DOCUMENT RESUME

ED 113 533 CE 005 065

AUTHOR	McKnight, A. James; And Others
TITLE	The Development of Job-Oriented Examinations for Postal Equipment Maintenance Positions: Subtask Report.
INSTITUTION	Human Resources Research Organization, Alexandria, Va.
SPONS AGENCY	Post Office Dept., Washington, D.C.
PUB DATE	Jun 69
NOTE	64p.; For related document, see CE 005 064
EDRS PRICE	MF-$0.76 HC-$3.32 Plus Postage
DESCRIPTORS	*Employment Qualifications; Equipment Maintenance; *Government Employees; Item Analysis; *Machine Repairmen; Manpower Development; Multiple Choice Tests; Occupational Tests; *Performance Tests; *Test Construction
IDENTIFIERS	Mail Processing Equipment Maintenance Personnel; *Post Office

ABSTRACT

The report discusses the development of written job proficiency examinations for four Mail Processing Equipment (MPE) positions (MPE apprentice, MPE mechanic, MPE senior mechanic, and MPE supervisor). After a brief introductory chapter, the next chapter describes the determination of examination objectives and the desirability of testing specific job knowledge and the aptitude for acquiring future job qualification knowledges, but not job skill, personality factors, and physical characteristics. Chapter 3 discusses the preparation of the examination items (of the five-alternative multiple choice variety) based on an analysis of maintenance and supervisory skills, the purpose of which was to test minimum knowledge standards while at the same time assuring an adequate supply of successful applicants. Chapter 4 discusses the preliminary administration of the 384 items to a sample of maintenance personnel currently engaged in job activities that correspond to the proposed job descriptions at the 13 most highly mechanized post offices. Chapter 5 covers the selection of 339 test items, the assignment of the items to the four qualifying examinations, and the minimum standards for determining job qualification. Chapter 6 discusses test validity and makes suggestions to the Post Office Department for broadening the assessment of worker qualifications. (JR)

SUBTASK REPORT

The Development of Job-Oriented Examinations for Postal Equipment Maintenance Positions

by

A. James McKnight, Richard D. Behringer,
J. Robert Lodge, and Miriam Safren

~~NOT FOR USE~~

June 1969

~~OUTSIDE OF HumRRO~~

United States Post Office Department
Contract No. RE 73-67
PPBS No. 70-80

HumRRO Division No. 1 (System Operations)
The George Washington University
Human Resources Research Office
300 No. Washington St.
Alexandria, Virginia

ABSTRACT

In this report the development of job-oriented qualification examinations for postal mail processing equipment (MPE) personnel is described. The examinations were prepared for the MPE Apprentice, Mechanic, Senior Mechanic, and Supervisor positions recommended in a companion HumRRO report dealing with MPE personnel classification. Content of the examinations was derived from a comprehensive and detailed analysis of MPE maintenance tasks. A pool of 384 test items was administered to current personnel at the 13 most highly mechanized U.S. post offices. Analysis of the results showed that the items constituting the preliminary pool possessed an above-chance relationship to worker proficiency as the latter was indicated by supervisor proficiency ratings and worker position in the job hierarchy. From the original pool, the 339 most valid items were selected on the basis of standards developed jointly by the Bureau of Personnel, U.S. Post Office Department, and HumRRO. Two alternate test forms were prepared for each of the designated job positions. Recommendations for future improvements in the assessment of job qualifications are provided in this report.

FOREWORD

The Mail Processing Equipment (MPE) test development subtask described in this report was conducted by the Human Resources Research Office as part of HumRRO Task MPE, the overall objective of which was to develop improved classification and selection procedures for Mail Processing Equipment (MPE) maintenance positions. This report deals with the development of examinations for four MPE positions recommended by the classification subtask report submitted to the U.S. Post Office Department in April 1969. A third Task MPE report concerns the analysis of postal maintenance jobs and presents data developed as a result of an analysis of MPE maintenance job positions.

The MPE test development effort was initiated in the Fall of 1968 on completion of a detailed analysis of MPE maintenance positions. Preliminary administration of test items was conducted in January 1969 and final examinations were prepared during the Spring of 1969. The research was performed by HumRRO Division No. 1 (System Operations) under the overall direction of Dr. J. Daniel Lyons, Director of Research. Dr. A. James McKnight was the Study Leader and members of the research staff included Dr. Richard D. Behringer who conducted the item analyses, Mr. J. Robert Lodge who prepared scheduled maintenance items, Dr. Miriam Safren who prepared supervisory items, and Mrs. Lola Craw who performed tabulations and various statistical analyses. Mr. William A. Carswell of Carswell, Vandiver and Associates, prepared items concerned with unscheduled maintenance.

In addition, the following representatives of the Post Office Department provided assistance both as advisors in planning the study and as participants in the administration of preliminary tests: Miss Ruth O. Peters, Employment and Placement Division, Bureau of Personnel; Mr. Marlin Burkhart, Compensation Division, Bureau of Personnel; Mr. David McCutcheon, Maintenance Division, Bureau of Facilities; and Mr. Erwin Vollmer, Bureau of Operations. Mr. Vincent J. Chirichella and members of his maintenance force at the Washington, D.C., Post Office served as technical advisors in the preparation of test item content. Progress of the study was greatly facilitated by cooperation of Postmasters, key maintenance staff personnel, and MPE Mechanics at the 13 mechanized post offices where preliminary tests were administered.

HumRRO research was conducted under Contract No. RE 73-67, PPBS No. 70-80, U.S. Post Office Department.

Meredith P. Crawford
Director
Human Resources Research Office

SUMMARY AND CONCLUSIONS

Problem

The postal service, one of the nation's largest single employers, relies heavily on its program of service-wide written examinations in assessing the qualifications of personnel for employment or promotion. With a program of such magnitude, it is difficult to assure that the qualifications by which workers are judged actually reflect the needs of individual jobs. It was for this reason that HumRRO undertook, at the request of the U.S. Post Office Department, a study aimed at improving the job relevance of examinations used to assess the qualifications of mail processing equipment (MPE) maintenance personnel. The new examinations were to be prepared for the MPE Apprentice, MPE Mechanic, MPE Senior Mechanic, and MPE Supervisor positions recommended in a companion HumRRO study concerned with classification of postal equipment maintenance positions.

Method

The following major steps were taken to help assure the development of qualification examinations capable of assessing an individual's ability to perform the particular job for which he was a candidate:

(1) Literature relating to measurement of job qualifications was surveyed to identify those qualifications most amenable to measurement through formal examination and to assess the types of examination that have proven most valid.

(2) A detailed analysis was performed of job activities and their associated knowledges and skills as a means of securing appropriate examination item content.

(3) A preliminary pool of examination items was assembled and administered to a sample of maintenance personnel currently engaged in job activities that corresponded to the proposed job positions. An attempt was made to obtain judgments of job relevance and to determine the usefulness of the items in distinguishing among individuals at different levels of rated ability and different positions within the job hierarchy.

Results

Analysis of the preliminary item pool as a whole showed an above-chance relationship between item responses and the two criteria of rated and classified (job hierarchy) ability. Individual items were selected

for final qualifying examinations according to the following HumRRO/POD standards of acceptability:

(1) The correct answer to each item must have been selected by individuals having the highest mean supervisor rating (as compared to those selecting other alternative answers).

(2) More than 50% of jobholders must consider the item relevant to their job.

(3) The item must be answered correctly by more than 20% and less than 95% of individuals in the job for which the examination is intended.

(4) Content of the examination item must be appropriate to the job in accordance with the results of the job analysis.

A total of 339 items met criteria for inclusion in the final examination. These items were assigned to the four qualifying examinations with the proviso that no item appear on more than two tests. A pair of alternate forms was prepared for each examination. One subset of items was designated "minimum standard" to be used in determining whether a candidate for a position should be considered as "qualified." An item had to be passed by 80% of workers in the job for which it was intended and judged by more than 75% of them as representing basic information that every qualified jobholder should possess, in order to be included as a minimum standard item.

Conclusions

Principal conclusions reached in this study are:

(1) The qualification examinations, because of the manner in which they were developed under this study, represent valid indices of ability to perform the specific jobs. However, no statement can be offered as to the level of validity or associated probabilities, and it has been recommended that all examinations be applied to an independent sample of MPE maintenance personnel for an assessment of overall validity. Such a program of validation was proposed as a part of the study but was omitted at the request of the U.S. Post Office Department when the overall HumRRO study effort was reduced in scope.

(2) Efforts should be undertaken by the U.S. Post Office Department to broaden the assessment of worker qualifications in the following manner:

(a) Use of performance tests to assess skills not measurable by written examinations.

(b) The application of job analytic procedures to the determination of required worker background

characteristics (e.g., education, job experience) as a means of improving "assembled tests."

(c) Use of behavior ratings to assess critical, job-related, personality characteristics.

(d) Use of a differential aptitude measure to assess the ability of postal employees to acquire various types of job skills.

(3) The implementation of qualifying examinations must be closely coordinated with other personnel activities such as recruitment, training, job classification, and work operations, if the examinations are to be of substantial benefit in filling individual job positions with qualified personnel.

TABLE OF CONTENTS

THE DEVELOPMENT OF JOB-ORIENTED EXAMINATIONS FOR POSTAL EQUIPMENT MAINTENANCE POSITIONS

Chapter 1

INTRODUCTION

As an organization grows in size, it becomes increasingly difficult to assess the qualifications of individual jobholders within it. Yet, some program of individual worker assessment is necessary if jobs are to be filled by the most capable people. Most large organizations devote considerable time and energy to the development of formal, systematic worker assessment programs. Questions inevitably arise as to the ability of such large-scale programs to cope with the peculiarities of each individual job. It was such concern over the job-relatedness of examinations that led the U.S. Post Office Department to request HumRRO, in 1967, to undertake a program of research aimed at improving qualification examinations administered to postal equipment maintenance employees. This report describes the nature and the results of that research program.

The effort was part of HumRRO Task MPE, a broad personnel research program in the area of Mail Processing Equipment (MPE) maintenance, which included, in addition to revision of examinations, the preparation of detailed job descriptions, and the reclassification of jobs with improved job career ladders. These additional MPE study efforts are described in two reports:[1]

Fink, C. D., and Hibbits, F. L. *Classification, Career Structure, and Job Analysis of Mail Processing Equipment Maintenance Personnel*, Subtask report, Human Resources Research Office, April 1969.

McKnight, A. J., Fink, C. D., *et al. Analysis of Postal Equipment Maintenance Positions*, draft report prepared for the U.S. Post Office Department, Human Resources Research Office, June 1969.

---

[1] Two additional HumRRO efforts in the training area are described in the following reports:

Lyons, J. D., and Williams, L. W. *Development and Initial Presentation of an Advanced Maintenance Management Course for the Post Office Department*, draft Technical Report prepared for the U.S. Post Office Department, Human Resources Research Office, submitted March 26, 1969.

Trexler, R. C., and Butler, P. J. *Task SABER: The Development of a Technical Training System*, draft Technical Report prepared for the U.S. Post Office Department, Human Resources Research Office, submitted September 1968.

The job positions for which improved qualification examinations were to be prepared are those concerned with the maintenance of mail processing and related equipment. One reason for giving initial priority to Mail Processing Equipment (MPE) is the fact that this equipment is almost entirely unique to the Post Office Department. Letter Sorting Machines, Facer Canceller Machines, or Parcel Sorters are not often found outside of postal installations. For this reason, the knowledges and skills associated with maintenance of this equipment are not as well understood as would be the case for other types of postal equipment, such as elevators, air conditioners, or trucks. As a matter of fact, owing to the recency with which some of this equipment has been introduced, the maintenance requirements are not universally understood even within the postal service. The second reason for giving primary attention to MPE is the critical role this equipment plays in the current program of mechanization by which the Post Office Department (POD) seeks a step improvement in the efficiency of its mail distribution. Any program aimed at facilitating the maintenance of such critical equipment may be expected to pay large dividends in the overall effectiveness of the postal operation.

At the time the research program was instituted, the maintenance of MPE was performed by personnel in the following positions, covered by the examinations indicated:

Position	Examination
Helper PFS-4	No exam.
General Mechanic PFS-5	Mechanical Aptitude
MPE Mechanic PFS-6	3-part Electromechanical
MPE Mechanic PFS-7	" " "
MPE Foreman PFS-9	MPE Supervisor

Under the classification project, the scope of MPE maintenance was broadened somewhat to include a variety of electromechanical equipment items not directly connected with the processing of mail. A survey conducted early in the research program (see *Analysis of Postal Equipment Maintenance Positions* referenced above) had shown that MPE personnel at many post offices maintained such equipment as tying machines, electric trucks and lifts, cleaning equipment, and communication equipment, owing largely to the similarity of skills involved. This is particularly true at smaller post offices where limited manpower requires that mechanics be able to work on a wide range of equipment.

The specific positions for which tests were required are described in the companion Classification report referenced above. The following position descriptions are excerpted from that report.

Mechanic Helper (PFS-4). Performs, independently, a variety of simple nontechnical and semiskilled tasks that are incidental to recognized trades or crafts, or similar maintenance repair functions. Assists craftsmen and mechanics in performance of maintenance tasks that require skill and knowledge of the function.

Apprentice Mechanic, Mail Processing Equipment (PFS-5). As an Apprentice, learns and performs routine preventive maintenance, troubleshooting and repair of one type of mail processing equipment. Assists mail processing equipment mechanics in complex troubleshooting, diagnosis and correction of equipment malfunctions. Performs cleaning, lubricating, inspection, and simple maintenance tasks under supervision, following detailed instructions or established procedures. Learns and complies with safety regulations. Learns use of tools and simple measuring and test equipment.

Apprentice Mechanic, Mail Processing Equipment (PFS-6). As an advanced apprentice, learns and performs routine preventive maintenance, troubleshooting, inspection, and repair of a second type of mail processing equipment. Assists mail processing equipment mechanics in complex maintenance actions and performs standard maintenance tasks such as preventive maintenance and troubleshooting, diagnosis and correction of equipment malfunctions of mail processing equipment on which trained. Prepares simple written maintenance activity reports and records.

Mechanic, Mail Processing Equipment (PFS-7). Performs standard maintenance, inspection, troubleshooting and repair of two types of mail processing equipment, working independently. Repairs, removes, installs, modifies, assembles, and disassembles, any mail processing equipment when working under supervision as a member of a work team. Reads, comprehends, and utilizes manuals, schematics, diagrams, and drawings to diagnose and correct equipment deficiencies. Learns routine preventive maintenance, troubleshooting and repair of a third type of mail processing equipment. Prepares or completes written equipment status reports, maintenance logs, and work order documentation.

Senior Mechanic, Mail Processing Equipment (PFS-8). Performs, independently, any inspection, malfunction diagnosis, and repair, on at least three types of mail processing equipment to include their electrical control subsystems. Acts as a work crew chief in the repair, removal, installation, modification, assembly and disassembly of any mail processing equipment. Utilizes and understands any type of POD maintenance reference materials. Provides technical assistance and training guidance including safety precautions, use of hand and power tools and measuring and test equipment, to lower-level mechanics. Prepares complex, written, equipment work orders, status reports, and maintenance procedures.

Foreman, Mail Processing Equipment Maintenance (PFS-10). Plans activities of the mail processing equipment maintenance work force, assigns duties and tasks to mechanics and crews and directs training of personnel. Inspects work and equipment to verify quality of maintenance and completion of tasks. Analyzes equipment status and implements plans for correction of deficiencies and replacement actions. Prepares reports, evaluates and counsels personnel, recommends promotions and personnel actions and conducts safety training and enforces safety regulations.

The tests to be developed would be administered to candidates for each of the posts indicated with the exception of the 6 level Apprentice whose advance from the 5 level position would be based entirely upon progress within the apprentice program.

In addition to the positions noted, larger post offices would be allotted the following MPE-related positions:

(1) MPE Maintenance Technician (PFS-9). A specialist on a particular equipment item (e.g., LSM) promoted from Senior MPE Mechanic without examination.

(2) MPE Training Officer (PFS-11). A supervisor of formal MPE training programs, promoted from MPE Foreman or Maintenance Technician without examination.

(3) General Foreman of MPE Mechanics (PFS-12). Supervisor of all MPE Foremen, promoted from MPE Training Officer, MPE Foreman without examination (in the case of the MPE Training Officer, the Foreman examination must have been passed).

A group of Electronic Technician positions was created, as an offshoot of the MPE career ladder, primarily to handle the new Optical Character Reader, a device to sort mail automatically installed at a number of larger post offices. The Electronic Technician would also assume maintenance responsibility for certain computer equipment and electronic memory systems. A test for entrance to this field already having been developed, the Electronic Technician positions were omitted from the present test development effort.

This report describes the process by which examinations were developed for the Apprentice, MPE Mechanic, Senior Mechanic, and Supervisor positions. Chapter 2, "Determination of Examination Objectives," explores, at length, the characteristics of workers that enable them to perform their jobs, and identifies those considered suitable for assessment through formal examination. Chapter 3, "Preparation of Examination Items," describes the systematic techniques used to develop examination content and the preparation of a preliminary set of examination items. Chapter 4, "Preliminary Administration," describes the administration of the preliminary test items to a representative sample of postal equipment maintenance personnel and the collection of criterion information. Chapter 5, "Development of Final Qualification Examinations," describes the results of the preliminary administration and the selection of a series of test items having demonstrable job relevance. Chapter 6, "Suggestions for Further Development," outlines techniques, beyond the scope of the present study, that might be used to improve the validity of individual worker assessment within the postal service.

Chapter 2

DETERMINATION OF EXAMINATION OBJECTIVES

The first step in the process of developing a set of examinations was to determine what types of measures are appropriate to the postal service. This process involved (a) a study of the characteristics that define an individual's qualifications for a job, and (b) selection from these characteristics those that are suitable for measurement. This effort was extremely broad and included consideration of qualifications and measures that could not feasibly be covered within the scope of the present study. The purposes of such a comprehensive investigation were both to help make clear the limitations of whatever measures were developed from the present study, and to assist the postal service in charting the course of further progress in development of worker assessment techniques. While much of the contents will be familiar to those well versed in the measurement of job proficiency, this chapter provides a foundation for the recommendations offered in Chapter 6.

The term "job qualifications" in the context of this report refers to those characteristics of an individual that enable him to perform his job successfully. These include (a) job abilities -- the specific knowledges and skills that enable the individual to carry out job activities; (b) aptitudes -- those general abilities that enable an individual to acquire specific job knowledges and skills; (c) motivational characteristics -- those interests, drives, and values that determine how well the individual will apply his abilities to the demands of the job; (d) personality characteristics -- those normal or characteristic ways of behaving that determine how likely the individual is to carry out job activities; and (e) physical characteristics -- those aspects of physical structure and function that are related to the individual's ability to perform the job. Each of these human characteristics was examined to determine its relation to the maintenance of electromechanical equipment. At the same time, research literature was reviewed in detail to determine what relationships between each of these characteristics and job performance had been demonstrated in the past.

JOB KNOWLEDGE

Job knowledge, that is, the possession of information concerning the job, is clearly one major determinant of how well an individual will perform his work. An experienced mechanic has typically accumulated a considerable store of job information that he can call upon in solving maintenance problems as they arise.

Types of Job Knowledge

Most of an individual's work activities are guided by procedures, that is, information which specifies the particular series of steps required to perform a work activity. Maintenance technicians generally acquire the ability to service equipment, replace parts, make adjustments, and perform other relatively routine activities by learning a set of procedures. Sometimes the procedure is one that has been specified in writing, while in other cases it is learned by watching someone else perform it. Some procedures follow a pre-determined sequence while in others the sequence will vary depending upon what happens as the steps are performed. An example of the latter is diagnosis of an equipment breakdown wherein the results obtained at one step of the process will determine what step to take next.

In addition to learning what to do, the technician must generally acquire a considerable body of factual information concerning the nature of the equipment with which he works. He must know, for example, where things are located, what they are called, normal operating conditions, critical values and tolerances, failure symptoms, safety hazards, and the idiosyncrasies of particular pieces of equipment. At a higher level, supervisors of maintenance must know maintenance policy, servicing schedules, and the strengths and weaknesses of their individual subordinates.

Where a technician cannot be provided specific procedures or facts concerning his equipment, he may have to draw upon a theoretical understanding of how equipment works in order to figure out for himself what is required. For example, a mechanic attempting to track down the source of an equipment failure may be called upon to apply concepts of electronics, mechanics, or hydraulics. In general the importance of theory to the maintenance of equipment seems to have been somewhat overrated. Studies of maintenance have repeatedly shown that most maintenance problems are amenable to pre-established procedures or, at most, relatively simple theoretical concepts. Judging from the rate at which it tends to be forgotten over time on the job, complex theory is not only unnecessary, but not particularly useful.

Knowledge Tests

Tests of job knowledge date back to early trade tests developed by Chapman in 1921 (*6*). He found that questions could be phrased which evidenced a significant relation to job experience as well as to supervisor ratings of ability. Many years later the worker analysis section of the U.S. Employment Service developed a series of 15 test items for each of 126 jobs. The items were selected on the basis of their ability to discriminate among groups of supervisors, apprentices, and people in related occupations. However, according to Anastasi (*1*), no validity data was provided on the final tests.

Tests of job knowledge have also been used extensively by the military services. Bellows (*2*) describes tests developed by the Air Force covering 97 different jobs or occupational areas, and reports median correlations of .54 with job performance. Morsh (*21*) reports

similarly high correlations between the written mechanic proficiency test and the supervisor ratings of performance. The U.S. Army reports validity information on 15 of the evaluation tests administered to enlisted personnel to assess their specific job ability. Correlations of test score with co-worker rating range between .05 and .54 with a median r of .29.

Although most trade tests are developed in written form, Hausman, *et al*, (*14*) note that oral tests, while relatively costly to administer, offer a number of advantages including (a) lowered reliance on verbal ability, (b) ease of revision in the face of job changes, and (c) flexibility in content and manner of presentation, making it easier to adjust the test to varying local conditions. Pictorial items have proven useful where job information cannot be readily accommodated in a written form.

Correlations between tests of job knowledge and measures of job proficiency are quite modest, particularly when proficiency is reckoned by observed or tested performance. Brown, *et al*, (*5*) found that the correlation between a written test of radio maintenance and measured performance was too low to warrant use of the former achievement test. A number of investigators including Ryans and Fredericksen (*22*), Meister and Rabideau (*20*), and Skinner (*23*) have called attention to the discrepancy between verbal behavior and job performance and have cautioned against uncritical use of job knowledge tests.

Since test items are typically developed by subject matter experts rather than jobholders, they often tend to be more academic than practical. Maintenance examinations generally emphasize theory of operation, nomenclature, technical details, and other forms of "book learning" instead of such job specifics as servicing and repair procedures, use of tools, location and identification of parts, failure symptoms and their causes -- all those factors that enable the experienced mechanic to perform his work quickly and accurately. Foley (*12*) points out that while knowledge examinations have some value, "...when their scope is limited to questions regarding the desired performance...there seems to be no evidence that broad theory questions or questions regarding peripheral materials have any correlation with ability to perform." One need only observe the following two sample items from Chapman's test, referred to earlier, in order to see the difference in approach:

What happens to the breaker points if the condenser is bad?

Burn ____ Pit ____ Foul ____ Corrode ____

What two metals are camshaft bearings usually made of?

Bronze ____ Brass ____ Babbitt ____ White Metal ____

While one may legitimately expect an auto mechanic to identify symptoms of a bad condenser, the appropriate metal for camshaft bearings would seem to be of greater concern to the automotive designer than to the mechanic.

Despite the low esteem in which tests of job knowledge have frequently been held, an assessment of the individual's job-related knowledge appears potentially one of the most efficient measures of his ability to perform. However, to attain its full potential the content of a knowledge test must be derived from a thorough analysis of the job to which it is to be related.

JOB SKILL

Part of the difference between knowledge and performance is the difference between knowing what to do and being able to do it. This difference has often been referred to as "skill." One type of skill is perceptual in nature and appears to involve the ability to interpret sensory stimuli. An example from the area of maintenance would be the ability to determine how hot a motor may become before it is symptomatic of breakdown or how discolored an electrical contact should be before it is declared unserviceable. While the mechanisms underlying these perceptual skills are not understood, it is clear that they involve more than mere information. A great deal of practice in dealing with stimuli is generally necessary before the required discriminations may be made.

Another set of job skills appears to be motor in nature, having to do with the ability to make appropriate physical responses. One type of motor skill is that which involves extremely rapid responses as, for example, the quick and deft movements required in soldering. Another form of motor skill involves simultaneous execution of multiple responses as in aligning very delicate equipment. The one distinctive feature that appears to underlie both these motor activities is an essentially "automatic" response, a response that appears to occur without conscious thought[1]. While the mechanisms underlying these automatic or reflex responses are not understood, it is clear that the continued repetition of a stimulus-response pattern may be necessary before a smooth "skilled" performance is obtained.

A third type of skill is of a cognitive nature and is associated with the ability to carry out the mental processes that intervene between stimulus and response. A mechanic, for example, may have all of the technical information needed to locate the source of a breakdown yet be unable to relate it to the problem at hand. Just how an individual learns to see relationships, make inferences and deductions, and carry out the covert activity involved in decision making or problem solving is not well understood. Yet, like perceptual and motor skills, these cognitive skills improve with practice.

---

[1] The fact that the responses are tied to external or internal stimuli has led to use of the term "perceptual-motor" in referring to these stimuli.

Performance Testing

It is clear that "skills" as here defined can be assessed only through actual performance. As Bellows (2) points out, performance tests, ranging from informal probationary procedures to objective systematic measures, form one of the oldest approaches to personnel evaluation. One would expect performance tests to provide a better estimate of job ability than written tests if for no other reason than what Cronbach (7) called "the common sense rule that the test which resembles the job ought to predict the job." In fact a sample of job performance, if properly taken, constitutes the best available criterion of job ability and has been used for that purpose in validating other types of job tests. To serve as a measure of ability, a job sample must (a) be representative -- it must represent a sufficient range of job activities in the proper proportion, (b) be reliable -- it must sample a sufficient number of activities to provide a stable estimate, and (c) have fidelity -- it must represent all critical conditions that would influence performance in the actual job situation.

The prime liability of performance tests is their cost. One cost is that of materiel. For example, an individual post office could scarcely maintain a Letter Sorting Machine solely for the purpose of assessing proficiency in maintaining that piece of equipment. Nor is there any guarantee that an operational machine could be provided whenever it was needed for testing purposes. The other major cost is time. Where a job represents a complex of many diverse tasks, a sizeable segment of job performance must be measured before a valid overall estimate of job capability can be obtained. For example, an accurate measure of a postal equipment mechanic's ability to diagnose failures and make repairs would require exposing him to a sizeable portion of the many thousands of breakdowns that can occur. It is unlikely that either examiner or examinee could be spared for the length of time required unless the test were part of a training program.

Partial Performance Tests

One way of reducing the cost of performance testing is to limit its application as much as possible to measurement of skills and to resort to more economical written tests to assess job knowledge. In maintenance, for example, the skills associated with symptom recognition (perceptual), soldering (motor), or failure diagnosis (cognitive) could be assessed through a battery of performance tests, while knowledge of maintenance procedures, normal operating characteristics and tolerances, or proper supply procedures could be dealt with through written tests. Many mixed batteries of written and performance-oriented job tests have been administered in a variety of operational and research settings. However, no cases were discovered in which tests designed to measure specific job components were correlated with a measure of total job performance.

Simulated Performance

Another popular approach to dealing with job skills is to simulate job requirements and measure resulting performance. Simulators have been used extensively in training and in instructional testing and have

shown a fair degree of correspondence with actual performance as measured by the ability of trainees to transfer skills from simulated to job situation. The degree of correspondence is largely a function of the fidelity of simulation, fidelity in this case being measured in terms of the extent to which simulated tasks call for the same performance as do operational tasks (not pure physical similarity). A dynamic operational simulator such as the Detex trainer used to train Letter Sorting Machine operators provides a more valid assessment of letter-sorting skills than would a static mock-up with a simple dummy keyboard. Since a mechanic must deal with all of the individual equipment piece parts, a fully dynamic simulator for him would be almost indistinguishable from an item of operational equipment.

Simulation can be applied to components of a task as well as the entire task. The use of "part task simulation" in maintenance has been concentrated on the cognitive skills involved in troubleshooting as this has been viewed as the most difficult aspect of the job. The Most elaborate of these simulators displays failure symptoms in terms of voltages, RPM, pressure, and so on, in a manner and according to a pattern that simulates the pattern of real equipment breakdown. The mechanic must formulate a diagnostic approach, collect and interpret symptoms, and infer trouble causes just as he would for operational equipment.

At the other extreme are simple paper-pencil devices that describe symptoms in verbal terms but still call upon the examinee to describe the diagnostic checks to be made. The simulation in this latter case is not substantially different from the common "open book examination." What is critical is not the format of the examination, but the fact that it requires application rather than simple recall of information. Data provided by Johnson (*16*) supports the value of the simulation or "situational" test, as he describes it, over the pure knowledge test.

Simulation may also be applied to the measurement of perceptual skills. Pictorial tests, for example, may be used to test the individual's ability to recognize symptoms of excessive wear, misalignment, discoloration, and so on. Motor skills such as those involved in soldering or in calibrating delicate equipment might be measured using other than operational equipment.

Assessment of Postal Maintenance Skills

As desirable as tests of real or simulated performance might be in the assessment of postal maintenance skills, development of such tests within the confines of the present study was not feasible. Not that the tests themselves would be a particular problem; performance tests are no more costly or time consuming than written tests to construct. However, the administrative problems that would have to be overcome to permit the large scale application of lengthy performance tests would have been immense. Therefore consideration of job skills was limited to those that could be assessed through written examinations. Further discussion of the potential role that performance tests might play in the evaluation of postal employees is included in Chapter 6.

Motivational Factors

Successful job performance depends not only on the individual's ability to do the work required, but also on his motivation and on his desire or willingness to undertake it. In some cases there may be a considerable gulf between the two. It is possible to acquire substantial ability in a field in which the person has little interest. However, the maintenance positions to be covered by the proposed examination program all require extensive maintenance experience -- even at the entry level. The chances that an individual could survive the period of work or study required to develop this experience without a reasonable level of motivation seems sufficiently small to warrant the exclusion of interest tests from the examination program. It is true that, as technicians ascend to positions of supervision and management, the nature of their work changes and they may find themselves undertaking tasks that are not inherently motivating to them. Yet, such changes are rarely abrupt and it is unlikely that a candidate for a higher level position will lack a reasonably clear picture of what that position entails. While the administration of interest tests on a voluntary basis might be of great personal benefit to postal employees in planning their careers, their use to determine who is entitled to employment or promotion appears inappropriate. For this reason interests were not added to the characteristics to be included among measures of job qualification.

JOB PERSONALITY

Over time, individuals develop characteristic or habitual ways of dealing with other individuals and things. These characteristic ways of reacting constitute what we generally call "personality."[1] An individual's personality has generally been viewed as having some bearing on his job performance. For some characteristics the influence is a very general one -- such traits as "honesty" or "dependability" are prized in connection with almost any job. In other cases the relationship is more specific. Such traits as patience, attentiveness to detail, and neatness, for example, would seem to be of particular value in an individual whose job it is to maintain equipment. For a supervisor, on the other hand, those characteristics may be less important than things like initiative, assertiveness, social sensitivity, or other qualities of "leadership."

By the time an individual reaches maturity he is generally rather set in his ways and his personality characteristics become rather stable -- so stable in fact that they are frequently treated as abilities, for example, "the ability to get along with others." This view has encouraged attempts to measure personality characteristics and use the information as a basis for personnel actions.

---

[1] The psychological use of the term personality is distinguished from its more popular use to refer to a particular form of personality, namely an extroversive, outgoing, sociable character.

The relationship between published measures of personality and success on the job has been discouragingly low. One reason for the low relationship is the fact that the majority of tests used have been rather general in scope, derived from some fundamental concept of personality organization rather than from a study of job behavior. Indeed, it is rare that practitioners or scientists using personality inventories in the job context know what characteristic ways of behavior are related to job success. Their criteria have been global ratings, job longevity, income, or some other remote index of success rather than specified job behavior. While personality measures are quite general in nature, research has shown that an individual's behavior characteristics are rather specific. A worker's "attention to detail" will, for example, greatly depend upon what kind of detail is involved. In view of these factors, it is not surprising that personality measures have not proven particularly useful in establishing an individual's fitness for a particular job.

The lack of a high correlation with performance is not the only argument against the use of personality tests -- many knowledge tests have similarly low correlations. An equally important objection is the lack of any demonstrated causal relation. One may survey the items that constitute most personality tests without uncovering a single question of direct functional relevance to any specific job. On the other hand, an item in a knowledge test is there because its possession is considered to be necessary to job performance, even if the causal relation is not evident in a correlation with a particular criterion. If the individual fails a knowledge test he can seek to remedy his deficiency through study. However, all a personality test establishes is that the examinee resembles people who have, in the main, realized a particular degree of success. While mere association may make the use of personality tests actuarily sound, the practice of discriminating against an individual because he seems to closely resemble an unsuccessful person seems inequitable. Moreover, regardless of how personality tests are applied, objections have been raised to the seemingly job-unrelated inquiries into an individual's private life. This objection has recently crystallized in the form of a ban on the use of personality tests for government employment. For these reasons, the use or development of personality "tests" to establish suitability for postal maintenance positions was rejected.

Job Behavior

Despite the inadequacies of personality tests, some estimate of the individual's characteristic job behavior is desirable. Such estimates have traditionally been provided in the form of ratings of an individual's past behavior by supervisors, and occasionally by colleagues. A major objection to supervisor ratings is their susceptability to prejudice on the part of the supervisor. Yet, the importance of certain personality characteristics to job success is sufficiently great as to make some assessment of them a valuable complement to objective examinations. Within the Army, the proficiency of enlisted personnel is assessed jointly through objective tests of knowledge and performance, and through a subjective "commander's evaluation rating."

The preparation of behavior rating scales was considered not to lie within the scope of the present study and therefore no effort was made to incorporate them into the development of qualification examinations. However, some measure of job-relevant behavioral characteristics is believed to be indispensable to a sound program of worker assessment and suggestions for future development of behavior rating scales appears in Chapter 6.

PHYSICAL CHARACTERISTICS

Physical characteristics that are related to job performance include those of structure, such as height, weight, and reach, and those that are functional, such as acuity, strength, and endurance. The magnitude of the relationship between physical factors and job performance will vary greatly from one job to another. In the Post Office Department, the importance of physical characteristics is probably greatest in the handling of mail, for example, lifting mail sacks. On the other hand, few maintenance tasks are physically very demanding. Therefore, the preparation of physical examinations for mail processing equipment personnel was not considered a worthwhile undertaking.

APTITUDES

The worker characteristics thus far described, knowledge, skill, motivation, personality, and physical characteristics, are all factors that influence the individual's ability to perform a particular job. In considering an applicant for employment or promotion, it is generally his job qualifications that one wishes to assess. However, there are circumstances under which a personnel manager may be as much concerned with the individual's ability to acquire specific job qualifications as he is in the individual's possession of them. If the job under consideration is a highly specialized one for which a course of instruction is required, it would be desirable to determine which candidate represents the most promising trainee. Or, if he is farsighted, the personnel manager may be concerned with a candidate's potential for future growth, the ability to assume positions further up the career ladder than that for which he is presently an aspirant.

An individual's ability to acquire the qualifications demanded by a particular job is generally called his "aptitude" for the job. On the whole, aptitudes seem to be a function of the same sorts of characteristics as are job qualifications. An individual's future acquisition of knowledge, for example, appears highly related to whatever knowledge he possesses at the time. His ability to learn new motor skills is generally discernible in his performance on a variety of manipulative tasks. So closely, in fact, do abilities and aptitudes parallel one another that it is often difficult to draw a distinction between the two. A particular knowledge or skill may both support present job performance and form a foundation for the acquisition of knowledges and skills relating to other jobs.

Early treatments of the subject by Hull (16) and Bingham (4) reserve the use of the term "aptitude" for highly stable characteristics that are presumed to be either inborne or the products of early learning. Abilities that resulted from specific job-oriented instruction would not be considered aptitudes. A distinction such as this is harder to make in practice than in theory, and Wesman (29) points out the difficulty in attempting to differentiate between tests of basic aptitude and tests of past achievement. He points out that both types of tests measure what has been learned up to the time that the test is taken. Recognizing this, Anastasi (1) distinguishes between aptitude and achievement tests on the basis of their intended use rather than how or when the ability was acquired. A test used to measure an individual's present ability to perform is an achievement test; one that is used to forecast future achievement in a new situation may be considered an aptitude test.

Since the objectives of this study were practical rather than theoretical, the functional view of aptitudes was adopted. In deriving tests, an item that is highly predictive of immediate ability to perform would be appropriate for an achievement test; that which is highly predictive of performance at some specified time in the distant future would be considered indicative of an aptitude. On the surface it would appear that immediate performance is best predicted by job-specific tests and long range attainment by tests of more basic, more stable characteristics. While there may be some test items that serve in both capacities, they are likely to be few. In general, the more directed an item is to a specific job, the more it is likely to discriminate against promising candidates who have not had an opportunity to acquire the relevant knowledge and skill. What is important, in any case, is that the distinction between aptitude and achievement is to be made in terms of predictability and not in terms of any psychological concept as to the genesis of the two attributes.

Aptitude Tests. A review of available aptitude tests failed to uncover one designed specifically to predict performance in maintenance of electrical or mechanical equipment. The mechanical aptitude examination administered for entry maintenance positions in the postal service is weighted with mechanical information items of varying relevance to maintenance. More fundamental abilities assessed by this examination include arithmetic, form recognition, size measurement, perceptual speed (locating letters), code learning, spatial relations (figure matching), and spatial visualization. No validity information on this examination is currently available.

Tests commonly used outside the postal service to assess aptitude for work of a mechanical nature include the following:

(1) Mechanical information -- Information relating to mechanical procedures, tools, procedures, vocabulary, and so on.

(2) Spatial relations -- Ability to perceive shapes, forms, and sizes.

(3) Mechanical assembly -- Ability to perceive mechanical relationships involved in the assembly of objects.

(4) Mechanical reasoning -- Ability to discover and/or apply mechanical principles to the solution of problems.

(5) Dexterity -- Ability to perform hand, finger, and arm coordinations required for rapid and simultaneous manipulation of mechanical objects.

Tests of the above nature are contained in various aptitude batteries including the "Minnesota" series of mechanical tests, the Flanagan Aptitude Classification Tests, the Differential Aptitude Test, the General Aptitude Battery, the Army Classification Battery, and the Airman Classification Test, among others.

Validity of Aptitude Tests. While personnel engaged in mechanical occupations generally score higher than the general population in the various functions described above, the relation of mechanical aptitude test scores to performance is weak. Low correlations are reported by Ghiselli (*13*), Krech and Crutchfield (*17*), Trattner (*25*), and Biesheuvel (*3*). It is very difficult to evaluate the worth of existing aptitude tests for predicting progress in maintenance of postal equipment. One reason is that aptitude tests, like personality tests, have largely derived from what Thorndike labels the "trait" approach, one in which "test development is based on the general qualities of the individual rather than on the characteristics of a specific job" (*24*); and as Dunnette (*9*) points out, the traits and their names are "based on the investigator's knowledge or presumptions about the content of the tests making it a factor, rather than on any effort to classify observed behavior outside the test." Thorndike found in reviewing the use of aptitude tests to select World War II aviators, that tests constructed from a trait approach failed to attain the predictive validity of a more complex test designed to reflect the combination of demands that existed in a particular job.

Another deficiency of aptitude tests, as we have defined them, is the criteria used to validate them; often they are highly dubious indices of job potential. One common criterion is success in training. There is nothing inherently wrong with this criterion since it is the ability to learn, after all, that aptitude tests are supposed to forecast. However, the content of training, particularly in technical fields, is often highly theoretical and places a premium on verbal skills that are not always relevant to ultimate job performance. Where criterion measures are acceptable indices of job success, they are often collected at the same time the tests are administered or a short time thereafter. It is not surprising that these tests emerge laden with information items that are heavily weighted in terms of specific job information.

Aptitude Tests for Postal Maintenance. The qualifying examinations currently administered to personnel engaged in maintenance of postal equipment appear to embrace both aptitude and achievement. It is the

aptitude items, calling for verbal and numerical reasoning, facility with spatial relations, and form perception, among other skills, that have been most severely criticized by maintenance personnel and supervisors. Both in their public testimony and interviews conducted during the early stages of this study, maintenance personnel have contended that an individual who bids for a vacant position should be evaluated on his ability to handle that position. This contention is voiced not only by technical personnel wishing to advance, but by supervisors whose work loads demand that positions be filled by personnel capable of assuming assigned responsibilities. The fact that 85% of postal workers are employed in the lowest five grades and that 80% never progress beyond the PFS level at which they enter the service would support a major emphasis on the satisfaction of immediate job needs.

On the other hand, the postal service does have a legitimate concern for the source of its future supervisors and maintenance managers. This source cannot be choked off by over attention to immediate needs of lower level technical jobs. Both aptitude and achievement must be considered. What is important is that the two be clearly differentiated so that personnel managers can weigh both factors in considering a particular applicant.

Since aptitude measures are generally believed to predict a broader range of behaviors than achievement tests, it seemed likely that many of the aptitude tests currently on the market could be used to predict success in the electromechanical area. Any attempt, therefore to develop and undertake long term validation of a specially prepared aptitude measure did not appear an efficient use of research resources. However, it was recognized that the selection of an appropriate aptitude measure must be based on consideration of the specific job qualifications that individuals must acquire. Therefore, one objective in the identification of job qualifications became that of providing information that would assist representatives of the Post Office Department in selecting an aptitude measure that was better suited to its mission than the aptitude measure now in use.

SUMMARY OF POSTAL EQUIPMENT MAINTENANCE EXAMINATIONS

It is apparent that out of the broad range of characteristics that influence a worker's ability to maintain electromechanical equipment, only a few are suitable for inclusion in a large scale program of qualification examinations under the conditions that now prevail within the postal service. Others might prove feasible for inclusion in such a program given substantial changes in these conditions, while still another set of job-related characteristics seems to lie completely outside the zone of consideration.

The major conclusions resulting from the study of worker characteristics described in this chapter may be summarized as follows:

(1) Tests of job knowledge can be developed which will show an acceptable relation to job performance so long as they emphasize job

specifics rather than general facts, simple terminology, or broad theory.

(2) Job skills can be measured only through some form of actual or simulated job performance. The development of total job performance measures does not lie within the scope of the present study. However, certain perceptual and reasoning skills may be assessed through written tests that approximate certain aspects of job performance.

(3) An individual's aptitude for acquisition of future job qualification knowledges is important to advancement and therefore should be assessed. However, aptitudes should be clearly distinguished from measures of achievement or present ability.

(4) While personality factors are related to job performance, they are not readily amenable to assessment through tests. Ratings of specific behavior are likely to provide more valid indices of future job performance than test scores and should at some time be entered into a program of personnel assessment.

(5) Job motivation is reasonably well assured through the worker's efforts to acquire the skills and knowledges needed for a particular position and his application for that position. Objective measures such as interest tests, while informative, should not be used as a means of selection.

(6) Aptitudes related to maintenance of electromechanical equipment are likely to prove predictable by various of the aptitudinal batteries now available. The present study of job related characteristics should aid the Post Office Department in seeking appropriate measures.

(7) The physical factors associated with postal equipment maintenance appear to be negligible and therefore not of major concern in a program of qualification examinations.

Chapter 3

PREPARATION OF EXAMINATION ITEMS

Having surveyed the qualifications that underlie maintenance of electromechanical equipment, and having selected those that appeared amenable to assessment through a formal program of examination, the next step in the development of qualification examinations was to prepare an appropriate set of test items. In this chapter the methods of generating the test items to measure the skills and knowledges involved in the maintenance of electromechanical equipment will be described.

GENERAL APPROACH

The conventional approach to the development of a job qualification test is to assemble a group of content specialists or "experts," and to ask them to prepare a set of questions that they believe will measure the skills and knowledges involved in a particular job. The test item "pool" that results is then submitted to test specialists who eliminate items that appear to be ambiguous, too easy, too hard, irrelevant, or, for some reason, unacceptable. The remaining items, after some editing, are usually administered to a sample of personnel representing the workers to whom the test will ultimately be applied. Individual items are then examined for their correlation with some measure of job competence or with the overall test score, and those with the highest correlation are chosen for the final tests.

While a great deal of attention is typically given to the selection of items from the item pool -- they may be tried out many times before a test is ultimately constructed -- relatively less consideration has been given to how the items entered the pool in the first place. The essential job relevance of the original item pool is assumed more or less out of respect for the experts who contributed them. However, there are grounds for questioning the ability of many "experts" to prepare job-related questions. First, even when the expert is a job-holder, one may legitimately question whether he is able to provide an unbiased and accurate description of the qualifications required in his own work. He cannot be expected to "follow himself around" making careful notes of the skills and knowledges he applies to each task. Rather, he is likely to rely upon what he has been taught, what he finds most interesting, what he thinks will impress, or some other body of qualifications that may not be representative of those that guide his day-to-day activities.

Secondly, it often turns out that the "expert" is not really a worker but an individual believed to be qualified in the subject matter

of the job. For example, items for maintenance examinations are frequently prepared by instructors, engineers, or technical writers. Although they may have been employed as mechanics at one time, their memory for the knowledges and skills they once had is likely to be vague. It is for this reason that maintenance examinations often lean heavily on engineering characteristics of the equipment, or on various technical facts or theory of operation, rather than on maintenance itself.

No amount of psychometric manipulation can accord validity to a set of test items that was not job-related in the first place. Correlation with some criterion of job proficiency can only skim off the best of what was furnished. Internal consistency statistics, such as item-test correlations, cannot even do this, but simply succeed in orienting the final test to those knowledges and skills that fortuitously dominated the original item pool.

Analytic Approach

If the test developer cannot rely on others to tell him what personnel qualifications underlie job success, he must discover them himself. The only way he can do this is to examine the job behavior in question and make inferences as to the qualifications that guide this behavior.

The analysis of job qualifications has been the subject of widespread attention during the past two decades, particularly in the military services where it is central to programs of selection, training, classification, and assignment. Unfortunately, this attention has not produced an abundance of methodology or data (*8*). The problem lies not so much in the process by which qualifications are inferred from job descriptions but rather with the job descriptions themselves. Generally speaking, the more detailed the description of behavior, the more accurate will be the inferences that are drawn from them. The statement "repair the letter sorting machine" or even "replace the coding bar" tells little about the mental and physical equipment a worker needs to carry out these activities. It is only when a step-by-step process, that is, the actual job behavior, is described that the knowledges, skills, personality factors, and physical characteristics that relate to the behavior can be reliably inferred. Highly detailed job descriptions are not unknown -- they characterized the earliest time and motion study. But as jobs have become more numerous, and more varied in character, it has been increasingly costly to maintain a highly detailed level of description.

In recent years a return to a more detailed level of job description may be observed in connection with the development of large military equipment systems. Two factors seem primarily responsible. First, in order to assure that qualified personnel would be available to man new systems, it has been necessary to select, train, and classify personnel while the system is still on the drawing board. Since there are no workers to be studied, the job analyst has no alternative but to study the characteristics of the system in order to determine what specific

behaviors it will demand. Second, the development of an equipment system is usually managed by a single agency that is given the responsibility for drawing together all requirements and resources pertaining to it. By combining all requirements for job analytic data and pooling available resources, it has been possible to furnish personnel and training agencies information at a greater level of detail than they could have obtained by operating independently.

It was the conviction of the researchers that the analysis of job behavior provided a surer route to the improvement of postal examinations than did any statistical reworking of the old test items or an attempt to obtain new items from the old sources. The fact that the development of examinations was combined with an attempt to prepare more detailed job descriptions and to improve classification of postal positions, and could share the cost of a job analytic program with these efforts, obviously weighed heavily in the decision to adopt this approach.

THE JOB ANALYSIS PROGRAM

An analysis of electromechanical maintenance jobs within the postal service was launched in July 1968. This analysis was performed to obtain data for use in preparing detailed job descriptions and revision of the job classification structure, as well as in the development of new examinations. It is fully described in a separate report (*19*) and will be only briefly summarized here.

Identification and Analysis of Tasks.

The term "job analysis" has been broadly applied to any collection of information about jobs. We have used the word "analysis" in its strictest sense, meaning a reduction into basic elements. The job analysis that was performed was therefore a reduction of jobs into their fundamental elements of performance. In truth, it was not existing jobs, but rather a broad range of electromechanical maintenance activities toward which the analysis was directed. This enabled the analysis to be used in the creation of new jobs under the classification phase of the project.

Identification of Tasks. The first step in the analysis was the classification of maintenance tasks. A task is a specific thing to be done, such as repairing a faulty vacuum pump on the Letter Sorting Machine, or replacing a bad bearing in a conveyor system. A task has a separate beginning and termination; it is not part of another activity. Since maintenance is directed toward equipment, it was primarily through study of the equipment that the maintenance tasks were identified.

Equipment maintenance requirements within the Post Office Department have been classified into the following major categories:

Routine Preventive Maintenance	Inspection	Modification
Cleaning and Lubricating	Troubleshooting	Installation
Alignment and Adjustment	Repair	Overhaul

These major categories of activities are frequently called "duties" or "responsibilities." Strictly speaking, installation is not a form of maintenance, but since it is frequently assigned to maintenance personnel, it becomes a maintenance duty. The actual activities that comprise the various duties will naturally differ from one item of equipment to another. It was therefore necessary to examine each piece of equipment separately. One may view the duties and the equipment items as constituting the coordinates of a matrix, with each cell encompassing a specific set of tasks.

In seeking out maintenance tasks, reference was made to manufacturers' engineering and technical manuals, post office route sheets (checklists describing scheduled maintenance tasks), blueprints and drawings, and interviews with maintenance personnel. Scheduled maintenance tasks were easily identified through the various scheduling documents. However, unscheduled maintenance tasks -- maintenance arising through the occurrence of breakdowns -- are considerably more difficult to identify. In theory, each of the thousands of electromechanical parts from which post office equipment is assembled represents a potential task in the sense that each requires a somewhat different set of activities for its diagnosis and correction. To identify and analyze each separate task would be prohibitively expensive. More importantly, it would be unnecessary since the skills and knowledges required of maintenance personnel are highly similar across repair tasks. This means that a reasonably large sample of repair tasks would suffice to show needed skills and could serve adequately as the basis for developing examination items.

The supervisor's job, while it encompasses all that a mechanic does, also includes a variety of administrative tasks that are not so conveniently described as are those tasks concerned directly with equipment. A portion of the supervisor's responsibility is codified in the form of official Post Office Department policy and the National Agreeement of postal employee organizations. However, much of what the supervisor is called upon to do from day-to-day is not written down and could be identified only through extensive interviews with and observation of postal maintenance supervisors.

Once it had been identified, each task in the sample was analyzed into the steps required to carry it out. These steps became the elements of the task analysis. An example of a task element would be "opens panel door and feels for loose components and connections," or "removes set screws from wheel collar and slips the collar off the axle." Note that the element includes a detailed description of the behavior involved. A description such as "checks panel" or "removes collar" would not have been sufficiently detailed to provide any real idea of what was required. Any "cues" that guided the mechanic were also described, for example, "moves sprocket laterally until chain has 1/2-inch sag."

Identification of Qualifications

The purpose in analyzing tasks was to identify those characteristics that enabled men to carry them out. These enabling characteristics constituted the individual's qualifications for the job of which the particular tasks were a part. Foremost among these qualifications were the job knowledges that guided the individual's activities. Each item of procedural, technical, or theoretical information was recorded adjacent to the task element to which it corresponded. This was done while the task was being analyzed, since in most cases the knowledges that enabled the worker to carrry out the task were the same that enabled the analyst to perform his analysis.

All perceptual, motor, and cognitive skills were described in as much detail as possible. Descriptions of perceptual skills were concentrated on the nature of the stimuli to be perceived. Motor skills were described in terms of the individual movements to be performed or coordinated. Descriptions of cognitive skills dwelt primarily on the elements of the reasoning process. Since "skills" as they are defined in this report involve processes that are not readily describable, the descriptions that were offered were not intended to communicate any deep understanding of the skills involved. The purpose in providing the descriptions was less to explain them, than simply to identify for users of job data those tasks that involve more than the mere acquisition of information -- tasks that would require considerable practice before they could be performed adequately. It was believed that this intelligence would be of value to personnel managers in identifying tasks that (a) would require practical, hands-on instruction, (b) require performance tests for proficiency assessment, and (c) would be appropriate for higher skilled, experienced, senior personnel.

THE NATURE OF ELECTROMECHANICAL MAINTENANCE QUALIFICATIONS

The results of the job analysis are best viewed in the position descriptions and in the job analysis data sheets provided as a part of the job analysis report (*19*). This chapter will merely summarize the types of knowledges and skills found to be related to the maintenance of postal equipment.

Scheduled Maintenance

The knowledges involved in the performance of scheduled maintenance are largely procedural. Mechanical inspection is primarily concerned with equipment deficiencies that can be readily observed and identified without elaborate testing. These deficiencies include breakage, misalignment, dirt, grease, excessive noise, vibration, heat, and looseness. The routine preventive maintenance, cleaning and lubrication procedures, required to overcome these deficiencies are, on the whole, about as obvious as the deficiencies themselves. Examples are removing dirt and grease, tightening bolts, belts, and chains, lubricating bearings, pulleys, chains, and correcting minor misalignment.

While a majority of corrective maintenance time is consumed in relatively routine tasks, some portion is spent in activities that demand a degree of technical knowledge of the following type:

(1) General Maintenance Practices: Knowing the proper methods of cleaning and lubricating, knowing safety practices in dealing with mechanical and electrical equipment, and knowing the indications of wear to belts, pulleys, bearings, and so on.

(2) Location and Identification: Knowing where assemblies and parts are located and being able to identify them.

(3) Normal Operating Procedures and Characteristics: Knowing the steps involved in operating equipment and the characteristics of normal operating, for example, RPM or response time.

(4) Specific Maintenance Practice: Knowing procedures for testing, adjusting, cleaning, lubricating, and servicing specific equipment items.

(5) Equipment Idiosyncrasies: Knowing the operating peculiarities or particular maintenance needs of an equipment item at a particular installation.

Some of the indications of potential or real equipment breakdown are not readily expressed in verbal terms and therefore require a degree of perceptual skill. For example, how hot is too hot? How much wear constitutes a "worn" bearing? How much vibration is acceptable? The development of the appropriate "mental images" demands considerable experience in perceiving normal and abnormal indications, and therefore qualifies as perceptual skill in terms of this discussion. Turning to motor skills, few tasks other then soldering, welding, and a few delicate adjustments require highly specialized or complicated response patterns.

Unscheduled Maintenance

Unscheduled maintenance, arising from equipment breakdowns, tends to be somewhat more varied and is therefore somewhat less easily reduced to procedures than is scheduled maintenance. The most challenging aspect of unscheduled maintenance is that of troubleshooting, that is, locating the source of the breakdown. Sometimes the source is readily observable as a broken conveyor belt, a burned-out motor, or a parted cable. However, where the cause is some relatively small part, the repairman must seek it out. In a few cases, manufacturers have prepared a set of diagnostic procedures that will lead the repairman to the source of most troubles. Unfortunately, this type of job aid is not commonly furnished with postal equipment.[1]

[1] See Trexler and Butler (*26*) for a discussion of troubleshooting procedures in connection with postal equipment.

Where troubleshooting procedures are not provided, the repairman must apply some knowledge concerning the nature of the equipment to isolation of the faulty part. In some cases, the cause may be inferred directly from observed symptoms. For example, mail accumulating on top of the Letter Sorting Machine is likely to be caused by misalignment of the decoder assembly. Sometimes the symptom-cause information is generated from the mechanic's experience, while in other cases it derives from a knowledge of what each part is supposed to do.

As equipment grows in complexity, the interrelationships among the various parts make it difficult to associate a particular symptom with any one part. Rather, the repairman must undertake a series of checks to progressively narrow down the trouble source until the faulty part is pinpointed. The repairman's ability to plot and carry out an efficient series of checks depends on his knowledge of the equipment's internal operation, his ability to perceive the interrelationships involved, and his capacity for making the logical inferences necessary to identify the cause of the failure.

Repair, that is, the removal and replacement of a faulty part, is generally less complicated than troubleshooting. In many cases the steps involved are rather obvious from the way the equipment is put together. Where a mechanism or assembly is particularly complex, diagrams or instructions are generally provided to aid the repairman in assembly and disassembly. Sometimes a degree of perceptual skill is involved in seeing how parts fit together, but often the only motor skills required are those involved in soldering or welding.

Supervision

Like his mechanics, the Mail Processing Equipment supervisor[1] devotes a considerable amount of his time to actual maintenance. The tasks that occupy his attention are guiding difficult repair jobs, expediting emergency repairs to urgently needed equipment, inspecting critical items for incipient breakdowns, establishing safety precautions to prevent personal injury and damage to the equipment, installing and modifying equipment and submitting occasional recommendations for minor design changes. To perform these duties well, the supervisor must possess a thorough understanding of his equipment, its design characteristics, its theory of operation, and its individual idiosyncrasies. Having little to do with the more routine aspects of maintenance, the supervisor would not be expected to have retained a detailed knowledge of routine maintenance procedures or of the volume of technical information that accompanies their application.

As the first line administrator of maintenance policy, the supervisor is called upon to establish work assignments, help assign

---

[1] Throughout the remainder of this report reference will be made to "supervisor" rather than foreman since the foreman term is generally used in connection with qualification examinations.

priorities of repair, adjust work schedules to cope with absences and other contingencies, conduct training, and prepare equipment reports and recommendations. The ability to administer effectively requires at the very least a knowledge of the maintenance policy to be administered, primarily that set forth in *Maintenance Management Facilities Handbook* (*28*). However, since official policy cannot be expected to anticipate all eventualities, the supervisor must exercise considerable judgment in dealing with individual maintenance problems as they arise. As a decision maker and problem solver he must possess and be able to apply a knowledge of individual worker strengths and weaknesses, the roles of various equipment items in mail processing, and various costs including both maintenance costs and operating costs associated with equipment downtime.

As a manager of people, the supervisor becomes an executor of Post Office Department personnel policies, as described in the Postal Manual and in various local guides to personnel procedures. In his capacity as a personnel manager, the supervisor reviews leave requests, deals initially with unauthorized absences, recommends workers for promotion, processes employee suggestions, listens to grievances, and counsels workers on a variety of job-related problems. Effective personnel administration is important to any organization, however, maintenance supervisors have objected to the degree of emphasis given to questions on personnel administration in existing supervisory examinations -- at least in relation to the meager coverage of maintenance administration. Their objections are supported by the results of the job analysis survey (*19*) which showed that relatively little time was devoted to personnel matters.

Underlying all of the supervisor's actions is a need to maintain an acceptable level of productivity on the part of his work force -- to establish a working relationship between himself and his workers, as well as among the individual workers, that will lead to an effective maintenance operation. To assess the supervisor's capacities in this area, existing supervision tests contain a substantial number of items dealing with the supervisor's ability to provide effective "leadership," that is, items dealing with his "interpersonal relations."

The desirability of this type of question was challenged in the present study on two counts. First, the existence of any single set of behaviors that could be said to constitute "effective leadership" can be questioned. Individual employees respond differently to a particular approach, so that what constitutes good leadership for one employee may not be effective for another. Secondly, in this area particularly, there is often a substantial difference between knowing what is desirable behavior and actually exhibiting it. Whether a supervisor attempts to "understand a worker's viewpoint" or attempts to "involve him in decision making," for example, is as likely to be a function of his own basic personality patterns as it is his knowledge that such is considered good leadership. To what extent tests of interpersonal relations measure actual supervisor qualifications as opposed to simple "book knowledge" is an open question. In any case,

because of the investigators' doubts as to the validity of this type of test item, it was eliminated from the study with the approval of the Post Office Department.

PREPARATION OF TEST ITEMS

The basic source of content for the test items was the description of knowledges and skills that grew out of the analysis of maintenance and supervisory skills. In some cases these descriptions furnished all of the information required for the test item, while in other cases the descriptions referenced information contained in related technical manuals, policy manuals, and textbooks. In general, the types of items assigned to each form were as follows:

Apprentice Test

1. Common maintenance procedures including use and care of tools, safety precautions, preventive maintenance and repair of common components such as motors, bearings, and mechanical linkage.
2. Preventive maintenance and minor repairs of common equipment in use by the post office including conveyors and communication equipment.

Since the Apprentice test would be administered to applicants from outside the maintenance craft, and in some cases outside of the postal service itself, a "passing score" would be based solely upon items from the first category.

MPE Mechanic Test

1. Items similar to but more complex than those appearing in the Apprentice test.
2. Preventive maintenance and minor repairs to equipment that is unique to the post office.

To permit application of the MPE Mechanic test to personnel outside of the postal service and outside of the maintenance crafts,[1] only items from the first category would be used in determining the "passing" score.

MPE Senior Mechanic Test

1. Items similar to but more complex than those appearing in the Intermediate test.
2. Items related to the major repair of all types of equipment including items dealing with such information as normal operating values and tolerances, symptom-cause

[1] Many promising candidates for MPE Mechanic positions would not be willing to enter the maintenance craft at the PFS 5-6 level.

relationships, troubleshooting procedures, and principles of operation.

Supervisory Test

1. Items similar to the more complicated items from Category 2 of the Senior Mechanic test.
2. Factual items concerned with maintenance and personnel administrative procedures.
3. Items calling for judgment in applying maintenance and personnel policy to the solution of supervisory problems.

GENERAL CONTENT CONSIDERATIONS

It was not possible within the time and money constraints of the present study to pioneer novel item formats. Under agreement with the Post Office Department the five-alternative multiple-choice type of items was continued. However, a number of steps were taken to improve the value of the multiple-choice items in assessing underlying skills and knowledges. The most important of these steps were the following:

(1) Readability. Vocabulary and grammar were made as simple as possible, within the limits imposed by the technical nature of the work, in order to minimize the role of verbal intelligence and maximize that of specific job knowledge and skill. Where practical, diagrams were used in lieu of verbal descriptions.

(2) Terminology. The role of terminology was greatly reduced. Items involving pure nomenclature were eliminated. While this type of item has been the mainstay of technical examinations, there is little evidence that knowing what something is called is critical in dealing with it. Moreover, terms tend to differ from one situation or location to another. In addition to eliminating pure nomenclature items, an attempt was made to reduce the dependency upon terminology in general. The use of diagrams instead of terms to represent things was a step in this direction.

(3) Application. Where possible, questions called for the application of information rather than its mere recall. The purpose of this was to allow the exercise of reasoning skills involved in solving troubleshooting and repair problems. While it would have been desirable to require the examinee to work problems from the actual technical manuals used on the job, the availability of such manuals was found to be very uncertain. Therefore, mechanical and electrical schematics were provided with the test. Some of these represented items of postal equipment, while others were schematics created solely for test purposes.

MINIMUM STANDARDS

To be employed as a selection device, a test must have a "passing" score, a score below which candidates will not be considered for employment or promotion. The cutting score itself is usually set at a level that will assure an adequate supply of personnel and provide an acceptable probability of success. However, the usual passing score represents a total of correct answers and can be attained on any combination of test items; it cannot be directly related to any particular body of knowledge that the examinee possesses. The passing score has no "absolute" meaning. This is a distinct handicap to a personnel manager attempting to relate an individual applicant's abilities to the needs of the job.

To help overcome this apparent deficiency, an attempt was made to prepare, for each PFS level, a set of items that could be logically viewed as representing minimum knowledge standards. To be considered "qualified" for the position he seeks, the candidate would have to be able to answer correctly all of these items (or almost all, allowing for a small margin of error). Items to be counted toward a passing score would meet the following criteria:

(1) The general area of content should be one with which all applicants taking a test can be expected to be familiar. This excluded items dealing directly with equipment specific to the Post Office in the cases of the Apprentice and Intermediate tests.

(2) The items should be viewed by job incumbents at the level for which the test is intended as an item that everyone should be capable of passing.

(3) A sufficient number of examinees should pass the item to assure an adequate supply of successful applicants.

Chapter 4

PRELIMINARY ADMINISTRATION

Using the results of the job analysis, a pool of 384 multiple-choice test items was developed. The next step in the study was to administer the items to a sample of maintenance personnel in order to determine the relation between performance on the item and indices of job proficiency.

PREPARATION OF PRELIMINARY TESTS

Each of the 384 items was assigned to one of the four tests -- Apprentice (98 items), Mechanic (70 items), Senior Mechanic (136 items), and Supervisor (80 items). Since the examinations were being developed for a set of job positions whose status was only that of a proposal, a true sample of job incumbents did not actually exist. However, inasmuch as the proposed positions had been converted from existing positions, it was not difficult to pair each examination with the position representing the same set of job duties. The examinations and their corresponding positions were:

Proposed Position	Existing Position
Helper (PFS-4)	Helper (PFS-4)
Apprentice (PFS-5 & 6)	General Mechanic (PFS-5)
Mechanic (PFS-7)	MPE Mechanic (PFS-6)
Senior Mechanic (PFS-8)	MPE Mechanic (PFS-7)
Supervisor (PFS-10)	MPE Supervisor (PFS-9)

The plan of administration for the preliminary tests was to give each test not only to personnel in the appropriate position, but also to personnel at those PFS levels immediately above and below it. For example, the Senior Mechanic's test would be administered not only to the level 7 Mechanics, but to the level 9 Supervisors and level 6 Mechanics. The purpose of this multiple administration was to guard against loss of an item in the event the researchers had misjudged what level was appropriate. For example, an item intended for the 7-level position might become, on the basis of the preliminary administration, more appropriate for level 9 or level 6 personnel. The administration plan is depicted in Figure 1.

PLAN OF ADMINISTRATION OF PRELIMINARY TEST ITEMS

Test	Examinees: Helper (PFS-4)	General Mechanic (PFS-5)	Mechanic (PFS-6)	Mechanic (PFS-7)	Supervisor (PFS-9)
Apprentice	X	X	X		
Mechanic	X	X	X	X	
Senior Mechanic			X	X	X
Supervisor				X	X

Figure 1

Time limitations prevented the trial administration of entire tests to examinees at the 6, 7, and 9 levels. Therefore, all tests were divided into two forms of approximately equal length with one form of each examination being administered to random halves of the examinee sample at the three levels.

Sample. The preliminary examinations were administered January 21-31, 1969, to personnel at the 13 most highly mechanized U.S. post offices, located in Buffalo, Chicago, Cincinnati, Denver, Detroit, Houston, Los Angeles, Miami, New Orleans, Omaha, Portland, Sacramento, and St. Paul. All personnel on all tours were examined, excluding only those unavailable owing to annual or sick leave. The total numbers for each job position are: Helper (PFS-4) 89; General Mechanic (PFS-5) 32; MPE Mechanic (PFS-6) 351; MPE Mechanic (PFS-7) 118; Supervisor (PFS-9) 62. While several factors make it difficult to determine what percent of assigned personnel these numbers represent; it may be safely estimated that the figure is in excess of 80%. What is critical is that there did not appear to be any selective factors operating to make the examined sample unrepresentative of the total population of personnel defined by the post offices studied.

CRITERION INFORMATION

The purpose of the preliminary administration was to collect information that would be of value in determining the validity of items for the identification of qualified personnel. One index of validity would be the relation between an individual's performance on an item and his position within the career hierarchy. On the whole one would expect workers in a particular job to be better able to answer a job-related question than individuals who are somewhat lower in the hierarchy. If, for example, an item is to be used to select MPE mechanics, then MPE mechanics[1] should answer it correctly with greater frequency than Apprentices. If this were not the case, one might with good cause question whether the item assesses knowledges and skills that are related to the mechanic's job.

A second index of validity to be collected was a ranking of all personnel on a particular tour by that tour supervisor. The ranking was to be performed in terms of the overall proficiency of each individual in the maintenance of postal equipment. By making distinctions among individuals holding a given job position, the rankings would provide a more refined index of proficiency than would job position alone. It would also correct for situations where an individual at a lower level was for one reason or another more proficient than some

[1] To avoid using two sets of terms, reference will be made to proposed positions rather than those existing at the time of the study. The reader should bear the fact in mind when proposed position titles are used in connection with results of the preliminary administration.

individual holding a higher position. The use of ranks rather than an absolute rating was intended to force supervisors to make distinctions among subordinates and not to allow them to rate everyone as "good" or "bad." Since rankings provide a measure of relative status within a group, the method eliminates any differences among different groups. While this arbitrary equating of groups has the advantage of eliminating inter-rater differences, that is, differences due to variation in standards employed by individual supervisors, it has the undesirable effect of eliminating any true differences among the groups as well. Unfortunately, there is no way in a rating system of distinguishing inter-rater from true inter-group differences. After consideration of the situation in which ratings would be collected, the researchers chose the ranking approach as being the most likely to provide a valid indication of ability.

ADMINISTRATIVE PROCEDURES

Subjects were administered the preliminary examinations in small groups, generally from one-quarter to one-half of a particular tour at a time. They were informed that the purpose of the examination was not to evaluate the employees, but rather to "test the test," and that the results would be kept confidential. They were invited to make comments concerning the examination items, either orally or in written form on the back of the answer sheets. No time limit was imposed.

In addition to answering each test item, examinees were asked two questions about the item. First, they were asked whether they found the item to be related to their particular job. The obvious purpose of this was to provide a check upon each item's job relevance. The second question asked them to judge whether an item reflected a minimum standard for their job, that is, was the item one that any qualified individual at their level should be able to answer correctly. This information was required in the selection of items to count toward a passing score, as described earlier. Total administration time for the preliminary examinations ranged from approximately one-and one-half hours for the more rapid Supervisors to three-and one-half hours for the slower Helpers.

Upon completing the examination, each individual was administered a job activity questionnaire as part of the job analysis project described earlier. The one related item of information collected on this questionnaire was the indication of the individual's job position. Supervisors, in addition to completing the job activity questionnaire, were also asked to rank all personnel on their tours. The indication of job position and the ranking were the two items of criterion information mentioned previously.

ANALYSIS OF RESULTS

The following statistics were compiled for each examination item:

(1) Job relevance -- the percent of examinees in each job position indicating the item was considered job relevant.

(2) Minimum standard -- the percent of examinees in each job position indicating that the item reflected a minimum standard.

(3) Item response -- the percent of examinees in each job position selecting each of the five alternative answers.

(4) Ranking -- the mean rank of all examinees (excluding supervisors, who were not ranked) selecting each of the five alternative responses.

Before ranks could be averaged to obtain the mean ranks, it was necessary to convert each individual's rank to a normalized score. This was done in accordance with a table prepared by Fisher and Yates (*11*). By this process an individual who, for example, ranked first in a group of three individuals, would receive a score representing a point in a normal distribution corresponding to the midpoint of the upper third of the distribution. This conversion permitted individuals from different size groups to be directly compared and for statistical manipulations to be performed upon the rankings.

Data from answer sheets were entered into the HumRRO IBM 360/40 computer for compilation of necessary statistics. Since the desired information concerned the relation between individual items and criterion variables, no attempt was made to score entire tests or to determine relationships between individual items and a total score.

Chapter 5

DEVELOPMENT OF FINAL QUALIFICATION EXAMINATIONS

The development of a final set of qualification examinations involved (a) establishing the validity of examination items, (b) selecting a set of items for each examination, and (c) assembling selected items into a series of examinations.

OVERALL ITEM VALIDITY

When examining a large number of test items, one may expect that a certain number will exhibit a relationship with the criteria of validity through chance factors alone. For example, some items will be answered correctly more often by mechanics than apprentices even if all items were answered randomly by both groups. It was therefore necessary to inspect the overall distribution of item validity statistics before attempting to identify any individual items as "valid."

Figure 2 represents a distribution of items classified according to the average rank of individuals selecting the correct alternative. The bar labeled "highest mean rating" indicates the number of items where the correct alternative was selected by individuals having the highest mean normalized supervisor ranking -- in short, these are the items in which the "best" people selected the correct answer. Items in the "second highest" category are those in which the correct alternative had the second highest mean rank, that is the "best people" chose some other alternative. In preparing Figure 2, the highest ranking alternative was ignored if chosen by less than 10% of the total sample on the grounds that a mean based on such a small number would not be sufficiently stable to warrant consideration. The "less than 10%" category in the figure indicates the instances in which the correct alternative was selected by less than 10% of the sample.

For Figure 2, it can be seen that in 302 of the 384 preliminary items (79%), the correct answer was selected by those with the highest mean supervisor rating. While it would be difficult to determine what constitutes a "chance" distribution of item alternatives with respect to the rating criterion, it is clear that chance factors cannot explain the obtained result. The fact that the higher rated workers chose the correct answer on the overwhelming majority of items is an indication that the preliminary test as a whole possessed a degree of validity with respect to supervisor ratings.

The examination of each individual item alternative differs somewhat from the more conventional approach to item analysis in which

Distribution of Items According to the Mean Supervisor Rating of Those Selecting the Correct Answer

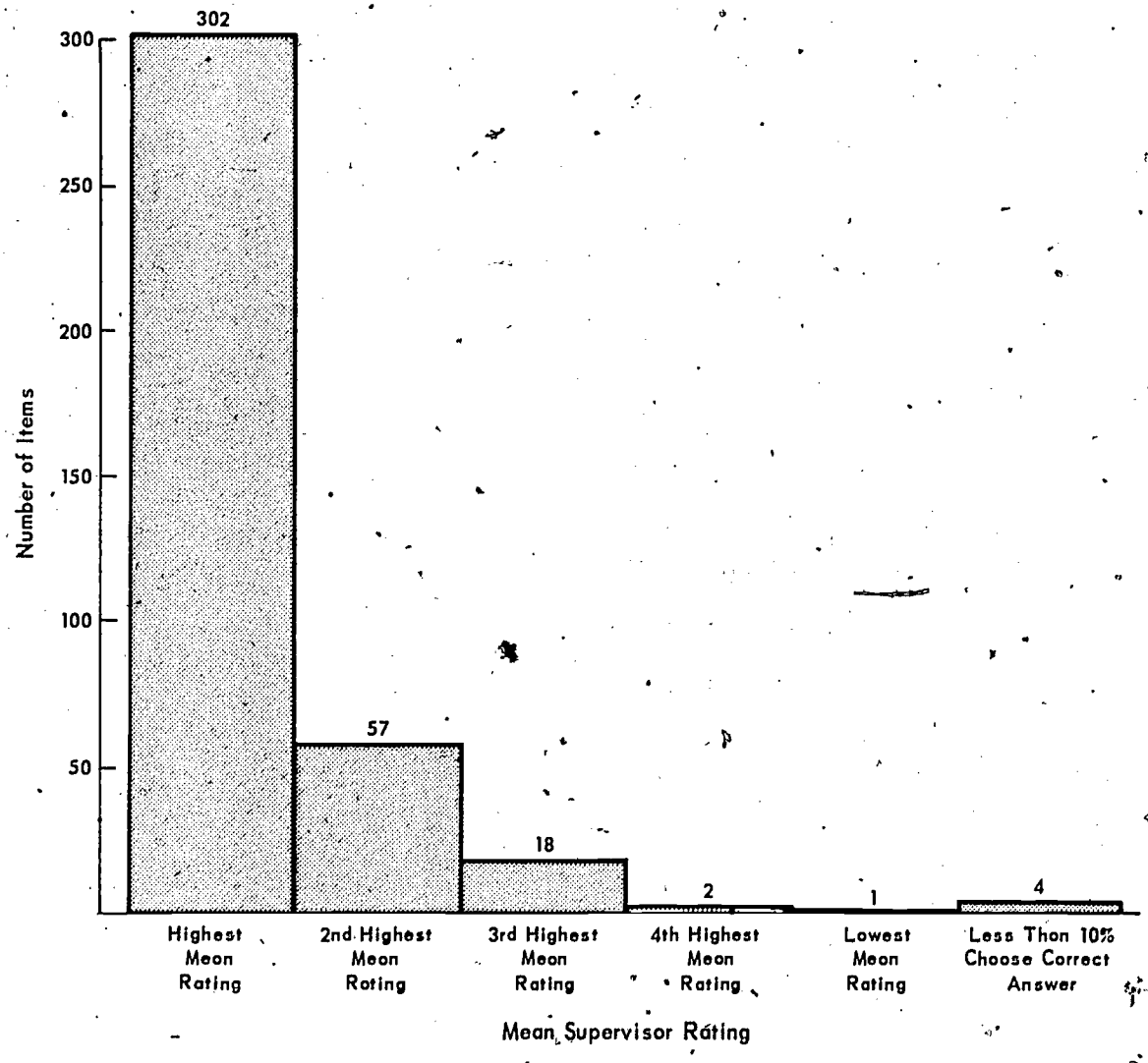

Figure 2

47

results are reduced to "pass" versus "fail" (all distractors are grouped together in a single "fail" category). By the latter practice, an item would be considered valid if the correct answer were favored over the grouped distractors, even though an individual distractor might constitute a better answer according to the criterion employed. The standard imposed in the present study, that the correct answer be favored over all distractors, while a more stringent requirement, seemed a logistically more defensible one.

Figure 3 displays a distribution of items ordered in terms of the difference in the percent of individuals in pairs of job positions getting the correct answer. A "+" score indicates that the difference favors the higher job position; a "-" score indicates that it favors a lower position. For example, were 85% of Supervisors and 73% of Senior Mechanics to have obtained the correct answer to a question, the difference of +12% would be entered in the +10 to 14% interval.

The preponderance of items show a difference in favor of the higher of the two jobs compared in each pair of positions. The most successful distinctions appear to be between the Apprentice and Helper positions, and between the Mechanic and Apprentice positions. The fact that the percentages for Apprentices are based on only 32 subjects accounts in some part for the large number of very sizeable (+20 and over) differences. While the real validity of items in this interval is undoubtedly less than that indicated, it seems safe to say that the majority of items are capable of making valid distinctions among the three lower level groups.

The ability of items to make valid discriminations appears much less at the higher than the lower level groups, although the majority of items still favor the higher of each pair of positions compared. In the case of Mechanics and Senior Mechanics, the relatively small differences reflect a well established similarity in job duties. As for Supervisors and Senior Mechanics, the differences appear to be related to the type of item involved. Those items that favor Supervisors most markedly primarily deal with administrative matters, information concerned with postal policies and procedures. Although certain of the Senior Mechanics, in the capacity of acting supervisor, may occasionally deal with such matters, their familiarity with this type of question should not be as great as that of supervisors.

Turning to the items that fall in or near the "no difference" category (-4 to +4), the majority of them are of a technical nature. While it is true that the Supervisor is expected to provide technical guidance and assistance to his subordinates, his heavy involvement in administrative matters will naturally attenuate to some extent his technical proficiency. It is understandable that items of a technical nature will, therefore, make little distinction between Supervisors and Senior Mechanics.

On the basis of the results displayed in Figures 2 and 3, it appeared reasonable to conclude that the preliminary tests, as a

Distribution of Items According to the Difference in Percent of Adjacent Job Groups Passing the Item

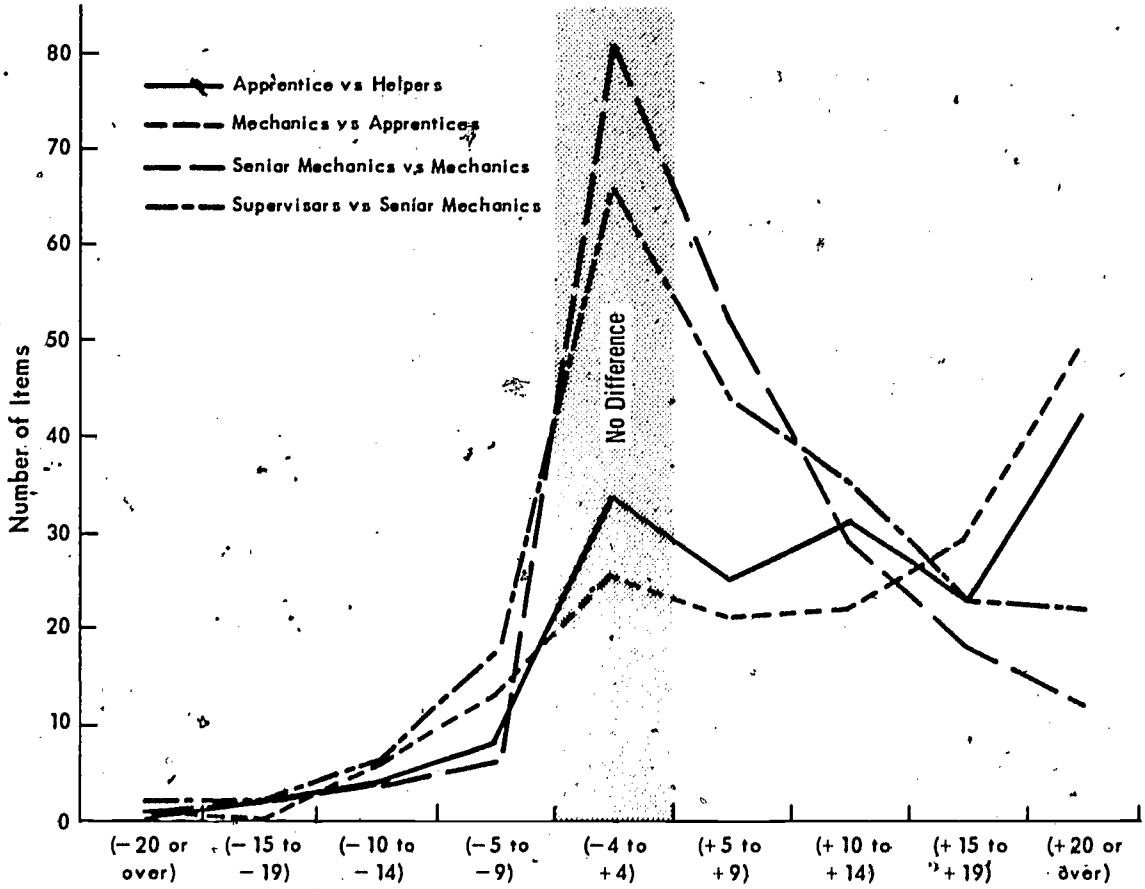

Figure 3

whole, were capable of making valid distinctions of differing levels of rated and classified (job position) proficiency. It was, therefore, possible to proceed with the selection of the most valid-appearing of the items with the expectation that they represented something more than chance fluctuations of invalid items. The degree of validity possessed by the final tests, that is, how accurate the tests would be in distinguishing the better from the poorer workers, would, of course, have to be established through their administration to an independent sample of workers.

SELECTION OF ITEMS

A set of standards for the selection of examination items was developed jointly by representatives of the HumRRO staff and of the Bureau of Personnel, U.S. Post Office Department. Under these standards, an item was to meet these conditions:

(1) The correct alternative must have had the highest mean normalized ranking, that is, be selected by the "best" people.

(2) The item must be considered job relevant by over half of the examinees in the position for which the examination is to be used.

(3) The item must be answered correctly by no less than 20% of the examinees in the position for which the examination is to be used, and no more than 95% of the examinees in the positions from which applicants would come.

(4) The content of the item must be appropriate to the job position for which it is to be used, as described in Chapter 3.

The above general standards did not fit all cases so that it was necessary to introduce certain qualifications to avoid discarding usable items. The first qualification was that an item in which the correct alternative was "second-ranked" might be used if (a) the number of individuals selecting the first ranked alternative was relatively small, that is, between 10 and 20%[1], (b) the mean rankings were close to that of the first ranked alternative, and (c) the percent of correct answers in the position for which the exam was intended was greater than that for the position below it by 10% or more. Where validity statistics favored both the correct answer and some distractor, the latter, while technically incorrect, was generally found to have some degree of truth. To improve such items, the incorrect alternative was replaced by a less attractive distractor. Where this was done, the percent choosing both alternatives was "credited" to the correct alternative in subsequent data treatment.

---

[1] Recall that an alternative selected by less than 10% of the sample was ignored entirely.

A second qualification was needed to deal with the fact that normalized ranks were not available for supervisors. In the case of administrative questions, the standard employed was that the correct answer should be obtained by a greater percentage of supervisors than mechanics. Certainly a supervisor should be more knowledgeable in administrative matters than an individual who is primarily a technician, and an item of this nature, in order to be valid, must reflect this. However, the same reasoning does not hold for technical items. While a supervisor should be technically proficient and able to guide technical activities, it is not reasonable to expect him to surpass the highest level of technician, although it is reasonable to expect him to be as competent as the best technicians. Therefore, an item in which the correct alternative was selected by the highest ranked technicians was considered appropriate for supervisors, provided it met all other standards.

FINAL QUALIFYING EXAMINATIONS

A total of 339 of the original 384 items met criteria for inclusion in final qualifying examinations. From these items two alternative forms were prepared for each of the four positions. In assigning items to tests, the following constraints were imposed:

(1) No item may be assigned to more than two different tests.

(2) No item should appear on both forms of the same test.

(3) An item assigned to two tests (201 items were so assigned) must appear on different forms of the test (e.g., Form 1 for the Apprentice test, Form 2 for the Mechanic test).

(4) The two forms of a particular test should have the same approximate mean and distribution of item difficulties (i.e., percent passing the item).

Item Difficulty

The distribution of item difficulties, that is, percent passing each item, is shown in Figure 4. The mean of individual item job relevance percentages is also shown. On the whole, the two forms for each test evidence a similar pattern. An exact match was difficult to achieve owing to the loss of flexibility which resulted from the requirement that an item appearing on Form 1 of one test be assigned to Form 2 of the other test upon which it appeared. The results of this restriction are most evident in the Senior Mechanic's test where there are sizeable differences in the numbers of items at each of the lower difficulty levels (higher percent passing). However, note that differences at one level are largely counterbalanced by differences in the opposite direction at the next highest level -- indeed, were 20% instead of 10% intervals employed, the distribution for the two forms of the Senior Mechanic's test would look almost the same.

Distribution of Item Difficulties (Percent Passing Item) for Final Qualifying Examinations

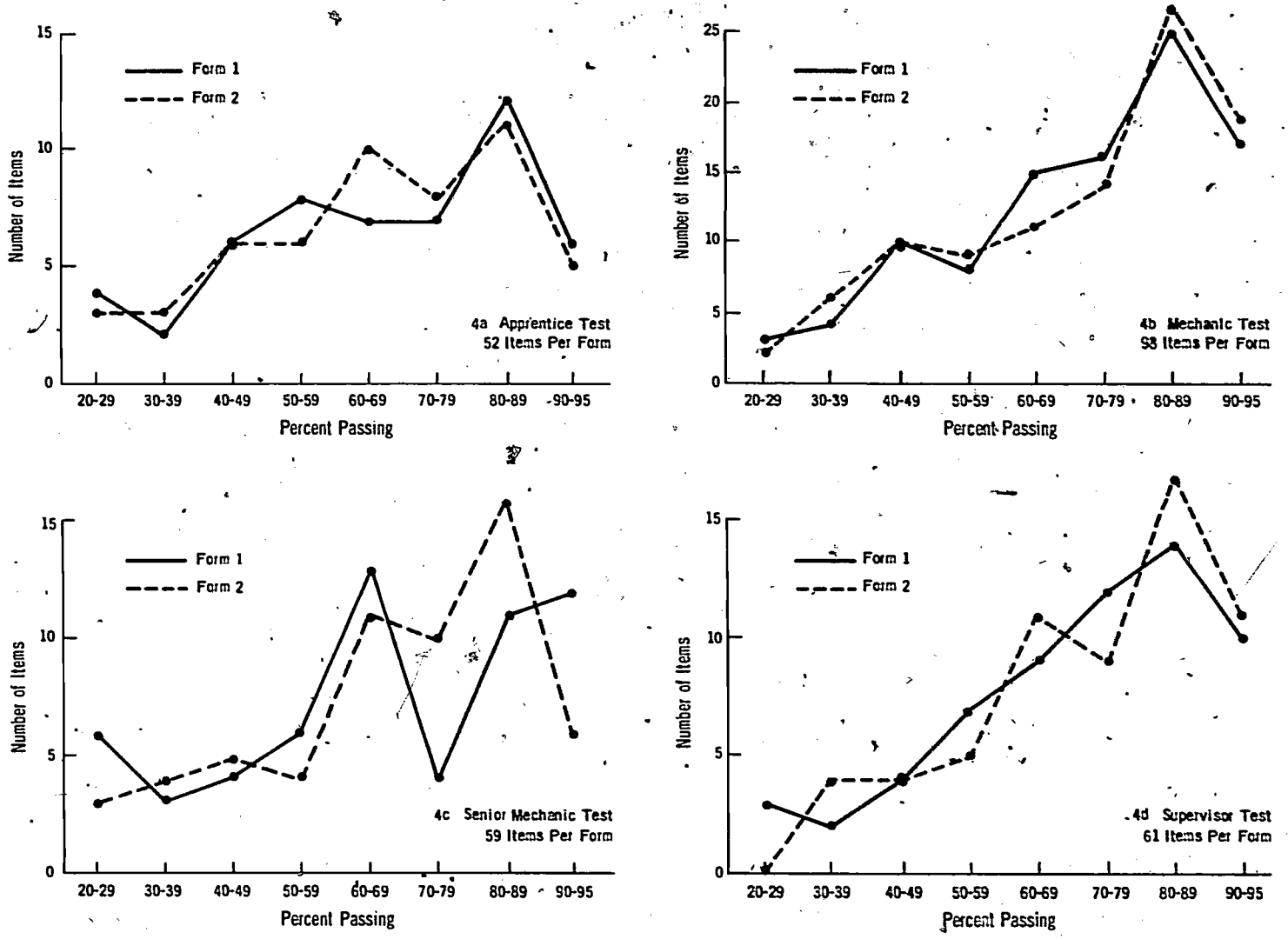

Figure 4

The distributions shown in Figure 4 are not offered as evidence of "real" inter-form equivalence; the matching process involved undoubtedly capitalized on chance similarities that would not appear in subsequent administrations. Each form must be standardized on an independent sample, and equivalence achieved through statistical adjustment. The disproportionate number of low difficulty (high percent passing) items reflects the presence of the "minimum standard" items discussed below.

Test Length

The number of items assigned to each test is indicated in Figure 4. With the exception of the Mechanic test, the examinations are about the same length. While the Apprentice test is a bit shorter than the other two, the examinees (current 4 level Helpers) are somewhat slower test takers and the administration time should be about the same -- one and one-half to two hours.

The fact that the Mechanic examination has almost twice the number of items as the others is a result of the fact that more items were appropriate to this position in content and their tendency to discriminate between current Apprentice and Helper level personnel. However, it is also true that no effort was made to reduce the length of this examination in recognition of the fact that the next lowest position, the PFS-6 Apprentice position, was not covered by an examination. If an examination is deemed necessary as a prerequisite to the 6-level Apprentice position, it should be possible to obtain a sufficient number of job relevant items from the Mechanic test to prepare an additional examination in alternate forms.

Job Relevance

To be included in any examination, an item had to be judged job relevant by more than 50% of those holding positions for which the examination was designed to select personnel. The mean of the actual percentages for all items in a test were as follows:

Test	Mean Percent	Test	Mean Percent
Apprentice		Senior Mechanic	
Form 1	78.3	Form 1	81.3
Form 2	80.1	Form 2	82.0
Total	79.2	Total	81.6
Mechanic		Supervisor	
Form 1	82.4	Form 1	93.9
Form 2	82.0	Form 2	92.3
Total	82.2	Total	93.1

For the three technical positions, items averaged about 80% of examinees judging them job relevant. The supervisors averaged 10% higher. The principal reason for this difference appears to be the difference in scope between maintenance and supervisor jobs. Those performing the maintenance often tend to specialize on a particular

group of equipments; test items that deal with other pieces of equipment, while perhaps job relevant in general, are not relevant to specific individual jobs and therefore were judged "not relevant." The duties of supervisors, on the other hand, are far more similar and it is therefore easier to find items of near universal job relevance. In any case it may be fairly said that the qualification examinations, as a whole, are judged by the great majority of trainees to be relevant to their particular jobs.

Minimum Standard Items

Some portion of the selected items were to be designated as "minimum standard items," that is, items to be counted toward a passing score. The requirements for a minimum standard item were that (a) its content be appropriate to all positions within a particular job, (b) the item be viewed by current jobholders as representing a minimum standard, and (c) a sufficient number of present jobholders have actually passed the item. The second two considerations were quantitative and required setting numerical levels. These levels had to be set sufficiently high to assure that the item did indeed reflect the minimum standard, yet low enough to provide enough items for reliable measurement. The standards ultimately arrived at by representatives of the research staff and the Post Office Department were that (a) at least 75% of current jobholders must view the item as reflecting a minimum standard, and (b) at least 80% of current jobholders should have passed the item. The number of minimum standard items for each test, along with the mean percent of examinees passing the items and the mean percent judging them as a minimum standard are:

	Minimum Standard		
Test	Number of Items	Mean Percent Passing	Mean Percent Judging
Apprentice			
Form 1	12	90	87
Form 2	12	89	86
Mechanic			
Form 1	17	92	80
Form 2	17	91	81
Senior Mechanic			
Form 1	12	91	80
Form 2	12	88	75
Supervisor			
Form 1	17	88	83
Form 2	17	89	84

INTERNAL CONSISTENCY

No attempt has been made to determine the internal consistency of items within the preliminary item pool or the final test. Such internal consistency measures as inter-item, part-whole, or split-half correlation would provide no information concerning the value of individual items or the overall test for the purposes for which the test was intended.

Where a test is designed to measure a variable representing some unitary construct such as "job skill," it is reasonable to expect a high correlation among the items or subtests comprising the test. However, the examination under development was not intended to measure such a variable. Rather, it was designed to sample from a defined population of knowledges and skills. While the inter-correlation of individual items may be of interest, it provides no index of the value of individual items or the overall test. On the contrary, were certain items to be discarded because of low internal consistency, the resulting sample would no longer represent the population of knowledges and skills. While the test might achieve psychometric reliability, it would do so only at a sacrifice of sampling reliability.

Chapter 6

SUGGESTIONS FOR FURTHER DEVELOPMENT

The efforts undertaken in this study represent only a step in the development of an improved system of worker assessment within the postal service. Further activity should be initiated under auspices of the Post Office Department toward (a) validation of the proposed qualifying examinations, (b) development of techniques to expand the assessment of worker qualifications, and (c) coordination of qualification examinations with other personnel activities.

VALIDATION OF QUALIFYING EXAMINATIONS[1]

In preparing the qualifying examinations described in this report, the following steps were taken to enhance the power of each measure to provide a valid assessment of the individual worker's ability to perform the job for which he is a candidate:

(1) A survey of prior research was conducted to identify approaches to the measurement of job proficiency that have proven valid in the past.

(2) The development of test content was based upon a comprehensive and detailed analysis of job tasks along with their associated skills and knowledges.

(3) All items were administered to a group of individuals representative of the population for which the examination is intended and the general relation of items to supervisor ratings and job position was examined.

(4) Only those items that appear to be "valid" in terms of the stated criteria were incorporated in the final qualifying examinations.

---

[1] The application of a developed test to an independent sample for the purpose of obtaining test-criterion correlations is frequently termed "cross-validation." We have avoided use of the term because of the unfortunate implication that a test may be considered as having been validated prior to undertaking this step. We would reserve the term "cross-validation" for application to a new population of a test already validated with respect to some initial population.

In view of the process by which qualification tests were developed, it appears highly likely that they are capable of making valid distinctions among workers in terms of their proficiency for the jobs for which the examinations were intended. However, the degree of relationship or the certainty of its existence cannot be determined without application of the proposed tests to an independent group of workers.

The validation sample cannot and need not be as large as the item analysis sample since the entire test may be expected to be more valid than any individual item. Because of the relatively small numbers of individuals involved, it should be possible to collect co-worker ratings as well as supervisor ratings.

Consideration should be given to a comparison of the proposed examinations with those currently in use. However, such a comparison should be approached with caution as it would require use of a criterion sufficiently reflective of job performance to show the specific job orientation of the proposed examinations. Either a work sample measure, or rating based on controlled observation of job behavior should be used if any improvements in job relevance are to be shown. Until some sort of validation program is conducted, the ability of the proposed examinations to assess individual worker qualifications must remain an unconfirmed hypothesis, regardless of the amount of a priori information that can be advanced in its favor.

SUGGESTED FURTHER DEVELOPMENTS

The examinations developed under the research program described in this text cover only a portion of the worker qualifications that were outlined in Chapter 2. To be specific, they are confined to a sample of the knowledges and of certain of the mental skills involved in maintenance of electromechanical postal equipment. If the postal service is to attain a valid and equitable program of worker assessment, efforts must be undertaken to extend the scope of examinations to cover a greater range of worker qualifications.

Performance Examinations

A program of performance examinations should be instituted with a requirement that job candidates demonstrate their ability to perform some sample of tasks called for in the jobs to which they aspire either as a condition of promotion or before the period of probation has ended. It is believed that a performance test program to be effective should have the following characteristics:

(1) Skill orientation. While a performance examination should call for application of all types of job-related knowledges and skills, it should give emphasis to perceptual, motor, and cognitive skills that are difficult to assess through written tests. These skills are most heavily represented in tasks associated with (a) diagnosis, removal, and repair of equipment breakdowns, (b) alignment, particularly of the more delicate mechanisms of complicated equipment such as the Letter

Sorting Machine, and (c) inspection skills, particularly those associated with the visual and aural detection of unserviceable parts.

(2) Individual administration. Because of the demands which they make upon equipment time, performance examinations are generally administered to one individual at a time. While more costly than group examinations, an individually administered examination permits the content of performance tests to be tailored to the needs of individual postal installations and specific job positions. The role of the Post Office Department in assembly of such tailored performance tests would be the establishment of guidelines covering the selection of tasks, preparation of administrative procedures, and development of procedures for interpretation of results, including the preparation of scoring standards.

(3) Coordination with training. In view of the expense involved in administering performance tests, it is desirable that their application be closely coordinated with training in order that training resources, both personnel and equipment, may be shared. Unfortunately, maintenance training throughout the postal service does not appear to be sufficiently standardized to permit its integration with a formal service-wide examination system at the present time.

(4) Research data. In addition to providing a means of assessing worker qualifications, performance tests provide a criterion for the validation of other types of tests including simulated performance tests and written examinations. Procedures for administering performance tests should be established with a view toward the possible role of performance tests in the validation of other types of tests by Post Office Department personnel.

ASSEMBLED TESTS

Where jobs are highly complex or extremely varied, as in the case of managerial positions, assessment by means of a relatively brief examination is inappropriate; an individual's prior education and experience provides a more reliable index of his potential. The use of "assembled examinations," as the process is called within the postal service, is also of value in dealing with highly specific positions for which preparation of a set of formal examinations is unfeasible.

Because an examination is of an assembled nature does not exempt it from the need to be highly job-related. Such examinations should be based upon the same comprehensive and detailed study of job activities and qualifications as were used in the preparation of the formal examinations described in this report.

APTITUDE TESTS

A fundamental premise underlying the development of qualification tests described in this report was that an aspirant for a job deserves an opportunity to demonstrate his ability for that particular job and that this demonstration should not be confounded with factors related to his ability to learn future jobs. On the other hand, a measure of ability to learn, of the individual's general aptitude for work in a particular area, can be a valuable aid in career planning, provided it is adequately distinguished from the individual's specific job qualification. It is suggested that, in pursuing development of aptitude tests, the Post Office Department switch from the use of specific aptitude measures, for example, Mechanical Aptitude Test, to a single differential aptitude test battery capable of assessing aptitudes for a variety of job areas. A differential aptitude test battery, administered to all postal employees, would aid personnel managers in (a) steering new employees down promising career paths, (b) identifying most likely candidates for supervisory and managerial positions, (c) selecting appropriate personnel for special training and educational programs, and (d) transferring personnel from one craft to another as situations demand.

In the same way that an attempt was made to purge job qualification tests of general aptitude factors, a differential aptitude test should be made relatively free of specific job content. If this is not done, an individual with a smattering of knowledge in a particular area may achieve a higher score than one who is fundamentally a more promising candidate. Each of the military services currently uses some form of differential aptitude test in the classification and assignment of its personnel.

BEHAVIOR RATINGS

The importance of personality factors was emphasized in Chapter 2. While current personality tests do not appear suitable for use in an employment situation -- particularly within the government service -- the use of some form of behavior rating appears both practical and desirable. It is suggested that the Post Office Department prepare rating scales to become a standard part of qualification assessment. It is believed that rating scales will achieve the highest possible validity and utility if these guidelines are followed:

(1) Ratings should be performed by immediate supervisors, section leaders, or other individuals who are closely associated with the individual being rated.

(2) Ratings should identify specific behaviors rather than general traits, that is, how often the individual is tardy rather than "punctuality" or "industriousness."

(3) While many relevant personality factors will be general in nature, an attempt should be made to tailor rating scales to

individual jobs. This requires that rating scales be developed from detailed job descriptive information.

(4) Personality scales should be validated against some overall criteria of job proficiency in the same manner as qualification examinations.

(5) Objective methods should be developed for combining the results of ratings with other assessment factors.

COORDINATION WITH OTHER PERSONNEL ACTIVITIES

Implementation of valid job-oriented qualification examinations may be expected to raise to some extent the level of proficiency represented in the jobs affected. However, full benefit of rigorously established qualification examinations and standards will not be realized unless certain aspects of other, related personnel activities are coordinated with the examination program. Personnel activities to be affected would include:

(1) Recruitment. Published qualification standards, educational prerequisites, experience requirements, and so forth.

(2) Training.

(a) The availability of courses or instructional literature to permit acquisition of knowledges and skills covered by qualification examinations.

(b) Use of qualification examinations in selection of training inputs and the certification and promotion of training outputs.

(c) The use of course-administered tests as qualification examinations.

(d) The application of training resources and data to the preparation of qualification examinations including both training-generated job descriptive information and student performance data.

(3) Job classification. The relation of qualification tests to the determination of appropriate PFS levels, steps in grade, and patterns of career progression.

(4) Work operations. The relation of qualification examinations to work assignments, job performance standards, staffing levels, and other aspects of the way in which work is performed.

The need for closer coordination of qualification examinations with various other personnel activities is but one facet of a larger problem concerning the overall coordination of personnel operations within the postal service. An effort is needed to unite into a well

ordered personnel system all of those activities concerned with assuring the continued availability of qualified personnel. Until such an integrated system is created and descriptions of characteristics disseminated in such a way as will enable individual workers and managers to understand its operation, the potential effectiveness of any improvements to individual components of the system will be compromised at the outset.

LITERATURE CITED

1. Anastasi, Anne. *Psychological Testing*, The Macmillan Company, New York, 1957.

2. Bellows, R. M. *Psychology of Personnel in Business and Industry*, Prentice-Hall, Inc., New York, 1954.

3. Biesheuvel, S. "Personnel Selection," *Annual Review of Psychology*, vol. 16, 1965, pp. 295-324.

4. Bingham, W. B. *Aptitudes and Aptitude Testing*, Harpers, 1937.

5. Brown, G. H., Zaynor, W. C., Bernstein, A. J., and Shoemaker, H. A. *Development and Evaluation of an Improved Field Radio Repair Course*, HumRRO Technical Report 58, September 1959.

6. Chapman, J. C. *Trade Tests*, Henry Holt and Company, Inc., New York, 1921.

7. Cronbach, L. J. *Essentials of Psychological Testing*, Harper and Brothers, New York, 1960.

8. Dreese, M. *An Analysis of the Present Status of Qualifications Analysis in the Military Departments with Some Principles and Recommendations for Future Development*, ONR Report ACR-41, Office of Naval Research, Washington, D.C., May 1959, p. 55.

9. Dunnette, M. D. *Personnel Selection and Placement*, Wadsworth Publishing Company, California, 1966.

10. Fink, C. D., and Hibbits, F. L. *Classification, Career Structure, and Job Analysis of Mail Processing Equipment Maintenance Personnel*, Subtask report, Human Resources Research Office, April 1969.

11. Fisher, R. A., and Yates, F. *Statistical Tables for Biological, Agricultural, and Medical Research*, 4th Edition, Oliver and Boyd, 1953.

12. Foley, J. P., Jr. "The Requirements for Performance Tests for Measuring Training and On-the-Job Achievement," *Proceedings, 7th Annual Military Testing Association Conference*, 6570th Personnel Research Laboratory, USAF, San Antonio, Texas, 25-28 October 1965, pp. 69-77.

13. Ghiselli, E. E. "The Measurement of Occupational Aptitude," *University of California Publications in Psychology*, vol. 8, 1955, pp. 101-216.

14. Hausman, H. J., Begley, J. T., and Parris, H. L. *Selected Measures of Proficiency for B-29 Mechanics: Study No. 1*, HRRL Report No. 7, Human Resources Research Laboratories, Bolling Air Force Base, Washington, D.C., 1 July 1949.

15. Hull, C. L. *Aptitude Testing*, World Book Company, New York, 1928.

16. Johnson, E. G. "Performance of Fact and Situational Type Items in Army MOS Evaluation Tests," *Proceedings, 7th Annual Military Testing Association Conference*, 6570th Personnel Research Laboratory, USAF, San Antonio, Texas, 25-28 October 1965, pp. 45-50.

17. Krech, D., Crutchfield, R. S., and Ballachey, E. L. *Individual in Society*, McGraw-Hill, New York, 1962.

18. Lyons, J. D., and Williams, L. W. *Development and Initial Presentation of an Advanced Maintenance Management Course for the Post Office Department*, Draft Technical Report prepared for the U.S. Post Office Department, Human Resources Research Office, submitted March 26, 1969.

19. McKnight, A. J., Fink, C. D., *et al. Analysis of Postal Equipment Maintenance Positions*, Draft report prepared for the U.S. Post Office Department, Human Resources Research Office, June 1969.

20. Meister, D., and Rabideau, G. F. *Human Factors Evaluation in System Development*, J. Wiley, New York, 1965.

21. Morsh, J. E. *The Development of the Written Evaluation of Mechanic's Proficiency (WEMP) Measure for B-50 Aircraft*, Technical Note No. AFPTRC-TN-56-75, Air Force Personnel and Training Research Center, Lackland AFB, Texas, 1956, pp. 11-13.

22. Ryans, B. G., and Fredericksen, N. "Performance Tests of Educational Achievement," in Lindquist, E. F., *Educational Measurement*, Chapter 12, American Council on Education, Washington, D.C., 1951, pp. 455-494.

23. Skinner, B. F. *The Technology of Teaching*, Appleton-Century-Crofts, New York, 1968.

24. Thorndike, R. L. *Personnel Selection: Test and Measurement Techniques*, John Wiley and Sons, New York, 1962, pp. 41-42.

25. Trattner, M. H. "Comparison of Three Methods for Assembling Aptitude Test Batteries," *Personnel Psychology*, vol. 16, 1963, pp. 221-232.

26. Trexler, R. C., and Butler, P. J. *Task SABER: The Development of a Technical Training System*, Draft Technical Report prepared for the U.S. Post Office Department, Human Resources Research Office, submitted September 1968.

27. U.S. Post Office Department, *Postal Manual*, Superintendent of Documents, Government Printing Office, Washington, D.C. 20402.

28. U.S. Post Office Department, *Maintenance Management, Facilities Handbook*, Series MS-10, TL-1, 7-25-64, U.S. Post Office Department, Washington, D.C. 20260.

29. Wesman, A. G. "Intelligence Testing," *American Psychologist*, vol. 23, No. 4, April 1968, pp. 267-274.