effectively. He describes performance management as the management of workers, which includes planning their performance, facilitating the achievement of goals and effecting the review of performance in such a way that it is both motivational as far as the worker is concerned and in line with the objectives of the organisation. The essential difference of the two approaches is that the goals and objectives of the organisation are far more clearly stated in the case of performance management, whereby workers' activities and development are brought more into line with the achievement of these goals and objectives. A more modern approach to performance appraisal includes the multi-source or 360 degree approach as advocated by London (1997).

Having been reminded of the negative sentiments regarding appraisal, it would be pertinent at this stage to consider the origin of these sentiments. A number of researchers have identified various reasons for resistance against performance appraisal processes from both the point of view of the ratee and the rater. Schneier, Beatty and Baird (1987) for example, suggest that resistance from the point of view of the rater is manifested *firstly*, in a normal dislike for having to criticise subordinates and then having to defend their point of view which invariably develops into a conflict situation. *Secondly*, there is often a lack of skill in the handling of the appraisal interview. *Thirdly*, there is often resistance towards new procedures and the concomitant changes which they bring about especially from the point of view of evaluation measures. *Fourthly*, there is often mistrust concerning the validity of the appraisal instrument in use. Moon (1997) takes the perceived problem areas a little further by stating that appraisals take too much time, that the paperwork is complicated, that the results are never tangibly observed, that appraisals are often overly subjective, and that some appraisers who are normally well balanced, tend to become officious in the appraisal situation. He adds that appraisal should be used as a tool to help managers manage and should not only be utilised to the advantage of the organisation but more importantly, for the needs, interests, and development of employees. Philp (1990) attributes resistance toward appraisal to the lack of responsibility on the part of raters in accepting ownership of the process. From the point of view of ratees, he suggests that their attitudes are affected by suspicion of why they are being appraised, concern for the appraisal process to be fair, the subjectivity of appraisal, being evaluated in terms of personality traits instead of performance results, and the assumption by supervisors that all factors affecting performance are under the control of ratees. This latter point is supported by Spangenberg (1994). Again in this case a multi-source appraisal approach together with intensive training of raters, would go a long way to eradicating the negative perceptions of performance appraisal.

Personal involvement with performance appraisal has shown that some, or all of the factors mentioned above, presented themselves in the appraisal of employee performance. A great many colleagues and associates of the authors appeared to have similar sentiments towards performance appraisal as a particular human resources management system.

These sentiments indicated that performance appraisal is regarded as:

- too time consuming and laborious having to record, prepare and list incidents.
- too subjective in terms of appraisals by supervisors.
- too confusing in terms of having different instruments for different purposes instead of a single multi-disciplinary instrument.
- insufficiently representative and unfair in that, at review level, ratees' evaluations are ratified and finalised by persons who don't always know them and/or are not always aware of their true capabilities. The outcome of their assessment can also be influenced by the oratory capabilities or personality

- of the person presenting it before the reviewing body.
- unclear in that incidents and weightings thereof are ill-defined. A general lack of training in the administration of the system and use of the instruments is apparent.
- non-beneficial in terms of notable career progress such as promotions and merit awards.
- inadequate in terms of feedback which is experienced as being generally weak and not regular or meaningful enough to make a difference in improving performance.
- inefficient and de-motivating in that feedback on ratified appraisals takes place too long after the appraisal has been carried out, or that feedback from supervisors is inadequate to make any meaningful impact on improving performance.
- unethical in that supervisors are often prejudiced for various reasons when evaluating personnel.

Gibb (1985) deems it important for managers to relate performance planning to performance review where the linking of performance planning and appraisal enables *firstly*, the promotion of understanding and acceptance of organisational goals. *Secondly*, the articulation of problems and *thirdly*, and most importantly for the individual employee, the enhancement of his or her skills assessment and growth. This approach is very much aligned with the concept of performance management. Exploratory enquiries made among a number of employees revealed that their general perception and experience was, that if the three points mentioned did exist, they were not always apparent, nor were they communicated clearly. The core problem always is, *what* is being assessed? Whatever is assessed should necessarily be relevant concerning the effect or output of the job.

Although the planning of performance expectations, as in management by objectives, is a concept more related to positions where performance can be measured against the achievement of strategic or operational goals, there should still be some pre-set standards for those jobs which can be described as mainly routine or duty orientated. In this context, performance management looks more toward a system whereby performance is defined by business strategy, team missions, clients, the situation, the nature of the system and roles (Spangenberg, 1994).

It was the intention through this research therefore, to explore the perceptions and experiences of supervisors and subordinates with regard to specific areas of performance appraisal such as utility, fairness, ethics, motivation, accuracy, validity, rater error, effectivity and feedback, as well as administrative and developmental aspects. It would be unacceptable to take for granted that performance appraisals tend to be problematical without investigating the reasons for problems in any given situation. This would especially be the case given the different methods, systems or approaches in use and the application or interpretation of them. The re-active element of the research was therefore important, from this point of view. The pro-active element was to be able to present the findings to the organisation with recommendations as to how problems could be avoided with their contemporary system as well as with the implementation of envisaged new systems.

In essence the study attempted to determine the extent to which perceptions of the performance appraisal system in the Public Service, with regard to utility, fairness, ethics, motivation, accuracy, validity, rater error, effectivity, feedback, administrative and developmental aspects, are dependent upon the post-level of workers and whether they are considered to be "achievers" or "non-achievers". Briefly, the concept post-level, determines whether an employee is a supervisor or subordinate while "achievers" and "non-achievers" are determined through their respective job-achievement levels. This is important in terms of establishing the potential for perceptual differences toward performance appraisal. A further consideration was to contribute towards the development of a measuring instrument that is both valid and reliable for the measurement of perceptions concerning performance appraisal in a South African context, as originated by Le Roux (1989).

Having briefly mentioned the problem areas and the aims of the study, it is important to consider what makes for effective assessment in order to place the utility of performance appraisal into perspective. The underlying success of evaluation is the ability of the appraiser to observe and record behaviours accurately in terms of constituencies and criteria. According to Milkovich and Wigdor (1991) for example, the goals of appraisal are *firstly*, to create measures which accurately assess the job performance levels of individuals and *secondly*, to create an evaluation system which enhances the operational functions of the organisation. Gathering and processing information, as well as the cognitive processes in making decisions and judgements are therefore, of paramount importance. London (1997) suggests that being a skilled observer is important in terms of having the ability to understand the effects of individual characteristics, the conditions within any given situation and the effects of these situations on individuals.

By means of introduction to the hypotheses, it is necessary to take note of the nature of the appraisal instrument and the aims thereof. The appraisal instrument and the appraisal process contribute to the way assessments are made in terms of the classification of appraisal methods in use (i.e. individual or multiple), the purpose for which appraisals are being carried out, types of ratings, and the effects of environments on role players. In this particular case, the instrument (Personnel Assessment Questionnaire) in use in the participating organisation is based on a combination of critical incidents which are evaluated or scored on the basis of the graphic rating scale method. The organisation makes use of two other separate processes for the awarding of 'merits' and salary notch increments. The aims of the main process for which the Personnel Assessment Questionnaire is used however, are the following (Chapter BX/II/1, Public Service Staff Code, 1994):

- to determine the promotability of employees to, and acceptability in, higher posts.
- the consideration of transfers within and between departments (correct placement).
- to identify supervisory and/or management potential.
- to determine the needs for training and development, especially those of under-achievers.
- to decide upon corrective action.

Although the above-mentioned aims of the system are acceptable from a human resources management point of view, the sentiments of public servants in respect of these instruments and processes as recorded earlier, concern more the application of the system than the system itself. Two hypotheses were developed from this and are stated as follows:

• Main Hypothesis 1

A significant difference in perception of the performance appraisal system with regard to utility, fairness, ethics, motivation, accuracy, validity, rater error, effectivity, feedback, administrative and developmental aspects, exists between supervisors and subordinates.

• Main Hypothesis 2

A significant difference in perception of the performance appraisal system with regard to utility, fairness, ethics, motivation, accuracy, validity, rater error, effectivity, feedback, administrative and developmental aspects, exists between the subgroups of subordinates and supervisors namely, "achievers" and "non-achievers".

A composite sub-hypothesis for those factors reflecting statistical significance, namely fairness, ethics, accuracy, rater error, and administrative aspects is stated and dealt with under "Results".

METHOD

Research Participants

Although a study on the performance appraisal system in the Public Sector should ideally have been conducted in the wider Public Service, this was deemed to be impractical due to the size

of the population. It was decided therefore to focus on a conglomerate of State Departments and when the population of this conglomerate again proved to be too large in terms of the relatively limited nature of the research, it was finally decided to concentrate more specifically on centralised Head Office components.

It is believed however, that the population from which the sample was drawn, was representative of the Public Service. The reason for this is that *firstly*, experience has shown that the specific comments or criticisms concerning performance appraisal are common throughout the Public Service. *Secondly*, with the constant migration of public servants through inter-departmental and inter-provincial transfers, it can be expected that the application of the system must become more and more stereotyped and therefore be subject to similar criticisms throughout the Service. *Thirdly*, because the system is universally applicable throughout the Public Service, training methods and information concerning the administration of the system are originated and controlled from a single central source. This means that individual departments cannot deviate from the rules as set out in the Public Service Staff Code, thus leading to a situation which also lends itself toward uniformity. This last point is critical in terms of the presence of feelings or attitudes of ownership toward the system and the extent to which the supervisory or management levels would probably be prepared to defend the system or really try to make it work.

In an attempt to improve the quality of results, it was decided to distribute questionnaires to a randomly selected sample population of at least six hundred employees representing subordinates and supervisors. The questionnaires were distributed to participants through liaison personnel who were identified by the personnel component. Respondents were supplied with envelopes in which the completed questionnaires could be returned.

The overall response to the questionnaires was very satisfactory in that 431 of the 600 questionnaires were returned (almost 72%) of which 44 were unusable. Therefore, 387 questionnaires could be used of which 186 were supervisors' and 201 subordinates'. Of the 186 supervisors' responses which could be used for statistical analysis, 80 (43%) were categorised as achievers and of the 201 subordinates' responses which could be used for statistical analysis, 38 (19%) were categorised as achievers. These percentages, albeit in raw form, were interesting in terms of the relationship of achievement and career progress, i.e., a notably greater percentage of supervisors were categorised as achievers.

Measuring Instrument

The diagnostic instrument used, was adapted from those used by Mount (1983) ("Satisfaction with the Performance and Development Appraisal") and Le Roux (1989) to include aspects which are more in line with features of the performance appraisal system unique to the participating organisation.

Two variations of the measuring instrument consisting of two sections each were prepared for the research. One questionnaire was designed for subordinates while the other was designed for supervisors. Section A in each questionnaire covered demographic information while Section B consisted of forty five items each, which are directly focused on aspects of performance appraisal. Some questions in Section B which covered general appraisal aspects, were set exactly the same in both questionnaires. The remaining questions in each questionnaire were essentially the same but differed in terms of relationship with a particular dimension or factor, either from the point of view of a subordinate or from that of a supervisor. Responses to Section B of the questionnaires were recorded on a seven-point scale which elicited a high degree of discrimination.

Apart from the questions that addressed general aspects of performance appraisal, which have already been mentioned, there were also questions which were aligned with other factors of the aims of the study. These factors were utility, fairness, ethics,

motivation, accuracy, validity, rater error, effectivity, and feedback. The administrative aspects were addressed through seventeen items, while developmental aspects were addressed through fifteen items. Reliability for the questionnaires was measured on all items and the results indicated an overall Coefficient Alpha of 0.77 (rounded), made up of subordinates 0.80 (rounded) and supervisors 0.74, all of which are sufficient for research purposes.

Research procedure

Data were collected from the two groups, namely supervisors (from the ranks of administrative officer and equivalent to assistant director and equivalent) and subordinates (from those ranks below the rank of administrative officer). Both groups were further divided into "achievers", or those who were expected to form a positively orientated sub-group, and "non-achievers", or those who were expected to form a negatively orientated sub-group. The division between these two sub-groups was made on the basis of their classification of assessment over the last three years and whether they had received a merit award over the last five years or not. The assumption was made that employees who had received a "Class 1" or "Class 2" assessment classification and/or a merit award, were more positively orientated towards the performance appraisal system than those employees who had received a "Class 3" assessment classification and had never received a merit award.

The classification of assessments can be explained as follows: a "Class 1" assessment means that an employee is considered to be fit for promotion far in advance (1 year) of his/her due date, or as a matter of 'priority'. A "Class 2" assessment means that an employee is to be considered for promotion in advance (6 months) of his/her due date, or as it is termed an 'out of turn' promotion. A "Class 3" assessment means that the employee is performing within the expected norms of his/her job and may be considered for promotion after the minimum required period of time has been served in a particular rank/post. It is apparent then that Class 1 & 2 classifications represent performances of a very high standard while Class 3 classifications represent the norm. Merit Awards are coupled almost exclusively to Class 1 & 2 assessments. The information for the determination of the different groups/sub-groups was obtained from the demographic section of the questionnaire.

Although the questionnaires cover a fair amount of common ground compared with those used by Mount (1983) and Le Roux (1989), the questionnaires used in this study have been adapted to include a wider range of dimensions and items which are more in line with features of the performance appraisal system unique to the Public Sector. An important factor to bear in mind is that in this particular case, junior and middle management do not have the necessary executive powers to amend existing systems or implement new performance appraisal systems, but are dependent on the systems recommended and enforced by a central body being the Department of Public Service and Administration.

RESULTS

In order to illustrate or operationalise the previously stated aim of the research, use was made of descriptive statistics as well as multiple comparisons. The descriptive statistics consisted of means and standard deviations while use was also made of One-way analysis of variance (one-way ANOVA) and *Post Hoc* tests. Howell (1990) states that analysis of variance is reputed to be the most used technique in contemporary psychological research. The statistical analysis included a multivariate test to establish the significance of the Hotelling T^2 value prior to continuing with the one-way ANOVA and *Post Hoc* tests.

The statistical analysis was carried out by means of the "Statistical Program for Social Sciences" (SPSS for Windows package, George & Mallery, 1999). After the raw data had been acquired, frequencies were calculated for each of the biographic items that were used in Section A of the questionnaires as well as for each of the questions in Section B. Frequencies were also calculated for each of the sub-groups and the total group. Frequencies were then calculated for each of the dependent variables (utility, fairness, ethics, motivation, accuracy, validity, effectivity, feedback, administrative, and developmental aspects).

T-tests were conducted in terms of differences between the independent variables, (supervisors and subordinates) with reference to each of the dependent variables. This procedure was also carried out for differences between the achiever/non-achiever groups with reference to each of the dependent variables. Considering that the hypotheses were stated in non-directional

TABLE 1
DESCRIPTIVE STATISTICS OF EACH GROUP REGARDING FAIRNESS, ETHICS, ACCURACY, RATER ERROR AND ADMINISTRATIVE ASPECTS

Variable	Group	N	Mean	Standard	Standard	95% Conf. Interval		Min,	Max,
				Deviation	Error	Lower Bound	Upper Bound		
Fairness	Sub. Achievers	38	29,47	6,18	1,00	27,44	31,51	17	40
	Sub. Non Achieve	163	27,67	7,17	,56	26,57	28,78	6	42
	Supe. Achievers	80	30,54	3,72	,42	29,71	31,37	22	39
	Supe. Non Achiev	106	30,21	5,16	,50	29,21	31,20	18	42
	Total	387	29,14	6,08	,31	28,53	29,74	6	42
Ethics	Sub. Achievers	38	19,82	5,63	,98	17,97	21,67	8	32
	Sub. Non Achieve	163	20,92	6,75	,53	19,88	21,96	5	35
	Supe. Achievers	80	13,68	4,73	,53	12,62	14,73	5	28
	Supe. Non Achiev	106	14,76	6,47	,63	13,52	16,01	5	35
	Total	387	17,63	6,98	,35	16,93	18,33	5	35
Accuracy	Sub. Achievers	38	27,74	5,68	,92	25,87	29,61	9	37
	Sub. Non Achieve	163	27,74	6,81	,53	26,41	28,52	6	42
	Supe. Achievers	80	29,21	4,33	,48	28,25	30,18	18	39
	Supe. Non Achiev	106	28,84	5,16	,50	27,85	29,83	17	42
	Total	387	28,23	5,85	,30	27,65	28,81	6	42
Rater	Sub. Achievers	38	20,97	3,44	,56	19,84	22,11	16	29
	Sub. Non Achieve	163	20,31	3,66	,29	19,75	20,88	5	29
	Supe. Achievers	80	22,75	3,11	,35	22,06	23,44	16	32
	Supe. Non Achiev	106	21,34	4,05	,39	20,56	22,12	11	32
	Total	387	21,16	3,75	,19	20,79	21,54	5	32
Admin,	Sub. Achievers	38	76,37	16,78	2,72	70,85	81,89	44	104
	Sub. Non Achieve	163	71,62	18,89	1,48	68,70	74,54	29	116
	Supe. Achievers	80	80,31	11,26	1,26	77,81	82,82	55	105
	Supe. Non Achiev	106	76,78	14,16	1,38	74,06	79,51	52	113
	Total	387	75,30	16,40	,83	73,66	76,94	29	116

terms, two-tailed (or non-directional) tests were carried out so that extreme outcomes in either tail could be rejected.

Comparisons between all the independent variables, namely, the achiever and non-achiever sub-groups representing subordinates and supervisors, with the dependent variables were determined by means of the One-way analysis of variance procedure. Analysis of variance is also suited to situations where unequal numbers of observations exist. Although the variables were initially investigated in terms of planned comparisons in accordance with the hypotheses, other possible significant relationships were investigated by application of *Post Hoc* tests using the Bonferroni technique. These were carried out by relating each one of the independent variables with the other three, in terms of each of the dependent variables. It was considered that the outcomes of these tests may also elicit possible themes for future research.

Only the results concerning those factors reflecting statistical significance have been recorded, but before continuing with the results, it should be noted that the multivariate test produced a Hotelling T^2 value of 0.266, significance 0.000 ($p < 0.001$). In addition and by way of introduction to the comparative statistics, a table displaying descriptive statistics (TABLE 1) has been included as an indication of the distribution of scores reflecting the perceptions of each group.

The hypothesis for each variable namely, *fairness, ethics, accuracy, rater error and administrative aspects* states that there is a significant difference in perception of the performance appraisal system between (a) supervisors and subordinates, and (b) achievers and non-achievers. The exposé of results of the tests and procedures incorporate sections (a) and (b) of the hypothesis.

TABLE 2
T-TEST : PERCEPTUAL DIFFERENCES BETWEEN SUPERVISORS AND SUBORDINATES AND THEIR SUB-GROUPS, IN TERMS OF FAIRNESS, ETHICS, ACCURACY, RATER ERROR AND ADMINISTRATIVE ASPECTS

Variable	Groups	T-test for Equality of Means		
		t	df	Sig - p (2-tailed)
Fairness	Supervisors / Subordinates	-3,842	385	,000***
	Sup: Achievers / Sub: Non-achievers	485	184	,628
	Sub: Achievers / Sub: Non-achievers	1,428	199	,155
	Supervisors / Subordinates	10,169	385	,000***
Ethics	Sup: Achievers / Sub: Non-achievers	-1,271	184	,205
	Sub: Achievers / Sub: Non-achievers	-,935	199	,351
	Supervisors / Subordinates	-2,508	385	,013*
	Sup: Achievers / Sub: Non-achievers	,522	184	,602
Accuracy	Sup: Non-achievers / Sub: Achievers	,227	199	,821
	Supervisors / Subordinates	-4,035	385	,000***
	Sup: Achievers / Sub: Non-achievers	2,589	184	,010*
	Sub: Achievers / Sub: Non-achievers	1,014	199	,312
Rater error	Supervisors / Subordinates	-3,517	383	,000***
	Sup: Achievers / Sub: Non-achievers	1,834	184	,068
	Sub: Achievers / Sub: Non-achievers	1,424	199	,156
	Admin Aspects			

* $p < 0,05$
** $p < 0,001$

Perceptual differences between main groups, and sub-groups

Fairness. The *t*-test revealed the results as indicated in TABLE 2. The perceptual difference between supervisor and subordinate groups was significant as indicated in the table. The alternative hypothesis concerning supervisors and subordinates was therefore, supported ($p < 0.001$). The hypothesis for the sub-groups (b) was not supported.

Ethics. The *t*-test revealed the results as indicated in TABLE 2. The perceptual difference between supervisor and subordinate groups was significant as indicated in the table. The alternative hypothesis concerning supervisors and subordinates (a) as stated, was supported ($p < 0.001$). The hypothesis for the sub-groups (b) was not supported.

Accuracy. The *t*-test revealed the results as indicated in TABLE 2. The perceptual difference between supervisor and subordinate groups was significant as indicated in the table. The alternative hypothesis concerning supervisors and subordinates (a) as stated, was therefore, supported ($p < 0.05$). The hypothesis for the sub-groups (b) was not supported.

Rater Error. The *t*-test revealed the results as indicated in TABLE 2. Perceptual differences were significant between supervisors and subordinates (a) on the one hand and between the supervisor sub-groups, achievers and non-achievers (b). The alternative hypothesis as stated for supervisors and subordinates was supported ($p < 0.001$) as well as that for the supervisor sub-groups, achievers and non achievers ($p < 0.05$). The hypothesis was not supported in respect of the subordinate sub-groups, achievers and non-achievers.

Administrative Aspects. The *t*-test revealed the results as indicated in TABLE 2. The perceptual difference between supervisor and subordinate groups were significant as indicated in the table. The alternative hypothesis for supervisors and subordinates (a) was therefore, supported ($p < 0.001$). The hypothesis for the sub-groups (b) was not supported.

Comparisons among all groups

TABLE 3
ONE-WAY ANOVA'S REGARDING RELATIONSHIPS OF FAIRNESS, ETHICS, RATER ERROR AND ADMINISTRATIVE ASPECTS WITHIN THE TOTAL GROUP

Variable	Group interaction	Sum of Squares	df	Mean Square	F	Sig - p
Fairness	Between Groups	631,180	3	210,393	5,914	,001**
	Within Groups	13624,562	383	35,573		
	Total	14255,742	386			
Ethics	Between Groups	4068,091	3	1356,030	35,301	,000***
	Within Groups	14712,327	383	38,413		
	Total	18780,419	386			
Rater Error	Between Groups	323,954	3	107,985	8,118	,000***
	Within Groups	5094,790	383	13,302		
	Total	5418,744	386			
Administrative Aspects	Between Groups	4494,371	3	1498,124	5,778	,001**
	Within Groups	99310,456	383	259,296		
	Total	103804,827	386			

** $p < 0,01$
*** $p < 0,001$

Fairness. The results of the one-way ANOVA regarding relationships of Fairness within the total group elicited significant perceptual differences as indicated in TABLE 3. The Bonferroni technique of the *Post Hoc* test was used in order to identify where the differences occurred. Statistically significant diffe-

rences were indicated between the sub-group, subordinate non-achievers, with both supervisor sub-groups (achievers and non-achievers), $p < 0.05$.

Ethics. The results of the one-way ANOVA regarding relationships of Ethics within the total group elicited significant perceptual differences as indicated in TABLE 3. The Bonferoni technique of the *Post Hoc* test indicated statistically significant differences between each subordinate sub-group with each of the supervisor sub-groups respectively, $p < 0.001$.

Accuracy. The results of the one-way ANOVA regarding relationships of Accuracy within the total group elicited no significant perceptual differences. It was not therefore, necessary to carry out a *Post Hoc* test.

Rater Error. The results of the one-way ANOVA regarding relationships of Rater error within the total group, elicited significant perceptual differences as indicated in TABLE 3. The *Post Hoc* test indicated significant statistical differences between the subordinate sub-group, non-achievers and the supervisor sub-group, achievers, $p < 0.001$.

Administrative Aspects. Results of the one-way ANOVA regarding relationships of Administrative Aspects within the total group elicited significant perceptual differences as indicated in TABLE 3. The *Post Hoc* test indicated significant differences between the subordinate sub-group, non-achievers and supervisor sub-group, achievers, $p < 0.01$.

DISCUSSION

Perception of *Fairness* in performance appraisal. The rationale of the hypothesis is that there are perceptual differences between the different groups in the way fairness is applied in the performance appraisal process. This refers to whether the subordinate feels that he or she is being treated fairly through the system on the one hand and whether the supervisor considers that he or she is applying the given system in a manner deemed to be fair, on the other. This particular situation also has a direct bearing on whether mutual trust exists between rater and ratee. It also concerns the effectivity of the system. The outcome of the *t*-test ($t = -3.842$, $p < 0.001$) indicated a significant difference between the two main groups, supervisors and subordinates (see TABLE 2). The one-way ANOVA ($F = 5.914$, $p < 0.001$) also indicated statistical significance (see TABLE 3). The *Post Hoc* test indicated that the difference lay between the subordinate sub-group, non-achievers on the one hand and both supervisor sub-groups (achievers and non-achievers ($\{p < 0.01$ for both)), on the other. The hypothesis is therefore supported.

Perception of *Ethics* in performance appraisal. The rationale of the hypothesis is that there are differences between the different groups in the way they perceive ethics to be applied in the performance appraisal process. The underlying reason for the assumption was based upon general comment and the feeling that manipulation of ratings takes place for the fulfilment of possible ulterior motives. The inference of a lack of mutual trust between ratee and rater is apparent.

The *t*-test (see TABLE 2) yielded statistically significant differences between the two main groups, subordinates and supervisors ($t = 10.169$, $p < 0.001$). The one-way ANOVA ($F = 35.301$, $p < 0.001$) also indicated statistical significance (see TABLE 3). The *Post Hoc* test revealed that the differences lay between the subordinate sub-group, achievers and both supervisor sub-groups, as well as between the subordinate sub-group, non-achievers and both supervisor sub-groups ($p < 0.001$ for all). The hypothesis is therefore supported.

Perception of *Accuracy* in performance appraisal. The rationale of the hypothesis is that there are perceptual differences between the different groups with regard to the accuracy of the mechanics of the performance appraisal process. It has a direct bearing on the reliability of the process. Balzur and

Sulsky (1990) urge that reliability and validity should become more prominent when measuring rating effectiveness. Factors such as the qualification of supervisors to rate accurately on a continuous basis (not to be confused with rater error), who would be best qualified to rate accurately, and the use of consistent standards and criteria in the appraisal process, are included in this aspect.

The outcome of the *t*-test (see TABLE 2) indicated statistically significant differences between supervisors and subordinates ($t = -2.508$, $p < 0.05$). The hypothesis is therefore supported. The result of the one-way ANOVA was marginal, but did not produce significant relationships between or within the sub-groups ($F = 2.171$, $p < 0.091$).

It is obvious from the results of the tests that there is a significant difference between how subordinates and supervisors perceive the accuracy or reliability of rating effectiveness. In this case parallels may be drawn with the variables of fairness and ethics and to assume that these factors may have an influence on how subordinates perceive accuracy. The inference is that in a process where fairness and ethics are perceived to be absent, it can be expected that subordinates will have little faith in the accuracy or reliability of that same process.

Perceptions of *Rater Error* in performance appraisal. The rationale of the hypothesis is that there are perceptual differences between the different groups with regard to rater error in the performance appraisal process. This is very much a traditional area of research and has to do with the type of errors that raters are inclined to make as a result of habit, personality, inclination, etc.

The outcome of the *t*-test (see TABLE 2) indicated statistically significant differences between supervisors and subordinates ($t = -4.035$, $p < 0.01$) on the one hand and between the supervisor sub-groups ($t = 2.589$, $p < 0.05$) on the other. The one way ANOVA (see TABLE 3) also indicated statistical significance ($F = 8.118$, $p < 0.001$). The *Post Hoc* test indicated that the difference exists between the subordinate sub-group, non-achievers and the supervisor sub-group, achievers ($p < 0.001$). The hypothesis is therefore supported.

Logically the difference exists between the extreme sub-groups. The results also support the literature. Another factor which could influence perceptions on this particular aspect, is a possible correlation with the other statistically significant factors already identified, being fairness, ethics, and accuracy.

Perceptions of *Administrative Aspects* in performance appraisal. The rationale of the hypothesis is that there are perceptual differences between the different groups with regard to administrative aspects in the appraisal process. This factor examines the different perceptions regarding aspects such as promotions, transfers, disciplinary measures, dismissals etc. There is a strong relationship between this factor and utility. The main difference is that in this case a wider spectrum is covered and hence, more questions are incorporated for measurement which also increases reliability. Characteristics of this dimension are that the different aspects are orientated towards the past and are aimed at general performance (Kirkpatrick, 1986).

The outcomes of the *t*-tests (see TABLE 2) indicated statistically significant differences between supervisors and subordinates ($t = -3.517$, $p < 0.001$). The one-way ANOVA (see TABLE 3) also indicated statistical significance ($F = 5.778$, $p < 0.01$). The *Post Hoc* test indicated that the differences in perceptions ($p < 0.01$) existed between the subordinate sub-group, non-achievers and the supervisor sub-group, achievers. The hypothesis is therefore supported.

Conclusions

The results of the research have clearly indicated that perceptual differences exist between different groups in various computations for certain of the variables. From this point of view, the aims of the study were achieved. It also means that the

alleged complaints which public servants have aired, can largely, be attributed to the application of the performance appraisal system. This is substantiated by the results of the research.

The variables for which differences were observed are those of Fairness, Ethics, Accuracy, Rater error, and Administrative Aspects. At this point, it is interesting to note that there is a disquieting relationship between all five variables. The reason for stating this, is that Fairness, Ethics, Accuracy, and Rater error allude to the perceptions subordinates may have of the manner in which their appraisals are handled by supervisors or through the system. In other words they possibly perceive their supervisors as not giving them a fair chance, manipulating their appraisals, giving inaccurate evaluations or simply making rating errors which under normal circumstances should be avoidable. The effects of these perceived deficiencies are possibly seen to manifest themselves under administrative aspects where the system fails to take cognisance of what they consider themselves to be worth, in the sense that promotions and/or merit rewards are not forthcoming. When comparing the responses, it appears that supervisors are not aware of the intensity of the apparent frustrations being experienced by subordinates.

The remaining variables, although not indicating statistically significant differences, may nevertheless, have significant implications. For example, means for the remaining variables for each of the two main groups, were virtually inseparable and despite experiencing the system from different perspectives and roles, the two groups displayed consolidation of opinion. Most of the factors reflected neutral dispositions and this may also suggest that there are areas which still require a fair amount of attention. A possible positive alternative is that, in the case of these particular factors, subordinates are satisfied with the performance appraisal system.

As has been mentioned earlier, the participating organisation envisages the implementation of a different appraisal system in the near future; a system that will be more orientated towards the 'management by objectives' method. The expectation is that this will improve performance appraisal and hence, attitudes towards the whole concept. However, it does not matter what system is to be implemented, appraisals are about relationships, communication, co-operation, mutual trust and respect. Therefore, taking current perceptions into account, a good deal of effort will have to be put into changing the perceptions and opinions of ratees. This is especially critical when considering that non-achieving subordinates form the largest single group of employees. A further concern is that, if problems are experienced with a system that has been in use for so many years, the chances of a new and more sophisticated system succeeding, must be doubtful. In this sense, the research results in the case of this specific organisation should assist them in being able to identify problem areas and to take the necessary steps to counter them. Consideration towards the training of raters and implementing a multi-source assessment system with greater participation by all concerned, would be a strong recommendation.

The *Post Hoc* tests revealed that the main area of significant differences exists between the subordinate sub-group, non-achievers and the supervisor sub-group, achievers which understandably, represents the 'poles' of the total sample group. The next most common area of differences lies between the subordinate sub-group, non-achievers and the supervisor sub-group, non-achievers. In the case of Ethics, the difference lies between the subordinate sub-group, non-achievers and both supervisor sub-groups, which indicates how strongly this variable is 'contested' and is one of the areas that requires attention.

When the initial planned tests were carried out, no significant differences between the two subordinate sub-groups were observed. This suggests that the measures used for discriminating between these two groups, namely appraisal classifications and merit awards, are inadequate. Only when an additional discrete variable ("length of service") was introduced into the equation, could discrimination between the two groups be observed. This

should be borne in mind for future research, in the case of researchers wanting to use this manner of group division.

The fact that the Public Service is planning the implementation of a new performance appraisal system, provides the opportunity for a comparative (longitudinal) research project to be carried out on the perceptions of employees regarding the two systems. A before and after type opinion/perception measurement.

Finally, according to Spangenberg (1994), organisations that have implemented performance management systems, incorporating performance appraisal, have been very successful. This dispenses with the problem of performance appraisal becoming an isolated, once-a-year event and instead, forms part of a continuous process which cannot be fobbed off as something of mere nuisance value. All role-players under these circumstances are forced to become involved on a *continuous* basis. Perceptions under such circumstances could be drastically changed and research in this direction could therefore, contribute significant findings in the interest of HR management in particular, as well as industrial growth and development in general.

REFERENCES

- Cascio, W.F. (1987). *Applied psychology in personnel management*. Englewood Cliffs, NJ: Prentice-Hall.
- Deming, W.E. (1986). *Out of the crisis*. Cambridge: MIT Press.
- Dipboye, R.L. & de Pontbriand, R. (1981). Correlates of employee reactions to performance appraisals and appraisal systems. *Journal of Applied Psychology*, 66(3), 248-251.
- Folger, R. & Lewis, D. (1993). Self-appraisal and fairness in evaluations. In R. Cropanzano (Ed.), *Justice in the Workplace*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- George, D. & Mallery, P. (1999). *SPSS for windows step by step: A Simple Guide and Reference*. Boston, MA: Allyn and Bacon.
- Gibb, P. (1985). Appraisal goals and controls. *Personnel Journal*, 64(8), 89-93.
- Howell, D.C. (1990). *Fundamental statistics for the behavioral sciences (3rd ed.)* Pacific Grove, CA: Duxbury Press.
- Kirkpatrick, D.L. (1986). Performance appraisal: When two jobs are too many. *Training*, 23(3), 65-69.
- Latham, G.P. & Wexley, K.N. (1994). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley.
- Le Roux, C.W. (1989). *'n Empiriese ondersoek na perseptuele verskille tussen meenters en ondergeskietes ten opsigte van die administratiewe- en ontwikkelingsdimensies van prestasiebeoordeling*. Unpublished masters thesis, University of Stellenbosch, Stellenbosch.
- London, M. (1997). *Job feedback: giving, seeking, and using feedback for performance improvement*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Longenecker, C.O. & Goff, S.J. (1992). Performance appraisal effectiveness: A matter of perspective. *SAM Advanced Management Journal*, 57(2), 17-24.
- Milkovich, G.T. & Boudreau, J.W. (1988). *Personnel/Human resource management (5th ed.)* Plano: Business Publications.
- Milkovich, G.T. & Wigdor, A.K. (Eds.) (1991). *Pay for performance*. Washington: National Academy Press.
- Moon, P. (1997). *Appraising your staff*. Derby: Clays Ltd.
- Mount, M.K. (1983). Comparisons of management and employee satisfaction with a performance appraisal system. *Personnel Psychology*, 36, 99-109.
- Murphy, K.R. & Cleveland, J.N. (1991). *Performance appraisal*. Boston, MA: Allyn and Bacon.
- Philp, T. (1990). *Appraising performance for results*. London: McGraw-Hill.
- Republic of South Africa (RSA) (1994). *Public service staff code*. Pretoria: Government Printers.
- Schneier, C.E., Beatty, R.W. & Baird, L.S. (Eds.) (1987). *The performance management sourcebook*. Amhurst, MA: Human Resources Development Press.
- Schuler, R.S. (1981). *Personnel and human resource management*. St. Paul: West.
- Spangenberg, H. (1994). *Understanding and implementing performance management*. Kenwyn: Juta & Co.