

Performance Appraisal

I Performance Appraisal: Determination and Documentation of an Individual's Performance on the Job

II The Uses of Performance Appraisal

A) Administrative decisions.

- 1) Assigning work more efficiently
- 2) Training
- 3) Setting Raises
- 4) Promotion & Demotion
- 5) Firing

B) Employee development and feedback

- 1) Meeting employees' needs for growth.
- 2) Assisting employees in setting career goals & in developing their work-related skills.
- 3) Keeping employees advised of what is expected of them.
- 4) Warning about unacceptable job performance.
- 5) Improving performance of employees.

C) Criteria for research

- 1) Testing whether the selection procedures actually are picking the best employees (Validity of Selection Procedures).
- 2) Testing the effectiveness of training programs.
- 3) Testing the effectiveness of other interventions.

D) Documentation for legal action

- 1) For use in defense against claims of wrongful termination or other wrongful personnel action.

E) Training

- 1) Identifying training needs.

III An Important Distinction:

A) The Ultimate or Theoretical Criterion (Criteria): A conceptual definition of performance. What do we mean by good job performance?

B) The Actual Criteria:

- 1) The *operational* definition of performance. How performance is actually measured. What do we do to get a good measurement of performance?

IV The Nature of the Theoretical Criterion

A) What is performance?

B) Single or Multiple Dimensions?

1) A Single Criterion: A *global* measure of performance.

2) Multiple Aspects of Performance:

(a) Composite Criterion: Take all the different components of performance, and combine them into a single score.

(i) How should they be weighted?

(b) Multidimensional Criteria

(i) Keep each subcomponent separate.

(ii) How do you make global judgements?

C) What are the elements of the Theoretical Criterion?

1) Task Performance: How well you do the tasks that are part of your job.

(a) The importance of Job Analysis

2) Contextual Performance: All other activities that help the organization. Activities performed by employees that help to maintain the broader organizational, social, and psychological environment.

(a) Job Dedication

(i) Organizational Citizenship: The willingness to go beyond the formal requirements of the job and to freely give time and energy to help promote organizational goals.

(ii) Commitment or loyalty to the organization.

(iii) Personal Initiative

(iv) Motivation

(b) Interpersonal Facilitation

(i) Helping Others

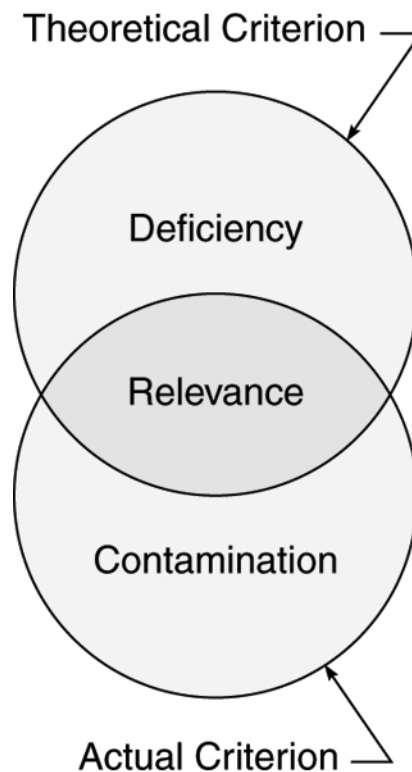
(ii) Maintaining Positive Social Relationships in the Workplace

(c) Job dedication is correlated with task proficiency, but interpersonal facilitation is more or less independent.

3) Counterproductive Behaviors: Behaviors that bring or are intended to bring harm to an organization.

V Evaluating Criteria

- A) Once we have a good *conceptual definition* of the criterion (the theoretical criterion), we will have to come up with real ways to measure the elements of performance that are important (the *actual criteria*).
- B) Our actual criteria are only good if they really capture the important aspects of the theoretical criterion. This is called the *relevance* of the criteria.
- C) We also have to make sure that the actual criteria capture all of the dimensions of the theoretical criterion. If not, our actual criteria are said to be *deficient*.
- D) In addition, we want to make sure that our actual criteria measure only elements of performance (the theoretical criterion), and not other things. If the actual criteria capture things not relevant to performance, then we say the actual criteria are *contaminated*.



E) Relevance is similar to *construct validity*.

1) Construct validity: The scores on a measurement can really be said to be measuring what we say they are measuring.

(a) In this case, it's the scores on the actual criteria are really measuring the theoretical criteria.

(b) Ways to assess construct validity:

(i) Convergent Validity: Does the measure correlate with other measures of the same trait.

(ii) Divergent Validity: Does the measure *not* correlate too highly with measures of *different* traits.

1. For example, an IQ measure should correlate more highly with another measure of IQ (convergent validity) than it does with a measure of memory.

(iii) Criterion Related Validity: The measure correlates with things that the construct being measured should correlate with.

1. For example, do scores on an IQ test correlate with school performance.

2. Or, do scores on an integrity test correlate with disciplinary actions.

(iv) Content Validity: Does the measure seem to represent all of the different aspects of the trait being measured.

2) What type or types of validity are related to contamination and deficiency?

Methods of Performance Appraisal

VI Behaviors versus Outcomes: What should be assessed?

- A) “Behaviors” means what the employee *does*.
 - 1) Sales calls made (and the things done on the call).
 - 2) Research papers submitted (and content of paper).
 - 3) Procedures followed on the assembly line.
- B) “Outcomes” refers to the results of the worker’s activities.
 - 1) Sales made.
 - 2) Papers published.
 - 3) Days without accidents.
- C) Outcomes versus Behaviors and Criterion Contamination.
 - 1) The focus should be on things that are under the worker’s control.

VII Four Types of Performance Appraisal Measures

- A) Objective Measures: Quantitative information taken from the records of employee performance (e.g., attendance, sales figures, disciplinary action etc).
- B) Tests of Job Performance (Work Sample Tests)
 - 1) Samples of work-type activities, but done under standardized conditions so that different people’s performance can be compared without contamination.
- C) Tests of Job Knowledge
- D) Performance Ratings (May be used in over 75% of performance evaluation situations).

VIII Objective Measures.

A) Are objective measures really objective?

- 1) Sometimes different things will be counted as an absence, or will be reported as a disciplinary action for different people.

- (a) Judgment is involved even here.

B) Absences & Lateness

- 1) Relevance?

- 2) Deficiency?

- 3) Contamination?

- 4) The problem of truncation of sample:

- (a) Many of the goals of performance appraisal require that different workers be evaluated relative to each other.

- (i) This requires *variability*: All of the employees being evaluated need to differ from each other.

- (b) With absences, only a few people will have a lot of absences. Most people will have only a few (unexcused) absences in an evaluation period.

- (i) Therefore, absences can only be used to differentiate among those few people who have differing amounts. All the rest will be given the same rating.

- (c) This problem is called truncation of sample because many of the people in the sample will be 'cut off' (Truncate = "To shorten by or as if by cutting off." <http://dictionary.com>).

C) Disciplinary Action or Disciplinary Reports

- 1) This can be appropriate in some situations, where formal disciplinary actions are reported (e.g., military), but many jobs don't have formal disciplinary reports.

- 2) Not all supervisors will use disciplinary actions in the same way. For example, some supervisors may report every infraction – even minor ones. Other supervisors may just handle minor things by talking to the worker.

- (a) An example of the subjective contribution to objective measures.

- (b) Who is the supervisor *contaminates* this criterion.

- 3) Truncation of sample is also a problem here.

4) Relevance? Deficiency & Contamination?

D) Production Rates: Counting how much is produced.

1) Relevance?

2) Contamination?

3) Deficiency?

4) Another problem: For some jobs there is nothing to count.

(a) For example, human resource director; Psychology professor.

E) Sales

1) When appropriate, this seems like a really good measure. (Relevance)

2) But, sales can also depend on many external factors (sales territory; time of year; store size, etc.). When these factors influence different people differently, then this can be a problem. (Contamination).

(i) Why doesn't it matter if everyone is being affected equally?

3) Deficiency?

F) Job Level & Salary Obtained

1) Seems like a good measure for managerial and professional staff.

(a) If merit is the basis of promotion, then this may reflect the overall evaluation of the employees work.

2) But, certain things need to be kept in mind:

(a) This measure must be adjusted for length of time in the job.

(i) Two people who are both promoted to Regional Manager may not be equally good employees if one has been working for 5 years and one for 10.

(ii) Can be strongly influenced by external factors such as the availability of positions and internal politics.

IX Tests of Job Performance (Work Sample Tests)

- A) This means a test that is set up to have the person do the tasks that are part of the job, but it is set up so that there is a clear-cut way to ‘score’ the performance on the job.
- B) The development of these kinds of tests needs to be done very carefully.
- 1) The first step is to develop (with job analysis!) a list of all the tasks that are involved in doing a job.
 - 2) Then, you have to have break these tasks down into the steps that each task takes.
 - 3) From this, you have to get the tasks that you will actually use in the test.
 - (a) These should be the tasks that are important for the job.
 - (b) Sometimes, it may not be feasible to have a certain task on the test
 - (i) because it takes too long, or
 - (ii) because the equipment may not be available
 - (iii) etc
 - (c) But, it might be possible to find some aspects of the task that reflect the same domain of knowledge, skills and abilities that are required to complete the job.
 - (i) The tasks should have *psychological fidelity*, even if they don’t have *physical fidelity*.
 - 4) After deciding on the tasks to be used, there needs to be some objective scoring criteria.
 - (a) You need to train scorers. Even if they are subject-matter experts, they need to be trained to make the scoring objectively.
 - (i) Scoring must have *intersubjectivity*.
 - (i) To see if the scoring does have intersubjectivity, you would look at *inter-rater reliability* (“The association between the ratings of two or more raters who rate the same subject on the same variable.”, Spector, p. 285)
 - (b) Scoring can be focused on the *process* or on the *product*.
 - (i) Product focused scoring can be more objective, but if the way in which someone does the task is important, then process oriented scoring is necessary.
- C) Work sample tests are probably not adequate as an only form of performance evaluation.
- 1) First of all, they are clearly maximum performance measures, rather than typical performance.

- 2) Secondly, while they are good for measuring tasks that can be clearly specified and clearly scored, not all tasks have such qualities. (For example, how would you include a test of a manager's job of encouraging her employees...).
- 3) Also, the limitations of what can be measured due to lack of time or lack of expensive or hazardous equipment limit the usefulness of work samples as the only method of performance evaluation.
 - (a) What category would this problem fall under?

D) Of course, if you are going to use a work sample test, its validity & reliability must be checked.

X Job Knowledge Tests

- A) Knowledge of things needed to know to perform the job or test of how you would do the job.
- B) Again, a good list of all the tasks required for the job is necessary. Then, items can be written that evaluate what you would do on the job.
- C) This method is good for jobs where knowledge about the steps to complete a task is important. But, it wouldn't be good where motor skills are particularly important.

XI. Performance Ratings: Employees' Performance is rated on a scale or questionnaire.

- A. Most commonly used method of performance appraisal.
- B. Referred to as *subjective* measure, because the (subjective) opinions of the raters affect ratings.
- C. Who does the rating?
 1. Usually ratings are done by supervisors. However, it is often good to get ratings from a variety of different people.
 2. The 360 Degree Feedback Approach
 - (a) This method is designed to get ratings from people all around (360 degrees) the employee.
 - (b) Ratings can be obtained from supervisors, subordinates, peers (others in the same or similar job), and even customers. Self ratings are usually included.
 - (c) This method can get ratings from different perspectives, and can help avoid biases because different people will have different biases.

D. Frequently (usually) uses a graphic rating scale, such as:

Name _____ Rating period from _____ to _____ Rater name _____ Rater title _____ Department _____		Rating Scale Key <input type="checkbox"/> 1 Fails to meet job requirements <input type="checkbox"/> 2 Meets most job requirements <input type="checkbox"/> 3 Fully meets all job requirements <input type="checkbox"/> 4 Exceeds all job requirements					
Part II: Rating Scales for Task Areas							
Position: Administrative Secretary Duties and Responsibilities							
Duty: Maintaining Departmental Personnel Records		PCT. (30%)	RATING:	1	2	3	4
				<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Using spreadsheets and other computerized forms: Ensure all recorded personnel information is up to date and accurate, including name, address, and contact information; maintain accurate records of all personal leave taken annually, as well as vacation schedules; maintain accurate records of all employees' benefits.			Comments				

E. The best performance rating scales use *behaviorally focused formats*.

1. This means that there are specific behaviors indicated on the scales to show raters what performance at that level means.
2. Use of a behaviorally focused format allows the raters to do a better job of being objective and consistent.
3. One example is the Behaviorally Anchored Rating Scale (BARS).
 - (a) The first step in creating a BARS is to do a job analysis to find out all of the dimensions of the job.
 - (b) Next, the Critical Incidents Technique is used to develop examples of effective and ineffective performance in each job dimension.
 - (c) Then, the incidents need to be judged for effectiveness so they can be placed on the appropriate place on the scale.

SALESMANSHIP SKILLS

Skillfully persuading prospects to join the navy; using navy benefits and opportunities effectively to sell the navy; closing skills; adapting selling techniques appropriately to different prospects; effectively overcoming objections to joining the navy.

- | | |
|---|---|
| 9 | A prospect stated he wanted the nuclear power program or he would not sign up. When he did not qualify, the recruiter did not give up; instead, he talked this young man into electronics by emphasizing the technical training he would receive. |
| 8 | The recruiter treats objections to joining the navy seriously; he works hard to counter the objections with relevant, positive arguments for a navy career. |
| 7 | When talking to a high school senior, the recruiter mentions names of other seniors from that school who have already enlisted. |
| 6 | When an applicant qualifies for only one program, the recruiter tries to convey to the applicant that is a desirable program. |
| 5 | When a prospect is deciding on which service to enlist in, the recruiter tries to sell the navy by describing navy life at sea and adventures in port. |
| 4 | During an interview, the recruiter said to the applicant. I'll try to get you the school you want, but frankly it probably won't be open for another three months, so why don't you take your second choice and leave now. |
| 3 | The recruiter insisted on showing more brochures and films even though the applicant told him he wanted to sign up right now. |
| 2 | When a prospect states an objection to being in the navy, the recruiter ends the conversation because he thinks the prospect must not be interested. |
| 1 | |

F. Subjective Performance Evaluations can lead to a lot of errors.

1. The notion of *error* has to be contrasted with *accuracy*.

(a) There is some true level of performance on each dimension of the job, and the goal is to get ratings of that. Anything that moves away from that is an error.

2. Types of Errors

(a) Halo effects: Raised or lowered ratings on all dimensions, when they don't really occur together.

(b) "Similar-to-me" Error: Assumption that person being rated is similar to the rater.

(c) Distributional Errors: Rating everyone the same way.

i Leniency errors: Rating everyone favorably

ii Severity errors: Rating everyone strictly.

iii Central Tendency errors: Giving everyone an average rating.

3. A number of these errors can be avoided by training the raters.
 - (a) It is important, though, if you are training raters to avoid a specific error, to assure that you don't introduce new errors.
 - i For example, if you are training people to avoid the halo effect, you don't want to introduce the error of overly spreading out people's ratings to avoid that effect.
 - ii For this reason and maybe others, some research shows that rating training may reduce errors, but it also reduces accuracy.
 - (i) It may be that the rating patterns (like Halo Effect or Leniency Effect) aren't as much of an error as thought. Maybe performance really *does* all go together, for example.