DOCUMENT RESUME

ED 393 879 TM 024 755

AUTHOR	Sheppard, Valarie
TITLE	Validation of Testing for a Multi-Disciplinary Technical Position.
PUB DATE	Jun 95
NOTE	12p.; Paper presented at the Annual Meeting of the International Personnel Management Association Assessment Council Conference (19th, New Orleans, LA, June 25-29, 1995).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Employment Qualifications; Job Analysis; *Occupational Tests; *Paraprofessional Personnel; Performance; *Performance Based Assessment; Safety; Scores; *Technical Occupations; *Test Construction; Testing Programs; Test Validity; Training
IDENTIFIERS	*Knowledge Skills and Abilities Method; Oral Examinations

ABSTRACT

This paper describes a testing program that established job-related standards of performance and testing procedures for the technician position within a gas transportation company. New standards were being established in response to safety concerns stemming from investigations that identified training inadequacy as a cause of accidents. The job title "Technician" in reality encompassed individuals in four separate disciplines, but considerable crossover of duties occurred. A job analysis was conducted to compile a comprehensive tasklist and to determine the criticality of tasks. To develop test items, it was then necessary to identify the knowledge, skills, and abilities (KSAs) needed to perform the critical tasks. Oral and performance tests were developed for the technician job, using the critical tasks and KSAs identified. The draft test was reviewed by workers in each discipline. Questions were developed to cover the overlapping task areas, and the test sections were administered to two persons from each discipline. Validation of the test was then accomplished through content and criterion-related strategies, with samples of 60 technicians. Cluster goal scores and minimum goal scores were then established. (Contains one table and one figure.) (SLD)

VALIDATION OF TESTING FOR A MULTI-DISCIPLINARY TECHNICAL POSITION

Valarie Sheppard
Human Performance Systems, Inc.

Paper presented at 1995 IPMA-AC Conference – New Orleans, LA

INTRODUCTION

In 1988 the National Transportation Safety Board (NTSB) found that the majority of the 110 safety recommendations issued between 1975 and 1986 the majority stemmed from pipeline accidents that were linked to inadequate training of personnel. Based on this finding the NTSB gave discretionary authority to the Department of Transportation to require pipeline companies to provide evaluation procedures for employees. Since that time the Department of Transportation Research and Special Programs Administration (DOT–RSPA) has published a document (49 CFR 192 and 195) entitled "Qualifications for Pipeline Personnel." The DOT–RSPA draft document expresses the need for qualification standards for personnel who perform regulated operation, maintenance, and emergency–response functions on gas and hazardous liquid pipelines.

After reviewing this document, the Office of Pipeline Safety (OPS) recommended qualification standards based on training and testing instead of licensing and certification programs. The proposed qualifications standards indicate that individuals who perform or directly supervise the performance of regulated operation, maintenance, or emergency response functions on pipelines must be trained and tested in the relevant areas related to pipeline safety. The OPS did not specify the method of training and testing, but left it to each company to submit a plan that outlines their training and testing program for approval.

These proposed regulations require that gas companies develop training and testing programs for workers involved in the operation and maintenance functions. When implemented such programs benefit the employer in terms of (1) an increase in worker effectiveness and productivity, and (2) a decrease in accidents and injuries.

This paper describes a testing program that establishes job–related standards of performance and testing procedures for the technician position within a gas transportation company.

OVERVIEW OF THE TECHNICIAN POSITION

The job title "Technician" may indicate one position, but in reality encompasses individuals who work in 4 separate disciplines. These disciplines are Communications, Controls, Corrosion, and Measurement, and each is responsible for performing different duties at the station.

Persons whose area of responsibility is Communications, are responsible for all station communication including telephone service, mobile radios and the microwave system. Tasks performed by Communications technicians include installing new phone lines, troubleshooting and programming the telephone system, and installing and servicing mobile radios. They are also responsible for the operation and maintenance of the microwave system which sends signals to different stations through microwave towers. These signals are divided into channels at each location and designated as voice or data. The technician is responsible for calibrating the microwave equipment to ensure the signal leaving and entering his station is at the proper level.

Technicians in the Control discipline are responsible for the operation and maintenance of the engines at the station. They perform tasks such as starting and stopping engines, calibrating equipment and conducting and/or overseeing engine repair. In the last 10 years the technical nature of the job has increased dramatically due to an increase in automated systems. This automation can vary from little more than a black box with a few buttons that control start-up and shut down of the engines, to elaborate computer systems that control all aspects of engine and compressor operation. Controls technicians' responsibility may vary from simple keypunch operations to start an engine to making major programming changes that affect station-wide systems.

Corrosion technicians perform most of their work outside the station. They are responsible for ensuring that corrosion of the pipeline and other equipment remains within government standards. Corrosion of the pipeline is caused by several factors including soil conditions, the composition of the gas (e.g., high sulfur content), and its proximity to other pipelines or equipment. The pipeline is divided into cathodic (positive) areas and anodic (negative) areas. In the anodic areas current flows from the pipeline steel into the surrounding electrolyte (e.g., soil, water) taking particles of metal with it, and causing the exterior of the pipeline to corrode. In order to protect those areas it is necessary to make the entire pipeline collect current (or become cathodic). This is accomplished by sending electric current from an electrical source (a rectifier) through the soil and onto the pipeline. The electrical current overpowers the current discharging from the pipeline and makes the entire surface cathodic, therefore providing protection to the entire pipeline. In order to set up this system, the technician must determine the location of the problem, or more specifically where the greatest amount of damage is likely to occur, the best location for the rectifier and its accompanying anodes, and the correct size of the rectifier bed.

The fourth discipline in which technicians work is the Measurement area. These technicians are responsible for the equipment that measures the flow of gas being transported and sold. The gas is measured by means of a device called a meter. Meters record the static pressure – the pressure of the gas in the pipeline, and the differential pressure – the pressure difference on both sides of the meter. There are currently 2 types of meters in use at natural gas stations. The first is the dry flow or mechanical meter which records the measurements using colored pens and a paper chart. The second type of meter is an Electronic Flow Meters (EFMs). These electronic meters record the flow of gas through transducers and saves the information until it is downloaded using a laptop computer or printer. The technician is responsible for the maintenance of both types of meters, therefore he must be able to work on the mechanical as well as electronic components.

Although technicians usually have one major area of responsibility, in most cases they are responsible for tasks that fall under the heading of another discipline. This crossover into another discipline can range from performing a single task in the secondary discipline to spending up to 50% of their time performing the tasks of that discipline. Due to this crossover of duties, it is conceivable that a technician could be performing tasks applicable to all 4 disciplines.

JOB ANALYSIS

Identification of Critical Tasks

In order to develop a test for the technician position, a job analysis was conducted to compile a comprehensive tasklist for the position. Job analysis provides the foundation for the development of the KSAs, and is the first step in establishing the validity of an evaluation/testing procedure. Not only is a job analysis necessary for developing evaluation procedures and standards, it is almost mandated by the Federal Uniform Guidelines (EEOC, 1978). The job analysis provides an empirical link between the testing procedures and the job requirements.

The job analysis was conducted in several phases. First, a preliminary tasklist was developed using past research on similar jobs and company operating and procedure manuals. HPS staff then visited several stations to observe technicians performing job tasks and to have the preliminary tasklist reviewed and expanded. A Subject Matter Expert (SME) meeting was then held in which technicians from each discipline reviewed and revised the task statements. Duties not included were added to the tasklist. After the SME revisions were completed, the finalized tasklist was sent to a random stratified sample of seventy technicians, across the 4 disciplines.

The technicians were asked to rate the tasks using two rating scales, frequency and importance. The frequency scale identified the number of times a task was performed in a set time period (e.g., a week, a month). The importance scale measured how important a task was to successful job performance as it related to safety, continuity of service, equipment damage, efficiency, and housekeeping. Technicians were instructed to give tasks they had not performed in the last 3 years or never performed a rating of zero on the Frequency Scale. This ensured that tasks were only rated by individuals familiar with the task.

Task Criticality Determination

Task criticality was defined as a function of the ratings on the Frequency and Importance scales. A linear combination of the rating means per task on these two scales was used to calculate criticality. The mean Frequency rating was summed with the mean Importance rating to yield a criticality index. The rating means were standardized (i.e., z–score) prior to the combination .

This equation combines importance with frequency and allows a rank ordering of the important, frequently performed tasks. If the composite score for the algorithm is below 0.0, the task was given either a low rating on importance or frequency or both. Since the goal was to select important and frequent tasks, a cutoff of 0.0 was selected because this score was the mean value for the standardized criticality values and indicated that the task was performed frequently, and was considered important. The use of the mean value indicated that tasks with a critical value of 0.0 or greater were critical and those below 0.0 were considered non–critical.

To ensure that criticality decisions were based on the data from individuals who performed the tasks most often, a separate criticality analysis was performed for each discipline using only those technicians, for whom that discipline was their major area of responsibility. Separate critical tasklists were developed for each discipline based on the results of the criticality analysis. These separate lists were then combined to form the master technician tasklist. In looking across the four critical lists, some tasks were shown as critical for multiple disciplines.

Knowledge, Skills and Abilities (KSA) Identification

In order to develop test items the knowledge, skills, and abilities (KSAs) needed to perform the critical tasks must be determined. A preliminary list of KSAs based on previous experience and the critical task list was developed. This KSA list was reviewed presented to a group of foremen, engineers and technicians at an SME meeting. They were first asked to review the KSA lists and make any revisions necessary. The SMEs were then asked to link the revised KSAs to the critical tasks. A KSA was considered linked to a task if it was linked by more than half of the SMEs. Both the critical tasks and the linked KSAs were used in the test development phase.

TEST DEVELOPMENT

Test Format

An oral and performance test format was selected for the technicians test. In this format, questions are presented verbally and answered either orally or by hands–on performance of a job task by the examinee. Although technicians tend to be more familiar with testing than individuals in other natural gas positions, it was determined that this was still the best format given the variety of equipment and procedures used across the NGPL locations and the potential for fairer treatment of the examinee.

Although the technician position is viewed as one job by the company, for the purpose of test development it was viewed as 4 separate jobs. Using this methodology test questions were developed for all critical tasks and KSAs pertinent to each discipline separately.

Test Clusters

The diversity of knowledges, equipment and duties found within each discipline of the technician position, poses a problem in the development of test questions. The job can be diverse at one location and restricted at another location due to size and equipment availability. It is important to evaluate an individual's knowledge of the equipment at his/her location, without penalizing individuals whose station is not similarly equipped. To accommodate geographical and location difference, a clustering system was developed.

Within each section of the test, the questions were clustered by major areas of knowledge identified in the job analysis. The use of this clustering system allowed for the identification of strong and weak knowledge areas for the examinee, and the identification of training needs for the examiner. In addition, this clustering system allowed an individual to be tested on only those areas of knowledge applicable to the job at his or her specific location (Figure 1).

Test Construction

Two or more questions were developed for each critical task and KSA within a subsection. The questions were developed based on the technician critical task list, on site interviews and observations, SME meetings, and written materials. Initial questions and answers were written by HPS staff. Where the same KSA and/or critical task was matched to more than one discipline the same or similar questions appeared in each preliminary test to assess whether these questions were accurate across all relevant disciplines.

Test Review

After the preliminary tests were completed, a separate SME meeting was held for individuals within each discipline to evaluate the draft test. SMEs were asked to review each question for: (1) adequate coverage of the KSA/task; (2) readability of each test question; (3) accuracy of the answer; and, (4) applicability of the questions to all locations.

Review of the draft questions. A five step procedure was used to review the test questions. First, the question was compared to the KSA and tasks associated with the question to determine whether the intent and form of the question reflected the actual job knowledge and responsibilities. Second, the wording and the accuracy of the question and answers were reviewed and revised, if necessary. Third, each question and answer was reviewed to ensure it was applicable equally across all locations. Fourth, the SMEs examined each question to determine how many responses the examinee should be required to provide. Lastly, the panel reviewed each KSA to determine whether additional questions needed to be developed. In addition, the steps required to complete a performance question were reviewed to ensure that they were in the proper sequence. Steps that posed a safety hazard if omitted or performed out of sequence were identified. Due to the importance of these safety steps, it was decided that the examiner could terminate a performance question if the examinee failed or incorrectly performed a safety–related procedure.

TECHNICIAN CLUSTERS

COMMUNICATIONS

- Microwave Systems
- Multiplex/Phone Equipment
- VHF Base Station and Mobile Radio
- Equipment Maintenance

CONTROLS

- Engine Operation
- Engine Preventive Maintenance
- Compressor Operation and Repair
- Engine Analysis

CORROSION

- Corrosion
- Cathodic Protection
- Surveys/Testing

MEASUREMENT

- Meter Maintenance and Installation
- Electronic Flow Meters
- Gas Analysis

Figure 1

For the questions presented to SMEs in more than 1 discipline, the original question and the revised version were presented in each subsequent meeting to ensure that the information was perceived the same across all disciplines. If the technicians in all relevant disciplines agreed on the format and answers for the question, the item remained part of the test. If however, the revisions of the item resulted in vastly different questions being developed for each discipline the critical task/ksa being tapped was re-evaluated, and the meaning of that critical task/KSA to each discipline was determined.

In some cases a specific statement represented different actions within each discipline. For example, the task statement "Program a computer system" could mean program a personal computer to a Communications or Corrosion Technician; program an Electronic Flow Meter to a Measurement Technician; and, use program logic to change the oil temperature of the engines to a Controls Technician. If it was determined that a different action was being measured in each discipline, individual questions were written for that sub-section of the test.

Overlap strategies. To handle situations where the same question(s) were applicable across multiple disciplines, three strategies were developed for incorporating these questions into the test. If the overlap was represented by a single question, the question was simply placed in each sub-section of the test with an examiner's note instructing the examiner to transfer the individual's score on that question to the same question in the other relevant clusters. If however, the overlap consisted of several questions that could not be grouped into an existing cluster of that discipline's test, a separate overlap section was developed within the sub-test (e.g., a Measurement overlap section was created in the Communications test). Thirdly, in those cases where the individual was responsible for a specific portion of a second discipline (e.g., a communication technician who is responsible for maintaining EFMs), the corresponding cluster of the second discipline's test was administered in its entirety.

Pretest of Examination

Each sub-section of the test was pretested on technicians from the various disciplines who possessed varying amounts of experience and knowledge. Each clusters of the test was administer to two examinees, to ensure the questions were understandable, and that they discriminated between knowledge levels. At the end of the pretesting session, the examiners and examinees evaluated the test and the testing procedures to identify any changes, additions, or deletions that would increase the clarity of the questions, the accuracy of the answers, and the effectiveness of the testing procedure. This discussion resulted in the further refinement of the technician examination (e.g., questions reworded, correct responses added) and testing procedures.

TEST VALIDATION

Validation of the Technician test was accomplished through content and criterion related strategies. The validation process involved providing evidence that the test was predictive of job performance. Since the test questions were derived from the KSAs and contained job content knowledge, content–validity was established. However, to determine the score that indicated successful performance, a criterion–related validation study was conducted. The test was administered to a sample of incumbent technicians and the relationship between their scores on the test and an evaluation of their job performance was determined. The design and conduct of the validation study is described below.

Criterion Measure Development

To determine the relationship between test performance and job performance, supervisor ratings of job performance were selected as the method for evaluating the incumbents. Two rating scales were developed to evaluate technicians' job performance across several job functions. A knowledge scale evaluated the technician s knowledge of each cluster of knowledge (e.g., Cathodic Protection, Gas Analysis). A performance scale evaluated the technician's ability to perform work in each knowledge cluster (e.g., Engine Repair, Multiplex/Phone Equipment).

Supervisors rated the technicians knowledge and performance on three levels. First the supervisor rated a technician's overall knowledge and performance within his discipline (i.e., Controls, Communication, Corrosion, Measurement). Secondly, the supervisor rated the technician's knowledge and performance on each cluster (e.g., Multiplex/Phone Equipment, Engine Operation, Cathodic Protection, Gas Analysis) within the discipline. Finally, the technician was rated on subsections of information within the cluster. For example, the Engine Analysis cluster was divided into: 1) Use of Oscilloscope, and 2) Use of Ultrasonic Flaw detector. These three levels of ratings provide an overall picture of the technician's on–the–job performance. These ratings were used as the criterion measure to assess the validity of the test.

Validation Sample

Technicians within each discipline accounting for 50–75% of the population, (Communications = 12; Controls = 20; Corrosion = 7; Measurement = 21) participated in the validation study. These technicians were tested on the sub–test for their specific discipline (e.g., Controls). Their job performance was rated by their supervisors. The number of individuals included in the analysis of each cluster varied as a result of the job tasks performed or equipment present at a location. For example, the engine analysis cluster in the Controls section of the test was not taken by all Controls technician examinees because many of these technicians do not perform engine analysis at their location. Therefore the sample size within the cluster is smaller than the sample size for the other Controls clusters.

Validity of the Technician Test

An analysis of the supervisor ratings determined that a combination of the ratings within the clusters for each discipline produced the best assessment of the employee's performance on the job. A criterion measure was calculated for each cluster, using the third level knowledge and performance ratings for that cluster.

The test cluster scores and the supervisor ratings of job performance were used to provide evidence of the criterion–related validity of the test through a multiple regression analysis (multiple R). A correlational analysis was performed between the test cluster scores and the respective supervisor ratings (the sum of Knowledge and Performance) based on the adjusted sample. Although most of the clusters were not significantly correlated with their corresponding criterion measure, most were moderately high, yielding correlations ranging from $r = .28$ to $.76$. Previous research on larger sample sizes have yielded comparable correlational values (Gebhardt & Sheppard, 1992). In addition, a correlation of $r = .25$ to $.35$ is considered substantial in cognitive testing. We believe this lack of significance is due in large part to our small sample sizes.

A stepwise regression was performed, for each discipline using all relevant subjects (i.e., technicians who perform those job duties), although our sample sizes were small. When all Communications clusters were included, the validity coefficient was 72. The Controls clusters yielded a validity coefficient of .71. The validity coefficients for the Corrosion and Measurement clusters were .48 and .65 respectively.

Although all clusters within each discipline did not add significantly to the prediction of job performance, it was important that all clusters be retained because of the need to demonstrate that the worker is qualified to perform the essential job functions in accordance with DOT regulations. Further, the job analysis showed that each cluster represented an important area of knowledge and skill that is necessary to perform the Technician job successfully, therefore meeting the requirements for content validity. If any cluster was omitted, an important aspect of the job would be missing from the test.

DETERMINATION OF CLUSTER GOAL SCORES

A multiple hurdle strategy was adopted for the Technician test in which a minimum score must be achieved for each cluster. When this strategy is used, an individual's performance (e.g., meeting goal score) on the test can be used to identify clusters in which knowledge is adequate and clusters in which training is needed.

Two procedures were used to determine the goal score for each cluster. First, an expectancy table was developed for each of the clusters within a discipline area. Expectancy tables provide an indication of the probability of different criterion (e.g., job performance) outcomes for persons obtaining specific tests scores. This basically allows one to determine an individual's expected job performance when various minimum test scores are required.

Second, the pass–fail ratios for different goal scores were investigated. This strategy involved identifying a minimum goal score that will maximize the number of valid passes (i.e., acceptable job performance and acceptable test performance) and the number of valid failures (i.e., unsatisfactory job performance and unsatisfactory test performance).

Recommendations for Minimum Goal Score

The identification of a minimum goal score that maximizes the expected utility of the test across all outcomes can be accomplished through the use of pass/fail tables. To determine the expected utility the number of incumbents who would pass and fail the test at different goal scores is determined. The utility of these cutoff scores is assessed in relation to the expectancy tables and the pass/fail rates.

The utility of diffe.:ent goal scores was evaluated by ranking the incumbents on the basis of their performance on the criterion measure across the distribution of the criterion. This distribution was divided into 10 intervals, each containing approximately 10% of the sample. For each cluster a job performance cutoff was set at the 10% interval nearest the job performance score indicative of satisfactory performance on that cluster. A knowledge or performance score of 3 on each individual rating item indicated satisfactory performance. To determine satisfactory performance for a cluster, the total number of ratings within a cluster was summed and multiplied by a score of 3. For example, the satisfactory job performance for the Equipment Maintenance cluster was set at 30 points: [5 knowledges + 5 performance] x 3 points = 30 points.

A sample pass/fail table is presented in Table 3. With a score of 67.25 on the Equipment Maintenance cluster, 83.3% of the technicians who obtained the goal score of 67.25 on the Equipment Maintenance cluster were considered successful on the job (≥ 30 on job performance). For the Technicians who did not obtain score of 67.25, 83.3% were rated unsuccessful (< 30 on job performance) on the job. Therefore, the valid pass and rejection rates were each 83.3%.

One must also investigate the gain in job performance with increases in a test score. This is accomplished by reviewing the expectancy table for that cluster which shows the percent of incumbents who score at a specific goal score. An expectancy table was generated for each cluster by ranking the incumbents on their cluster scores relative to the corresponding distribution of cluster scores. For each cluster, the distribution of cluster scores was divided into ten intervals with each interval containing approximately the same number of incumbents. For each distribution of test scores, the mean criterion rating was calculated at 10% intervals.

An example expectancy table is presented in Table 4. For the Equipment Maintenance cluster, if the minimum goal score was set at 67.25 (50% level), 50% of the sample would score 67.25 or higher. The job performance score for individuals who scored 67.25 or higher was 39.00 which is above the average level (i.e., 30.00). Using the data from the valid pass/fail rates and the expectancy tables, minimum goal scores were established for each cluster.

Table 1

Example of a Pass/Fail Table and an Expectancy Table
for the Equipment Maintenance Cluster
[0 to 50% only]

PASS/FAIL TABLE

Test Battery Cutoff Score	% Correct Decisions	% Incorrect Acceptances	% Incorrect Rejections
48.63	58.2	45.5	0
59.50	50.0	50.0	50.0
61.30	58.3	44.4	33.3
65.59	66.7	31.5	25.0
67.25	83.3	16.7	16.7

EXPECTANCY TABLE

Minimum Test Cluster Score	% of Techs Scoring over Test Minimum	Mean Criterion Measure Score
0	100.0	35.58
48.63	91.7	35.91
59.50	83.3	36.10
61.30	75.0	36.40
65.59	66.7	37.00
67.25	50.0	39.00