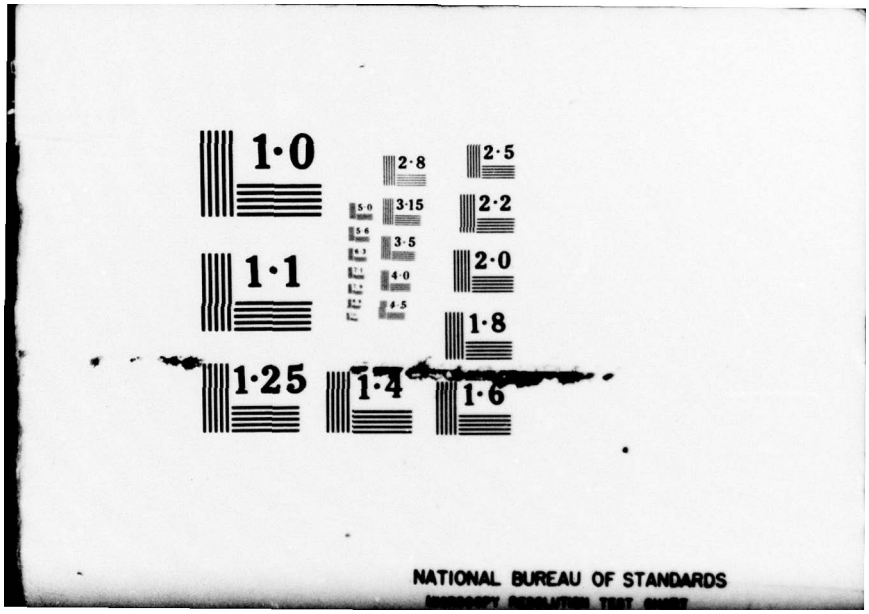
AD-A051 673 HUMAN RESOURCES RESEARCH ORGANIZATION ALEXANDRIA VA F/G 5/9
ALTERNATIVES TO PERFORMANCE TESTING: TESTS OF TASK KNOWLEDGE AN--ETC(U)
MAR 78 R VINEBERG, E N TAYLOR

UNCLASSIFIED HUMRRO-PP-6-78 NL

1 OF 1
ADA
051673

END
DATE
FILMED
4-78
DDC


1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5
5·0
NATIONAL BUREAU OF STANDARDS

HumRRO

# (6) Alternatives to Performance Testing: Tests of Task Knowledge and Ratings.

(10) Robert Vineberg and Elaine N. Taylor

(11) A[redacted] presented at a Conference on Defense Manpower Santa Monica, Ca[redacted] Feb[redacted]76,

(11) Feb 78

(12) 5 p.

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

(11) Mar[redacted]78

405 260

## PREFATORY NOTE

This paper is based on a presentation by Dr. Robert Vineberg and Dr. Elaine N. Taylor of HumRRO's Carmel Research Office at a Conference on Defense Manpower. The Conference, which was held at Santa Monica, Calif., in February 1976, was hosted by The RAND Corporation for the Defense Advanced Research Projects Agency (DARPA).

The Vineberg-Taylor paper is not based on any single HumRRO research project, but presents ideas generated in the course of several such projects. It is being packaged as a HumRRO Professional Paper to make the information more widely available than would otherwise be possible.

# ALTERNATIVES TO PERFORMANCE TESTING: TESTS OF TASK KNOWLEDGE AND RATINGS USING BEHAVIORAL ELEMENTS AND TASKS AS ENTITIES

Robert Vineberg and Elaine N. Taylor

I am going to talk about a combination of methods for assessing job proficiency that we have been working on at HumRRO. Part of the work is supported by ONR and part by the Army.

A convenient way to classify alternative approaches to assessing task or job proficiency is to consider what is being measured in terms of its remoteness from actual performance. Looking at what is being measured in this way we can identify at least five general strategies. In decreasing order of fidelity from actual task or job performance they are:

1. Measurement of performance in the actual job situation where the only change is recognition that a test or measurement is going on;
2. Measurement of performance on job sample tests, sometimes in an approximation of the job environment;
3. Measurement of performance using simulations involving varying degrees of degradation of the stimulus and/or response aspects of the actual performance;
4. Measurement, not of performance but rather of information about how a task or job is to be performed—knowledge that should correlate with actual performance; and finally
5. Neither the direct measurement of an incumbent's performance or of his knowledge but rather the appraisal by a second party, usually a supervisor or sometimes a peer, of how a person carries out his job.

Now, consider these methods. The first, the measurement of actual performance in the job is rarely done, in the sense of measuring actual job activities or processes. It is seldom feasible, there are problems of standardization and the cost is extremely high. The measurement of job output or the product of job performance, while limited to tasks that generate a permanent and objective record, and also facing problems of standardization, probably occurs somewhat more frequently.

The second method of assessing performance, job sample testing, is probably as close to the ideal as we can get from a measurement point of view. But as you know it is also extremely costly.

The third method, using some form of simulated job measures, is probably more feasible but can be quite risky given our lack of systematic knowledge about the properties of cues and responses that must be represented in the criterion situation in order to obtain valid measurement. While the adequacy of simulations for training has, of course, been studied through transfer experiments, there has been virtually no analysis of the use of simulations as criteria for assessing job performance.

Fourth, the measurement of job knowledge, while the most feasible and least costly approach to direct measurement, suffers in that it often does not provide an adequate

correlation with actual performance. However, knowledge tests can correlate fairly well with performance if they are constructed with care and if they cover content that is clearly relevant to performance. In a study that we conducted some years ago in which we administered knowledge tests and lengthy job sample tests to over 1,600 job incumbents in four different Army jobs, we obtained correlations between job sample and knowledge test scores ranging from .58 to .72.

And last, the use of ratings, while clearly the easiest and most frequently used method probably correlates least well with any of the direct measures of performance—a shortcoming that can be ascribed to difficulties of maintaining objectivity with indirect measurement. Also, ratings have often been fairly non-specific, perhaps intentionally, about the tasks or behaviors that comprise a job.

Our present work has focused on the last two methods, knowledge testing and ratings, the most feasible but, of course, the most remote from the job.

First we will consider knowledge testing. This is the work we are doing for the Army. It is based upon two notions: first, knowledge testing should focus on the performance of specific tasks and should consist of items that possess all or most of the knowledge that is relevant to the performance of these tasks. It should not consist of items of general job knowledge or individual elements of knowledge that have been isolated from the totality of information needed to perform a task. Second, knowledge testing should be restricted to tasks that do not involve skilled behavior.

A simple test of whether a task is skilled or non-skilled is to describe it in detail to a naive person. If he can perform it, the task is non-skilled and a knowledge test may be used. Examples of such tasks are changing a tire, dialing a long distance call, or keeping score in golf.

Skilled behaviors, on the other hand, require practice or rehearsal during learning. Examples are aiming a rifle at a moving target, manipulating materials with a crane, or hitting a golf ball where you want it to go. Practice is required in learning such behaviors for a variety of reasons—to discover the specific movements or actions themselves; to make adjustments in the behaviors; to gain speed, coordination or timing; and occasionally to provide for overlearning so there will be stability of performance under conditions of stress. While practice may accomplish different things during learning, the role that practice plays is not important for purposes of test construction. The mere fact that practice is required to learn a task is sufficient to classify that task as skilled and to indicate that something other than a knowledge test is needed. Even when it is possible to describe a skilled task verbally, such a description cannot be expected to impart that skill to another person. Likewise, a verbal description of skilled behavior given by a job incumbent cannot be used to infer that the job incombent providing the description can indeed perform the task.

Parenthetically, I should add that there are some tasks that require practice during learning but that are not properly classified as skilled. These are tasks that are perfectly communicable by verbal means but which are so lengthy or complex as to require several trials to be committed to memory. While there may be some practical problems in testing these tasks on the basis of information about them, they are measurable—in theory at least, with knowledge tests.

Now let us return to the notion that knowledge tests should focus on specific tasks and contain all or most of the information required for performing these tasks. As far as we know, this approach has never been attempted in any systematic way. However, over the last year or so the Army has been engaged in initiating a new system of testing to determine a soldier's job proficiency and whether he is qualified for promotion. In this program, referred to as Skill Qualification Testing, the Army has focused its attention on the tasks that it deems critical in each job. The emphasis is upon task performance and if it were possible the Army would use performance tests entirely.

We have just completed writing a manual for the Army on procedures to be followed in constructing both performance tests and knowledge tests of tasks to be used as Skill Qualification Tests. Both types of tests begin with the same materials: a detailed listing of the behavioral elements of a task. In the case of a knowledge test, these elements are then translated into descriptive information that mediates their performance. It is interesting to note that if a task has been properly analyzed and if a performance test is to be constructed, no further breakdown of the elements of the task is necessary. They translate directly into observable measures of performance. However, in the case of a knowledge test, many elements must be further broken down into finer sub-elements than those that emerge from the task analysis. For example, in adjusting the hydraulic brake on an M-60 tank, one of the behaviors is to "loosen both jam nuts on the brake pedal-master cylinder tie rod." The separate bits of knowledge that mediate the performance of this particular step are at least:

a. knowing the location of the tie rod
b. knowing the appearance of the jam nuts on the tie rod
c. knowing that the jam nuts need to be loosened.

As a matter of fact, while there are ten steps that should be observed in measuring adjustment of the hydraulic brake, there are at least 36 separate items of knowledge that can be identified. To find out if a job incumbent indeed knows what to do and how and when to do it, we need to assess almost all of these knowledges.

To keep the number of knowledge items within a reasonable limit, however, we have suggested a half dozen more or less common sense rules for sampling items that seem likely to give adequate coverage of a total task. For example, we recommend the use of items that test a knowledge of sequence when a task is procedural and consists of a large number of steps but recommend items solely about the content of individual steps when the task is short. In maintenance tasks, questions about actions and standards take priority over questions about the location of parts per se since knowledge of the location can often be assumed if a person knows what actions to take or what standards to meet.

Assuming that we can test proficiency for non-skilled tasks through these specially devised knowledge tests of entire tasks, what can be done short of performance testing for assessing proficiency in tasks that are skilled. Our work for the Navy is perhaps relevant here.

In this research we are trying to device methods for rating performance that may be more discriminating than traditional ratings. To do this we are exploring ways to be more elemental or specific in the rating process and are comparing two models of job analysis to arrive at more molecular descriptions.

In describing different approaches that have been taken to job analysis, Ernest McCormick has distinguished between "worker-oriented" descriptions of tasks and "job-oriented" descriptions of tasks.[1] Worker-oriented descriptions focus upon human behaviors that can be generalized across tasks. Job-oriented descriptions tend to focus upon job content that is characterized in terms of the specific technological objects of performance or achievements of the worker.

Examples of generalizeable elements or behaviors are "estimating the speed of moving objects," "obtaining information from written materials," "engaging in information exchange," or "activating fixed setting controls."

[1] McCormick, Ernest J., Jeanneret, Paul R., and Mecham, Robert C. "A Study of Job Characteristics and Job Dimensions as Based on the Position Analysis Questionnaire (PAQ)." *Journal of Applied Psychology*, Vol. 56, No. 4, 1972, pp. 347-368.

Examples of the more task specific job-oriented descriptions are "repairs coaxial cables," "anneals cooper tubing," "uses wiring diagrams," or "drafts business letters."

To develop rating procedures based upon a worker-oriented model, we used McCormick's Position Analysis Questionnaire to obtain job analysis data for 10 Navy jobs that we believe are quite different. The Position Analysis Questionnaire is an instrument for rating the relevance of 189 possible worker-oriented behaviors in a job.

We have also taken all of McCormick's items for describing the structure of jobs and translated them into items suitable for rating performance. We have constructed performance rating questionnaires for each of the 10 Navy jobs that contain only items for those elements that were identified as most relevant in the job analyses.

Our next step will be to ask supervisors (and perhaps peers) to rate the performance of men in these jobs with respect to these specifically selected elements of behavior. Depending upon the job, a man will be rated on 30 to 60 worker-oriented elements.

Now let us consider our other approach to performance ratings. To develop rating procedures based upon a job-oriented model we are using job task inventories that have been collected as apart of the Navy Occupational Training Analysis Program (NOTAP). From this program we have obtained lists of tasks performed by at least 50% of the incumbents in the jobs we are studying. We are now constructing job-oriented rating instruments using these specific tasks.

We will collect performance rating data with these instruments about the same incumbents from the same supervisors who used the worker-oriented instruments.

We anticipate that the worker-oriented ratings will distribute somewhat more normally than the job-oriented ratings. As you know, ratings generally have a tendency to pile up at the positive end of the scale. Since supervisors are responsible for insuring the effectiveness of their personnel, a poor rating can reflect upon a supervisor as well as an incumbent. Perhaps a supervisor can be more objective in his ratings when worker-oriented elements are taken from the entire job and when something less than perfect performance does not have to be viewed as failure in very specific tasks.

In our study we expect to compare the outcome from both kinds of instruments with the distributions obtained under the Navy's present performance rating system. We also plan to obtain some information on the concurrent validity of these instruments by comparing the performance of experienced and inexperience job incumbents.

Finally, as part of our work for the Army, we plan to conduct a study in which soldiers will take performance tests and knowledge tests and also be rated with worker-oriented and job-oriented instruments. This last study will give us our most definitive information about the efficacy of using knowledge tests of tasks, and worker-oriented or job-oriented ratings as substitutes for performance tests.