The development of a behaviourally-based performance evaluation system for research scientists

Amos Drory

The development of a behaviourally-based performance evaluation system for research scientists

The development of a behaviourally-based performance evaluation system for research scientists

Amos Drory

Amos Drory, Ph.D., Temple University, Philadelphia, USA; Senior lecturer,
Department Industrial Engineering and Management, Ben-Gurion
University of the Negev, Beer-Sheva

ACKNOWLEDGEMENTS

CONTENTS

SUMMARY

The purpose of this project was to develop a practical tool to enable
the assessment of research workers' performance for administrative and
feedback purposes, at the NIPR and comparable organizations. The items
for evaluation were generated following the behaviourally anchored
rating scales (BARS) approach. This approach calls for the assessment
of performance on the basis of actual activities which characterize the
incumbant's behaviour on the job rather than on general qualities and
traits.

In the first stage a representative sample of research and professional
employees at the NIPR were interviewed in an attempt to specify the
particular components of their job and to solicit a sample of critical
job behaviours which might lend themselves to direct observation by
superiors and peers and which could help to distinguish between poor and
good job performance.

The pool of critical behaviours was then reviewed by a group of Divisional
Heads for reliability and relevance and a preliminary set of assessment
items was prepared. The items were presented in a three-point scale form
representing high, moderate and low levels of proficiency with regard to
every job behaviour. This preliminary scale was then administered to all
NIPR Divisional Heads who were requested to assess all their professional
subordinates

The scale was also given to the subordinates themselves for self assessment.
Analysis of the responses revealed that although there were certain
discrepancies between superiors and subordinates, these were of manageable
magnitude. In contrast to previous findings subordinates did not rate
themselves significantly higher than they were rated by their superiors.
Subordinates' self-assessment was somewhat affected by the relative
importance which they attributed to the items. The scale was revised
in view of the respondent's comments and an updated version is presented
with a guideline for further refinements and the eventual implementation
of the evaluation system.

1. INTRODUCTION

The assessment of human performance at work has been a major
issue of concern for both researchers of human behaviour and
practitioners for the past few decades. It is widely agreed
that the proper measurement of employees' performance is an
essential tool for purposes of manpower management and an
important key to organisational effectiveness. A number of
tools and methods have been devised in order to allow an
effective assessment of workers' performance. Yet, in spite
of the considerable effort invested in this direction,
difficulties still exist both with the psychometric aspects
pertaining to the evaluation of human performance and in terms
of dealing with the informal political forces in the organisa-
tion which frequently inhibit the assessors from providing
honest and accurate appraisals.

Two major goals were set forth for the present project:

(a) The design and implementation of a performance appraisal
    system for researchers and professionals at the NIPR and
    comparable organisations taking advantage of methods
    representing the current state of the art and accommodating
    for the specific situation and needs of the NIPR. This
    objective was set with a more general perspective, namely
    developing a model of performance evaluation which may be
    adopted in the future for possible use in other research
    organisations.

(b) Testing some specific hypotheses with regard to the
    assessment of job behaviour and performance by superiors
    and subordinates, in research organisations. This
    objective was set in an attempt to contribute to the
    present state of knowledge in the area of performance
    evaluation.

Before the actual process of the scale development is described, attention should be directed to several relevant characteristics of the NIPR organisation as well as to some general considerations.

Firstly, most professional employees at the NIPR possess advanced training in the areas of human behaviour and while the psychological literature does not necessarily suggest that such training makes one a better judge of human behaviour, it is certainly expected that a relatively more critical and systematic approach to this task would be commonplace here in comparison to other professions. This characteristic may affect both the task of evaluating others as well as of self-evaluation. Secondly, being essentially a research organisation, the NIPR is character- ised by a relatively high level of individual freedom in terms of managing one's own work while the style of management, supervision and control is necessarily general rather than close and restrictive. The goals and objectives of such an organisation are by nature more general than would be expected in more structured profit-making firms. It is therefore quite natural to expect different and often contradictory interpretations of organisational goals coupled with a felt need for a clearer definition of these goals, a need which is almost impossible to meet. The ambiguity of goals and objectives may easily lead to incongruent role expectations among superiors and subordinates with regard to proficiency levels, task priorities, and the choice of research directions.

Another relevant organisation aspect is the considerable diversity of activities and tasks within the NIPR. Although most professional NIPR employees fall under the general category of research officers, the specific daily activities comprising their jobs are far from being similar. Some typical research activities such as literature reading, report writing, data gathering, statistical analysis and instrument construction are performed in different proportions by different units and individuals. While one unit is predominantly

2.

involved in test construction, another one is concerned most with methodological consultation and a third unit performs EEG analyses while still other units deal with training, counselling, basic research, and so on. Devising a single performance assessment scale to cater for the needs of all divisions is obviously not a simple task. In the process of developing the appraisal system an attempt was made to take into account the above characteristics and to provide practical solutions whenever needed.

The first step in the development of any performance appraisal system is the selection of the desirable measurement approach. A distinction can be made between two general types of appraisal methods. The first is referred to as the traditional or trait approach. Typically, a list of desirable qualities and personal traits considered relevant for work performance is drawn and a rating scale is attached to every item on the list. The immediate superior is then required to rate every one of his subordinates using the scales. Work-related characteristics such as 'understanding instructions', 'co-operation', 'persistence', 'originality', 'industriousness', and the like, are common examples. The task of the superior is essentially to judge his subordinates' qualities on the basis of his familiarity with them. The major disadvantage of this approach lay in the potential differences in standards and expectations among different superiors and in the questionable validity of such judgements being prone to various psychological biases.

A second more recent approach to performance appraisal which has become very popular is the Behaviorally Anchored Rating Scales (BARS). This approach which claims to overcome some of the major shortcomings of the trait approach is based on the sampling of actual relevant behavioural incidents occurring on the job. These incidents, after being properly screened and judged for their relevance and importance for the job, are then transformed into

a scale form. The superior is then requested to indicate the extent to which the incidents occur for each subordinate.

Performance appraisal systems based on the Behaviorally Anchored Approach have been reported in the literature in an increasing frequency during the last decade. It has been claimed that the behavioural approach is advantageous over the trait approach since it relieves the rater from the necessity of making a complex psychological judgement on his subordinate's qualities and only requires that he makes some straight-forward report on which critical behaviours were emitted. Some of the major biases associated with the task of person perception are more likely to be eliminated under these specific conditions and the results are expected to become significantly more reliable and valid in comparison to the trait approach. The items on which the subordinate is being evaluated provide a more relevant representation and generally make more sense both to the rater and to his subordinate.

In spite of the increasing popularity of this approach, a number of empirical studies failed to demonstrate its psychometric superiority over the trait based scales and the issue is still controversial (Bernardin and Smith, 1981). Current advocates of the BARS approach claim, however, that its advantages go far beyond the psychometric aspects (Jacobs, Kafri and Zedeck, 1980). They maintain that being based on actual job behaviours rather than general traits, BARS are clearly advantageous for purposes of performance feedback, as well as for organisational planning of training, placement, promotion, and job analysis. It was found, for example, that feedback to ratees on the basis of BARS generates greater behavioural changes than on the basis of standard scales (Hom, Denisi, Kinicki, and Banister, 1982).

On the basis of the above considerations, it was decided to adopt the BARS approach for the present project. It was believed that the BARS' advantages in terms of its potential vast

4.

applications for administrative and managerial purposes justify
the relatively extensive effort invested in the development of
the system.

2.   METHOD

The performance appraisal scale was developed through the
following steps.

2.1   Collection of critical incidents

In this step a sample of job incumbents in all professional
divisions at the NIPR were interviewed. In the first phase of
the interview the incumbent was asked to describe the specific
activities performed in his job and to indicate the relative
proportions of time devoted to the different activities. He
was then asked to provide specific behavioural examples of
performing the various task components. Examples of both high
levels and low levels of proficiency on the job were requested.
In addition, the incumbent was asked to suggest personal
qualities and talents needed to perform his job and to provide
some behavioural examples in which those qualities are reflected.
In some cases the interviewee was given a few days to produce
additional behavioural incidents in writing.

2.2   Construction of the initial item list

The interviewee process yielded a large number of critical
behaviours pertaining to the various aspects of the typical jobs
performed by research and professional employees in the different
divisions. The behavioural items were first sorted for
duplications and redundancies. The list was then content

analysed according to two content dimensions. The first dimension was "Nature of Task" following the logical sequence of research work. The specific categories which emerged from the content analysis along this dimension were:

i)   Reading professional literature

ii)   Initiating a research programme

iii)   Data gathering (including instrument construction, testing, interviewing, and so on)

iv)   Data analysis

v)   Report writing

Two additional tasks which did not fall under the logical sequence and yet were repeatedly referred to by job incumbents and were therefore included as separate categories were:

vi)   co-operating with and assisting others; and

vii)   liaison and public contacts

The second dimension of content analysis was "Personal Qualities" reflected in the critical behaviours. The following categories were identified:

i)   Skill and knowledge

ii)   Motivation

iii)   Conceptualisation and creativity

iv)   Interpersonal skills

v)   Thoroughness and rigour

Every behavioural item can be classified simultaneously under the two dimensions, yet not all the task categories require all the personal qualities. Table 1 presents the relationships between the two dimensions. The purpose of the content analysis process was to ensure that all major task and quality categories were

6.

sufficiently represented in the behavioural item pool. In
cases where it was felt that insufficient items were
available with regard to given category, additional interviews
were conducted in order to solicit additional items. The items
were then rewritten in a scale form with three levels of
proficiency for every item, following Rosinger et al.'s (1982)
procedure. Thus, for every particular item three behavioural
statements were presented, the first describing a high
proficiency behaviour in performing the particular task, the
second describing a moderate proficiency behaviour, and the third
describing a low proficiency behaviour. An item was to be rated
by selecting one of the three statements which best described
the ratee's behaviour on the job.

2.3   Initial judgement of the scale items

At this stage the pool of items was given to all divisional
heads for initial judgement. They were asked to first review
every item and suggest improvements in style and wording. Then
for every item, the judges were to rate the importance of it
for the incumbent's overall work performance on a scale from
1-4, and to rate the frequency with which the particular
behaviour described by the items occurs in their divisions.
The divisional heads were also encouraged to suggest additional
items of importance which were not presently covered. The
ratings for all items were tabulated and items which were
consistently rated as either unimportant or as very infrequent
were deleted. In turn, some new items suggested by division
heads were added.

2.4   Pilot run of the appraisal form

The items which survived the previous stage were arranged in a
scale form with the appropriate instructions for the rater. Two

overall performance appraisal items were also added to the form.
One item called for the rating of the incumbent's overall
performance on a five-point scale ranging from "Excellent" to
"Very Poor". The second item requested the ranking of the
incumbent's job performance in comparison with the total group of
NIPR research and professional staff on a five-point scale
ranging from "Top 10%" to "Lower 10%". The complete appraisal
form is presented in Appendix 1. This form was distributed
among all divisional heads who were asked to assess each one of
their subordinates individually by choosing for every item the
one statement which best described his behaviour on the job. In
addition, they were asked to rate the importance of the item for
the ratee's job performance on a four-point scale ranging from
"Not Important" to "Very Important". In their evaluation they
were asked to refer to the last 12 months of work. A similar
form was given to every job incumbent for self-evaluation and
comment. It was specifically stated in the instructions that
the responses would be used for experimental purposes only and
that the respondent's identity would be kept strictly confidential.

Responses were gathered from 11 divisional heads, evaluating 60
employees. Fifty-eight employees responded to the self-assessment
form. The following section will present the analyses and results
of the pilot data.

3.   RESULTS

As a first step in the analysis of the data, the means and standard
deviations of every item were calculated separately for superiors'
and subordinates' judgements. Table 2 presents these results. It
may be observed that the means shift towards the upper end of the
scale (1 being the highest score and 3 the lowest). An inclination
in this direction is expected for two reasons.

First, a leniency bias in performance rating is widely acknowledged
in the literature as a common characteristic of performance rating
scales. This bias may result from several sources such as the fear
of the superior that negative or low assessment of his subordinates

may reflect negatively on the quality of his leadership (Smith, 1976; Drory and Ben Porat, 1980). A more objective reason for the existence of this bias may be that through the organisational process of selection, coaching, promotion and dismissals, the employee's level of performance is shaped toward a more or less desirable level.

The comparison between the superior and subordinate assessment shows that for most categories the differences in ratings were minimal and non-significant. This finding stands in contradiction to previous research on this issue (Thornton, 1980; Kirchner, 1966; Holzbach, 1978) when it was consistently found that subordinates tend to significantly give themselves higher scores than given by their superiors regardless of evaluation method used and organisational rank.

The lower self-assessment found at the NIPR in comparison to other worker populations may be due to the fact that most professional employees at the NIPR were trained within the academic disciplines of human behaviour such as psychology, sociology and education. It may be speculated that as a result of such training a greater aware-ness and perhaps acceptance of norms of self-criticism is developed and as a result there is a greater tendency in this population to admit self-deficiencies and exhibit lower proneness towards social desirability bias in self-reporting questionnaires.

The comparison of average ratings obviously does not imply that the rate of superior-subordinate agreement on rating is either high or low. The only valid conclusion pertains to the comparison of average leniency effect in the two samples.

In order to further examine the correspondence between superior and self-assessment the responses of both were cross-tabulated for every item. Table 4 presents the rate of correspondence between superiors' and subordinates' ratings for every item. Column 1 in Table 4 shows the percentage of matched ratings.

9.

Column 2 presents the percentage of non-matching ratings where the superior's rating was higher than the self-rating. Column 3 shows the percentage of disagreement in rating where the subordinates rated themselves higher than their superiors. The overall picture is that on the average in approximately 57% of the cases subordinates and superiors agree on the rating of the subordinate's performance. In about 18% of the cases superiors rated their subordinates' performance higher than they rated themselves and in about 25% of the cases subordinates gave themselves higher ratings than given by their superiors. These results suggest that if the evaluation procedure includes a mutual discussion between the superior and his subordinate about the subordinate's performance, the potential disagreement is not expected to be very high assuming that the subordinate is not likely to contradict his superior when over-rated by him. In view of the advantages of such a procedure for purposes of feedback, guidance and goal setting, such a maximum rate of potential conflict may be considered quite acceptable.

It should be noted, however, that the rate of superior-subordinate agreement in a real life application of the appraisal system may be somewhat different from the present pilot run situation. The present results should therefore be viewed as a preliminary indication. A high rate of superior-subordinate agreement is by no means a necessary requirement for the implementation of the system. It is, however, an encouraging sign suggesting that it is feasible that performance feedback could take place relatively smoothly.

In order to examine the internal consistency of the scale, the items were grouped into seven subscales accordng to the task categories. Table 5 presents the items included in each subscale and the Cronbach Alpha coefficients for all subscales are sufficiently reliable. The intercorrelations between the subscales are presented in Table 6. In general, these intercorrelations are moderate in magnitude. This pattern of relationships suggests that the subscales are not independent of each other. This finding is not surprising since a considerable halo effect is claimed to exist among the various aspects of performance

evaluation (Hulin, 1982). It has been argued that the situation here is somewhat similar to the structure of intelligence where a general factor is suggested in addition to specific factors. It is therefore agreed that the interrelationships between different aspects in performance appraisal do not stem predominantly from a perceptual bias, but rather represent a true structure of human performance. Although statistical techniques were suggested to eliminate the common factor using a post hoc approach (Landy and Far, 1980; Landy, Vance, Barnes-Farrel and Steele, 1980) and others argued that such an approach is undesirable and distorting (Murphy, 1982; Hulin, 1982).

It may be concluded that the expected moderate inter-relationships among the performance assessment subscales do not invalidate the scale as a useful tool for personnel management purposes.

Table 3 presents the means and standard deviations of the importance ratings for the scale items for superiors and subordinates. It may be observed that the ratings range from 2.31 to 3.44. Apparently none of the items was considered totally or nearly totally unimportant either by superiors or by subordinates. The differences between the mean important scores of superiors and subordinates for all individual items were rather small. None of these differences was statistically significant. There was therefore no basis for the exclusion of any of the items for lack of relevance or importance for the NIPR. Certain items were however considered irrelevant for individual divisions. As will be suggested later, the actual employee assessment should be based on weighting the score of every rated item by its importance for every division with the exclusion of the irrelevant items in every case.

Another issue examined in this project is the relationship between the importance attributed to the performance items and the actual ratings of the items. It was hypothesised that subordinates will evaluate themselves higher on items which they consider to be of higher importance to their job. It is presumably easier for one to admit low proficiency on a task considered relatively unimportant than on highly important tasks. For every item, Pearson correlations were computed between the importance rating and the rating of the item itself. The correlations were computed separately for superiors

11.

and subordinates. Table 7 presents the correlations obtained for every item. It can be seen that for 18 out of the 30 scale items there were significant differences in the magnitude of the correlations in the expected direction.

Since the index of correlations does not suggest the direction of causality, it is impossible to ascertain on the basis of these results whether the perception of the items' importance affected self-rating or, possibly, the rating affected its attributed importance. This issue may be tested by using a different experimental design. At this stage, it may only be suggested that potential bias in self-assessment may be a partial cause for the discrepancies between superiors' and subordinates' assessment of the subordinates' performance.

Finally, the relationship between the five subscales and the global rating of performance was examined. It may be recalled that both superiors and subordinates were asked to assess the subordinate's overall performance twice, once by rating the overall performance on a five-point scale and then by ranking their performance in comparison to all NIPR professional employees. The correlation between the two measures was high (,79 for superiors and ,82 for subordinates). It was therefore decided to concentrate only on the global rating item for the purpose of the present analysis. The correlations between the seven subscales and the global item for superiors and subordinates are presented in Tables 8 and 9. For most subscales the magnitude of the correlations is considerably higher for the superiors group. Tables 8 and 9 present the results of a multiple regression analysis for the two groups using the global item as a criterion and the seven subscales as predictors. The multiple R's obtained for the two analyses are fairly similar (,87 for superiors and ,77 for subordinates) in spite of the higher simple correlations in the superiors group. In the superiors' sample analysis there was only one individual significant predictor in the multiple regression equation whereas

in the subordinates group three predictors contributed uniquely and significantly to the equation. The conclusion which may be drawn from this comparison is that there is a higher overlap among different aspects of performance as they are perceived by superiors whereas subordinates make a clearer distinction among performance aspects. Apparently, superiors are more prone to the halo effect in evaluating their subordinates' performance than their subordinates. This phenomenon may be attributed to the fact that a superior assesses the performance of a number of persons whereas a subordinate only assesses himself. It may also be noted that there is an apparent disagreement between superiors and subordinates with regard to the relative importance of the various performance aspects to the overall assessed performance. These differences are particularly reflected in the areas of data gathering, testing and instrument construction. While subordinates see no correlation between their own performance on that aspect and their overall performance the superiors' ratings suggest a very high correlation between the two. Similar discrepancies, though of a lower magnitude, were found in the areas of public contacts and relations, initiating research and catalysing others.

As a result of the pilot test and some of the comments made by superiors and subordinates, some minor changes were introduced to the scale and four problematic items for which the rate of importance and the variance in rating were relatively low (items 8, 17, 20, 21) were deleted from the scale. The updated version of the scale appears in Appendix 2.

4. PROPOSED IMPLEMENTATION OF THE PERFORMANCE EVALUATION SYSTEM

A word of caution is due with regard to the interpretation of the results. Both superiors and subordinates who responded to the assessment scale were aware of the fact that the scale was in a developmental stage at the time of data collection. Although it was made clear to all concerned that the scale is constructed with the intention of implementing it for practical purposes, it may be argued that their responses may not always resemble those which might be given in a clearly real life performance assessment situation. In this sense it would be important to replicate the data gathering and analysis phase after the system was implemented for a trial period in order to ascertain the validity of the present conclusions.

The scale development stages described above produced a behaviourally based scale which should not be looked at as a final product but rather as a basic version subject to further refinement. It is suggested that the next steps of this project will involve a comprehensive review of the scale with all potential raters at the NIPR. It is hoped that their insights resulting from the experimental practice will lead to further improvements of the scale. An implementation of the scale for a trial period should then take place during which the system's purposes such as feedback, goal setting, and incremental pay will be determined and routine evaluations will follow according to the guidelines presented in the following pages.

A number of practical issues regarding the implementaion of the scale remain to be discussed.

14.

4.1   The considerable difference in immediate goals, tasks and daily
       activities between and sometimes within divisions make it
       impossible to assess all professional NIPR employees by a single
       uniform set of standards. Some measure of flexibility must be
       allowed within the evaluation system in order to enable fair and
       realistic performance assessment. In order to secure such
       flexibility, it is suggested that importance rating be used as a
       weighting factor for the corresponding items. The head of the
       division should determine the importance and relevance of every
       item for his division and wherever appropriate should assign
       different importance ratings for different individuals in his unit
       in accordance with their particular function in the division. The
       assigned weights should be brought to the attention of the
       subordinates and preferably even discussed with them before the
       actual assessment takes place. Such communication will contribute
       to the clarification of priorities and objectives for the group.
       At the time of assessment the rating of each item should be
       multiplied by the importance rating for the final scoring of the
       assessment form. The importance rating should be reconsidered
       periodically in order to account for changes in individual
       positions and unit goals. The assessment in some divisions will
       remain somewhat problematic even after the weighting procedure is
       adopted. For example, two divisions in particular, being partly
       service divisions, differ considerably in their activities from
       most other divisions at the NIPR, and since the assessment form was
       designed primarily for the more research oriented division, a
       number of items are not applicable to them. It would be proposed
       that the evaluation system still be implemented in these divisions
       for a trial period after which the issue could be discussed with
       the relevant divisional heads in order to finally decide whether
       the scale in its present form can satisfactorily serve its purpose
       for these two divisions.

4.2   It was occasionally argued by subordinates and some supervisors, during the initial interviews, that the divisional heads are frequently not in a position to judge their subordinates' task behaviour, as in many cases the researcher's job is performed with a high degree of autonomy and with only infrequent reports to his superior. The results of the pilot study do not suggest that this is indeed a major problem. Apparently the discrepancies between judgements of superiors and subordinates were not very sizable in most cases. It is recommended, however, that some measures be taken to ensure a better and more intensive observation procedure to enable superiors to base their ratings on hard facts rather than on general impressions. Bernardin and Smith (1981) argue that the major potential advantage of the BARS approach over the more traditional strategies of performance appraisal lies in the standardisation of observation procedures which will eventually lead to a summative rating of behaviours. It is there-fore suggested that divisional heads attempt to gather observational information pertaining to the scale items on a continuous basis and keep a record of such observations to ensure the validity of their final ratings. It is acknowledged that direct behavioural observations are not always practical in various situations. However, an attempt was made in the construction of the scale to focus only on work activities which may be observed by the superior and avoid behavioural aspects which do not lend themselves to direct observation.

4.3   One of the major purposes of performance appraisal is to provide feedback to the job incumbent. Such feedback may serve as a basis for future goal setting and guidance and establish a useful tool for constructive two-way communication between superiors and subordinates. It was demonstrated that the BARS approach is particularly useful for this purpose (Hom et al., 1982), since it is based on observed behaviour rather than on subjective trait perception. In order to achieve this objective it is recommended

16.

that a standard procedure be instituted whereby every
subordinate assesses himself using the scale in every evaluation
period and then meets with his superior to compare their
assessments, discuss the differences and agree on future attempts
to improve effectiveness. Such a periodical encounter may be
beneficial for both sides and is strongly advocated in the
literature (Jacobs et al., 1980).

The results of the pilot study indicate that although the
structure of self-assessment differs in some ways from that of
superior assessment (i.e. there is a greater halo effect in
superiors' judgements and a greater interdependence between the
item importance and its actual rating in self-assessment), there
is still a considerable agreement between the two. In view of
the correspondence between superiors' and subordinates' ratings
it is unlikely that conflict of unmanageable proportions will
arise in the process of such encounters. It is therefore
believed that the adoption of the above procedure would be of
significant benefit in the management of NIPR personnel.

5.   A GUIDELINE FOR FURTHER REFINEMENT AND IMPLEMENTATION OF
     THE PERFORMANCE APPRAISAL SYSTEM AT THE NIPR

The appraisal scale may serve in the future for two major purposes
namely:

i)   feedback to subordinates and

ii) incremental pay and promotion decisions.

Although additional functions such as assessing training needs
and selection criterion may also be considered, the present
discussion will be restricted to the two major purposes
mentioned above.

It is the view of this author that the scale should be first implemented for feedback purposes and only after enough confidence is gained in its applicability should it be adopted as a data resource in determining pay increments and advancement. It is quite possible, however, to consider other options of implementation such as initially concentrating on the incremental pay aspect depending on directorial needs and preferences.

The following stepwise guideline is suggested in order to provide a structured procedure for the implementation of the appraisal system.

A.   Revision and refinement of the appraisal scale. The scale items will be reviewed and refined by the group of division heads on the basis of their experience in the pilot phase. Items' relevance and weights will be determined for each division. The specific purposes of the scale (feedback, incremental pay, etc.) for the implementation trial period will be discussed and determined at this stage.

B.   Implementation of the appraisal system for a trial period.

   i)   Division heads will discuss the appraisal system with their subordinates and present the items and their weights.

   ii) A timetable will be set for the first evaluation.

   iii) Divisional heads collect and record observational information using the scale items as a guide on a continual basis.

   iv) At the predetermined evaluation time, divisional heads and subordinates complete the assessment forms.

v)   Feedback interviews: divisional heads meet with every
       one of their assessed subordinates to discuss their
       evaluations and set specific goals for improved effective-
       ness for the future if necessary.

vi)   Completed assessment forms are reported to management.
       Item scores are weighted by their importance rating and
       the total score is expressed in terms of percentage out
       of the total number of items scored (excluding the irrelevant
       items which were not scored).

vii)   Appraisal data will be analyzed and reported.

viii) The scale, appraisal system, and its purposes will be
       finally discussed with the head of the organisation and
       divisional heads and final decisions will be taken with
       regard to the permanent implementation of the appraisal
       system.

References

1. Bernardin, H.J., and Smith, P.C. A clarification of some issues regarding the development and use of behaviorally anchored rating scales (BARS). Journal of Applied Psychology, 66, (4), 1981.

2. Drory, A., and Ben Porat, A. Leadership style and leniency bias in performance appraisal. Psychological Reports, 46, 1980.

3. Holzbach, R.L. Rater bias in performance ratings: supervisor, self and peers ratings. Journal of Applied Psychology, 63, 1978.

4. Hom, P.W., Denisi, A.S., Kinicki, A.J., Banister, B.D. Effectiveness of performance feedback from behaviorally anchored rating scales. Journal of Applied Psychology, 67, (5), 1982.

5. Hulin, C.L. Some reflections on general performance dimensions and halo rating errors. Journal of Applied Psychology, 67 (2), 1982.

6. Jacobs, R., Kafri, D., and Zedeck, S. Expectations of behaviorally anchored rating scales. Personnel Psychology, 33, 1980.

7. Kirchner, W.K. Relationships between supervisory and subordinate ratings for technical personnel. Journal of Industrial Psychology, 3, 1966.

8. Landy, F.J. and Far, J.L. Performance rating. Psychological Bulletin, 87, 1980.

9. Landy, F.J., Vance, R.J., Barnes-Farrel, J.L., and Steele, J.W. Statistical control of halo error in performance ratings. Journal of Applied Psychology, 65, 1980.

10. Murphy, K.R. Difficulties in the statistical control of halo. Journal of Applied Psychology, 67, 1982.

11. Rosinger, G., Myers, L.B., Levi, G.W., Loar, M., Mohrman, S.A., Stock, J.R. Development of a behaviorally based performance appraisal system.

Personnel Psychology, 35, 1982.

12. Smith, P.C. Behavior results and organizational effectiveness; the problem of criteria. In M. Dunnette (Ed.) Handbook of Industrial and Organizational Psychology. Chicago: Rand McNally, 1976.

13. Thornton, G.C. Psychometric properties of self appraisals of job performance. Personnel Psychology, 33, 1980.

Table 1

Relationship : "Nature of Task" - "Personal Qualities"

	1 Skill and Knowledge	2 Motivation	3 Conceptualisation and Creativity	4 Interpersonal Skills	5 Thoroughness and Rigour
1. Reading	X	X	X		X
2. Initiating research (ideas, design, samples)	X	X	X		X
3. Data collection (instrument construction and selection, testing, interviewing, training, EEG)	X	X		X	X
4. Data analysis (methodology, statistics, computer)	X	X	X		X
5. Report writing (research, counselling, training, selection, proposal-content and style)	X	X	X		X
6. Facilitating and catalysing others	X	X	X	X	
7. Liaison and public contacts	X	X		X	X

Table 2

Means and Standard Deviations for the Scale Items

Item	Supervisors		Subordinates	
	$\overline{X}$	σ	$\overline{X}$	σ
1	1.80	0.55	1.73	0.54
2	1.57	0.66	1.45	0.60
3	1.36	0.54	1.34	0.62
4	1.62	0.66	1.46	0.55
5	1.62	0.66	1.51	0.59
6	1.97	0.60	1.63	0.65
7	1.86	0.73	1.52	0.56
8	1.17	0.47	1.27	0.45
9	1.76	0.65	1.60	0.50
10	1.36	0.60	1.24	0.49
11	1.96	0.66	1.98	0.58
12	1.42	0.60	1.38	0.53
13	1.65	0.52	1.73	0.53
14	1.60	0.61	1.64	0.53
15	1.89	0.80	1.98	0.96
16	1.68	0.62	1.47	0.50
17	1.22	0.42	1.12	0.33
18	1.51	0.55	1.62	0.54
19	1.43	0.67	1.36	0.53
20	1.25	0.44	1.13	0.42
21	1.29	0.54	1.20	0.40
22	1.50	0.67	1.56	0.62
23	1.23	0.47	1.15	0.41
24	1.32	0.55	1.37	0.49
25	1.83	0.60	1.71	0.52
26	1.53	0.55	1.47	0.56
27	1.47	0.56	1.50	0.56
28	1.56	0.70	1.62	0.61
29	1.47	0.54	1.24	0.52
30	1.60	0.63	1.49	0.60

Table 3

Means and Standard Deviations for Items' "Importance" Ratings

Item	Superiors		Subordinates	
	$\bar{X}$	σ	$\bar{X}$	σ
1	3,11	1,11	3,18	1,05
2	3,09	1,20	2,78	1,20
3	3,07	1,20	3,04	1,14
4	3,05	0,95	3,12	1,00
5	2,98	0,95	2,91	0,94
6	3,11	0,89	3,02	0,93
7	3,28	0,96	3,06	1,07
8	2,37	1,32	2,58	1,30
9	2,62	1,37	2,87	1,17
10	3,06	1,30	2,83	1,34
11	2,89	0,83	3,12	0,98
12	3,09	0,89	3,08	0,89
13	3,27	0,82	3,26	0,87
14	3,18	1,06	3,30	1,00
15	2,98	1,10	2,88	1,13
16	3,37	0,87	3,26	1,12
17	2,96	1,17	3,17	1,14
18	3,06	1,13	3,06	1,07
19	2,88	1,23	3,00	0,97
20	2,75	1,20	2,86	1,19
21	2,94	1,28	3,16	1,03
22	3,14	1,07	3,44	0,70
23	3,27	0,97	3,34	0,75
24	3,14	0,95	3,04	0,92
25	3,02	1,11	2,91	1,13
26	2,94	1,15	2,72	1,24
27	2,98	1,25	2,97	1,12
28	3,36	0,65	3,22	1,04
29	3,22	0,86	3,10	0,95
30	2,96	1,05	3,21	1,10

24.

Table 4

Rate of Correspondence between Superiors' and Subordinates' Ratings (in percentage)

Item	1 % of agreement in rating	2 % of higher rating by superior	3 % of higher rating by subordinates
1	55.0	19.0	26.0
2	64.0	7.0	29.0
3	44.4	22.2	33.4
4	44.2	27.9	27.9
5	65.1	16.3	18.6
6	51.3	5.4	43.3
7	64.5	6.5	29.0
8	72.0	20.0	8.0
9	56.7	13.3	30.0
10	60.0	10.0	30.0
11	69.8	18.6	11.6
12	43.4	23.9	32.6
13	57.5	25.5	17.0
14	65.1	16.3	18.6
15	58.0	13.0	29.0
16	47.1	17.6	35.3
17	68.6	8.6	22.8
18	48.6	27.1	24.3
19	65.0	13.5	21.5
20	63.3	10.0	26.7
21	61.9	16.7	21.4
22	45.4	34.2	20.4
23	77.1	8.4	14.5
24	61.4	22.6	16.0
25	46.8	21.8	31.4
26	56.2	21.9	21.9
27	50.0	20.0	30.0
28	53.7	29.2	17.1
29	52.3	6.8	40.9
30	55.9	11.9	32.2
Mean	57.3	17.6	25.1

Table 5

Subscales' Items and Reliability

	Subscale	Items	α Coefficient
1.	Reading	1, 2, 3	.64
2.	Initiating Research	6, 7, 26, 27	.78
3.	Data Gathering	8, 9, 10, 21	.74
4.	Data Analysis	11, 12, 30	.71
5.	Report Writing	14, 15, 16, 25, 28	.71
6.	Facilitating and Cooperating	4, 5, 13, 23, 24, 29	.68
7.	Liaison and Public Contacts	17, 18, 19, 20	.74

Table 6

Intercorrelations among Subscales

	1	2	3	4	5	6
1. Reading	-					
2. Initiating Research	.61	-				
3. Data Gathering	.33	.52	-			
4. Data Analysis	.45	.44	.55	-		
5. Report Writing	.37	.52	.49	.34	-	
6. Facilitating and Cooperating	.56	.51	.56	.17	.44	-
7. Liaison and Public Contact	.48	.53	.42	.21	.58	.61

Table 7

Correlations between Item Rating and Item Importance

for Supervisors and Subordinates*

Item	r (Superior rating)	r (Self-rating)
1**	.35	-.04
2**	.59	.19
3**	.55	.20
4**	.35	.01
5**	.42	.04
6	.40	.45
7	.24	.27
8**	.42	.00
9	.32	.16
10	.41	.25
11	.21	.08
12**	.35	-.10
13**	.32	-.05
14	.19	.03
15**	.47	.07
16	.37	.18
17	.33	.17
18	.21	.08
19	.38	.15
20**	.38	.00
21**	.37	.01
22**	.30	-.08
23**	.50	.05
24**	.35	-.02
25**	.34	-.05
26	.48	.33
27**	.58	-.18
28**	.56	.13
29	.15	.36
30**	.51	.08

* The importance rating scale was reversed.

** The differences between the two correlations were significant at the .05 level
or above.

Table 8

Multiple Regression Analysis of Subscales on the Global Assessment

Criterion - Superiors' Assessment

Predictor	R	R^2	r	Significance of Predictor	Overall Significance
1. Data Gathering	.83	.69	.83	.000	.000
2. Facilitating	.85	.72	.59	N.S.	.000
3. Reading	.86	.74	.54	N.S.	.000
4. Report Writing	.87	.76	.71	N.S.	.001
5. Data Analysis	.87	.76	.34	N.S.	.003
6. Liaison	.87	.76	.76	N.S.	.01
7. Initiating Research	.87	.76	.64	N.S.	.025

Table 9

Multiple Regression Analysis of Subscales on the Global Assessment

Criterion - Self Assessment

	R	R^2	r	Significance of Predictor	Overall Significance
1. Report Writing	.60	.37	.60	.002	.002
2. Data Analysis	.69	.48	.30	.04	.001
3. Liaison	.74	.55	.31	.05	.001
4. Initiating Research	.76	.58	.28	N.S.	.002
5. Data Gathering	.77	.59	.08	N.S	.004
6. Reading	.77	.60	.30	N.S.	.009
7. Facilitating	.77	.60	.32	N.S.	.02

A P P E N D I X 1

ITEMS INCLUDED IN THE PILOT RUN

APPENDIX 1

CODE No.:   RANK:

1.   Rate the incumbent's overall performance during the last
      year on the following scale:

      1.   Excellent

      2.   Above average

      3.   Average

      4.   Below average

      5.   Very poor

2.   Rank the incumbent's overall performance during the last
      year in comparison to the total group of NIPR research
      and technical staff.

      1.   Top 10%

      2.   Between 70 - 90%

      3.   Between 30 - 70%

      4.   Between 10 - 30%

      5.   Lower 10%

A P P E N D I X 2

UPDATED PERFORMANCE EVALUATION SCALE

APPENDIX 2

PERFORMANCE EVALUATION FORM INSTRUCTIONS

This form contains items dealing with various aspects of job performance.
You are requested to complete the form as follows:

1. Go over every item and rate its importance to the overall job
    performance of the incumbent on a four-point scale, as follows:

1	2	3	4
not important or not applicable	low importance	moderate importance	highly important

    Mark your rating under heading "importance" next to every item.

2. For every item circle one of the three statements which best
    describes the incumbent's behaviour on the job (leave non-applicable
    items blank).

3. You may add verbal comments in the appropriate space next to each
    item wherever necessary.

4. In your evaluation, refer to the entire period from the last
    evaluation until now.

5. Evaluate the incumbent on the basis of his observed behaviour as
    you are aware of it. Avoid items which did not lend themselves to
    direct observation during the period from last evaluation.

6. After you have completed the form make sure that the incumbent has
    completed the similar self-assessment form.

7. Set a time to meet the incumbent for a feedback session.

8. During the feedback session, go over the form with the incumbent,
    compare your assessment to his self-assessment and discuss constructive
    ways by which needed improvements may be made. It is recommended to
    discuss specific goals for the incumbent by which he may be able to
    improve his performance. Be prepared to furnish behavioural examples
    to support your ratings for the different items.

9. After you have given feedback to all your subordinates, transfer the
    completed forms to the administrative person responsible.

PERFORMANCE SELF-EVALUATION FORM INSTRUCTIONS

This form contains items dealing with various aspects of job performance. You are requested to complete the form as follows:

1.   For every item circle one of the three statements which best describes your behaviour on the job.

2.   You may add verbal comments in the appropriate space whenever necessary.

3.   In your evaluation refer to the entire period from the last evaluation until now.

      Your supervisor will evaluate you on a similar form. Later you will have an opportunity to see your superior's evaluation and to discuss your performance with him.

		Importance 1-4	Comments

	1. Displays extensive familiarity with the literature in areas directly related to his/her tasks as well as in other areas. 2. Shows familiarity with the literature mostly in areas directly related to his/her task. 3. Has no familiarity with the literature in most areas.	□	
1.			

	1. The literature review part in his/her reports is adequately comprehensive and thorough. (Applies only to reports which should include literature review.) 2. The literature review in his/her reports is not very extensive but most of the important points are covered. 3. The literature review in his/her reports is frequently skimpy and inadequate.	□	
2.			

	1. Whenever discussing the literature orally or in writing he/she is critical and analytical (draws sensible conclusions, pinpoints misconceptions, integrates different pieces, identifies major issues and concepts). 2. When discussing the literature he/she makes some attempt to be critical and analytical but only with moderate success. 3. When discussing the literature he/she emphasises particular details rather than integrating and analysing themes.	□	
3.			

	1. Consults other colleagues on a wide variety of professional issues. 2. Consults other colleagues on one or two areas with which he is particularly familiar. 3. Normally does not consult other colleagues in any area.	□	
4.			

		Importance 1-4	Comments

5.	1. Often initiates professional discussion with colleagues and offers advice on his own initiative. 2. Doesn't normally initiate professional discussions but is always willing to discuss professional issues with others when asked. 3. Isn't always enthusiastic about spending time on discussing professional issues of other people.	☐
6.	1. Often discusses or suggests research ideas for the future. 2. Sometimes suggests research ideas for the future. 3. Never talks about possible research ideas for the future.	☐
7.	1. Handles the initial phases of problem definition sample selection and experimental design almost entirely independently. 2. Runs into problems in the initial phases of a research project and needs some help from others. 3. Relies heavily on others for help in the initial phases of research project.	☐
8.	1. When constructing a new research instrument he/she is very thorough and rigorous in searching for items and establishing reliability. 2. Reasonably thorough and rigourous in constructing a new research instrument. 3. When developing a new research instrument he/she does not make enough effort to construct new items and establish proper reliability.	☐

37.

		Importance 1-4	Comments
9.	1. When conducting an interview he/she manages to obtain all the necessary information from the interviewee. 2. His/her interviews do not always result in obtaining all the necessary information. 3. His/her interviews results are frequently partial and incomplete.	☐	
10.	1. Is very familiar with a wide range of statistical procedures, can plan and perform data analysis almost independently. 2. Has a limited statistical background, needs some help in selecting statistical procedures for his research. 3. His/her familiarity with statistical procedures is severely lacking, cannot handle data analysis without extensive help.	☐	
11.	1. Always interested in improving his/her data analysis skills (takes courses, probes colleagues, reads etc.). 2. Is not very enthusiastic about improving his/her data analysis skills and learns new things only when necessary. 3. Does not make any noticable attempt to expand his/her statistical and methodological knowledge.	☐	
12.	1. Very articulate and clear in discussing professional matters with colleagues. 2. Reasonably articulate and clear in discussing professional matters with his/her colleagues. 3. His/her professional communication with colleagues is not always very clear and sometimes he/she is difficult to follow.	☐	

		Importance 1-4	Comments
13.	1. His/her writing style is very concise and clear. 2. His/her writing style is reasonably concise and clear. 3. His/her writing style lacks conciseness and clarity.	☐	
14.	1. Published good quality papers. 2. His/her published papers are not of high quality. 3. Has not published any papers.	☐	
15.	1. His/her reports are very thorough and reflect high quality performance. 2. His/her reports are fairly thorough and reflect reasonably good performance. 3. His/her reports are frequently superficial and reflect less than adequate performance.	☐	
16.	1. Clients (trainees, outside organisations, counselling clients, etc.) are normally very impressed with him/her. 2. Clients respond reasonably well to him/her. 3. Clients are frequently dissatisfied with him/her.	☐	

39.

		Importance 1-4	Comments
17.	1. Very enthusiastic about establishing communications with outside clients. 2. Willing to communicate to outside clients when necessary but not very enthusiastically. 3. Sometimes tries to avoid direct contact with outside clients.	☐	
18.	1. Always meets deadlines when applicable to his work. 2. Sometimes fails to meet deadlines and needs an extension. 3. There are frequently severe delays and failure to meet deadlines on his part.	☐	
19.	1. Always prepared to co-operate with others on team work tasks. 2. Sometimes is reluctant to co-operate with others on teamwork tasks. 3. Frequently is not very co-operative on teamwork tasks.	☐	
20.	1. Always willing to take over someone else's task when it is necessary. 2. Sometimes reluctant to take over someone else's task when it is necessary. 3. Frequently attempts to avoid taking over someone else's task when it is necessary.	☐	

		Importance 1-4	Comments
21.	1. His/her project proposals are very well organised. 2. His/her project proposals are reasonably organised although sometimes there is a need for further elaboration. 3. His/her project proposals are not well organised thoroughly and frequently require further expansion	☐	
22.	1. Shows initiative and is assertive in attempting to organise samples for projects. 2. Shows moderate initiative and assertiveness in attempting to organise samples for projects. 3. Lacks initiative and assertiveness in attempting to organise samples for projects.	☐	
23.	1. When suggesting a new project he/she has a good conception of its objectives and has a clear step-by-step plan. 2. When suggesting a new project his/her ideas regarding its objectives and the methodology involved are not always fully developed. 3. When suggesting a new project he/she frequently does not have a clear understanding of its specific objectives and the methodological steps involved.	☐	
24.	1. Enthusiastic about writing reports and performs the task promptly. 2. Not very keen on writing reports and sometimes delays it on behalf of other activities. 3. Dislikes report writing and needs to be pushed to complete the task.	☐	

		Importance 1-4	Comments
25.	1. When assisting other colleagues (for example, on methodology or computer services) he/she conveys his ideas very clearly on their level. 2. There are sometimes indications that he/she does not manage to communicate well with other colleagues when assisting them on professional matters. 3. Frequently has communication difficulties when assisting colleagues on professional matters.		
26.	1. His/her data analysis is very comprehensive and covers all the relevant aspects. 2. Tends to analyse the major aspects but sometimes leaves certain issues unexplored. 3. Tends to overlook relevant possibilities in his data analysis.		

HSRC PUBLICATION LIST

A complete list of HSRC publications or a list of publications of a particular institute of the HSRC is available from the Publications Manager

HUMAN SCIENCES RESEARCH COUNCIL
RAAD VIR GEESTESWETENSKAPLIKE NAVORSING

Functions of the HSRC

The HSRC undertakes, promotes, supports and co-ordinates research in the field of the human sciences. It also determines research priorities, disseminates the findings of human sciences research, promotes and evaluates the implementation of research findings, stimulates the training of researchers, places the full spectrum of human sciences disciplines at the service of the inhabitants of the RSA and promotes science in general.

Funksies van die RGN

Die RGN onderneem, bevorder, ondersteun en koördineer navorsing op die gebied van die geesteswetenskappe, bepaal navorsingsprioriteite, versprei die resultate van geesteswetenskaplike navorsing, bevorder en evalueer die implementering van die resultate van navorsing, stimuleer die opleiding van navorsers, stel die volle spektrum van dissiplines in die geesteswetenskappe ten diens van die inwoners van die RSA en bevorder die wetenskap in die breë.

Institutes

Institute for Communication Research (ICOMM)

Institute for Educational Research (IER)

Institute for Historical Research (IHR)

Institute for Manpower Research (IMAN)

National Institute for Personnel Research (NIPR)

Institute for Psychological and Edumetric Research (IPER)

Institute for Research Development (IRD)

Institute for Research into Language and the Arts (IRLA)

Institute for Sociological and Demographic Research (ISODEM)

Institute for Statistical Research (ISR)

Bureau for Research Support Services (BRSS)

Administration

Institute

Instituut vir Geskiedenisnavorsing (IGN)

Instituut vir Kommunikasienavorsing (IKOMM)

Instituut vir Mannekragnavorsing (IMAN)

Instituut vir Navorsingsontwikkeling (INO)

Instituut vir Opvoedkundige Navorsing (ION)

Nasionale Instituut vir Personeelnavorsing (NIPN)

Instituut vir Psigologiese en Edumetriese Navorsing (IPEN)

Instituut vir Sosiologiese en Demografiese Navorsing (ISODEM)

Instituut vir Statistiese Navorsing (ISN)

Instituut vir Taal- en Kunstenavorsing (INTAK)

Buro vir Ondersteunende Navorsingsdienste (BOND)

Administrasie

Head office

Private Bag X41, Pretoria 0001
Republic of South Africa
Telegrams RAGEN
Tel. (012) 28-3944
Telex 3-0893

Hoofkantoor

Privaatsak X41, Pretoria 0001
Republiek van Suid-Afrika
Telegramme RAGEN
Tel. (012) 28-3944
Teleks 3-0893

NIPR

P.O. Box 32410 Braamfontein 2017
Republic of South Africa
Telegrams NAVORSPERS
Tel. (011) 33-94451
Telex 4-25459

NIPN

Posbus 32410, Braamfontein 2017
Republiek van Suid-Afrika
Telegramme NAVORSPERS
Tel. (011) 33-94451
Teleks 4-25459

Regional offices

Western Cape, Private Bag, 40, Parow 7500
Tel. (021) 92-1026

Natal, P.O. Box 508, Durban 4000
Tel. (031) 31-6926

NIPR Natal, P.O. Box 17001, Congella 4013
Tel. (031) 25-5531

NIPR Eastern Cape, P.O. Box 1124, Port Elizabeth 6000
Tel. (041) 53-2131

Streekkantore

Wes-Kaap, Privaatsak 40, Parow 7500
Tel. (021) 92-1026

Natal, Posbus 508, Durban 4000
Tel. (031) 31-6926

NIPN Natal, Posbus 17001, Congella 4013
Tel. (031) 25-5531

NIPN Oos-Kaap, Posbus 1124, Port Elizabeth 6000
Tel. (041) 53-2131