AL-TR-1992-0120

AD-A258 835

ARMSTRONG LABORATORY

JOB PERFORMANCE MEASUREMENT SYSTEM DEVELOPMENT PROCESS

Frances J. Laue
Jerry W. Hedge
Melody L. Wall
Larry A. Pederson
Barbara A. Bentley

Universal Energy Systems, Incorporated
Human Factors Division
4401 Dayton-Xenia Road
Dayton, OH 45432-1894

HUMAN RESOURCES DIRECTORATE
TECHNICAL TRAINING RESEARCH DIVISION
Brooks Air Force Base, TX 78235-5000

October 1992

Final Technical Report for Period January 1987 - October 1990

92-32638

92 12 23 004

AIR FORCE MATERIEL COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5000

NOTICES

This technical report is published as received and has not been edited by the technical editing staff of the Armstrong Laboratory.

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

Mark S. Teachout
MARK S. TEACHOUT
Contract Monitor

HENDRICK W. RUCK, Technical Director
Technical Training Research Division

Rodger D. Ballentine
RODGER D. BALLENTINE, Colonel, USAF
Chief, Technical Training Research Division

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1992	Final - January 1987 – October 1990

4. TITLE AND SUBTITLE

Job Performance Measurement System Development Process

5. FUNDING NUMBERS

C - F41869-86-D-0052
PE - 63227F
PR - 2922
TA - 01
WU - 01

6. AUTHOR(S)

Frances J. Laue
Jerry W. Hedge
Melody L. Wall
Larry A. Pederson
Barbara A. Bentley

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Universal Energy Systems, Incorporated
Human Factors Division
4401 Dayton-Xenia Road
Dayton, OH 45432-1894

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Armstrong Laboratory
Human Resources Directorate
Technical Training Research Division
Brooks Air Force Base, TX 78235-5000

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

AL-TR-1992-0120

11. SUPPLEMENTARY NOTES

Armstrong Laboratory Technical Monitor: Mark S. Teachout, (512) 536-2932; (DSN) 240-2932

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report documents the development process for each component of the Job Performance Measurement System (JPMS) for eight Air Force Specialties (AFSs). Procedures are described in general terms, with discussion included as needed to either explain deviations or highlight specific features. Input for this document was obtained through review of previous technical reports, technical papers, unpublished reports, and other informal documentation. Recommendations for future research and application of the JPMS methodology are discussed.

14. SUBJECT TERMS

Criterion development
Job knowledge tests
Performance measurement
Rating forms
Task analysis
Work samples

15. NUMBER OF PAGES

80

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

CONTENTS

CONTENTS (Continued)

CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Concluded)

LIST OF TABLES

PREFACE

As part of a joint-Service job performance measurement research and development program, the Human Resources Directorate of the Armstrong Laboratory (AL/HR) (referred to in this report by its former name, the Air Force Human Resources Laboratory) developed a new methodology, the Job Performance Measurement System (JPMS). The JPMS development process is the topic of this technical report.

The JPMS has been developed for eight Air Force specialties (AFSs) and data have been collected from first-term airmen. The prototype development involved the Jet Engine Mechanic career field (AFS 426X2). Following data collection with this initial set of JPMS instruments, development of the JPMS for three additional AFSs (AFS 272X0, Air Traffic Control Operator; AFS 328X0, Avionic Communications Specialist; and AFS 492X1, Information Systems Radio Operator) was simultaneously undertaken. A final set of four AFSs (AFS 122X0, Aircrew Life Support Specialist; AFS 324X0, Precision Measurement Equipment Laboratory Specialist; AFS 423X5, Aerospace Ground Equipment Mechanic; and AFS 732X0, Personnel Specialist) were then included to complete the development and administration of eight JPMSs that satisfied the Air Force's commitment to the joint-Service Job Performance Measurement Project.

This report documents general procedures used in JPMS development. An accounting and recording of these procedures is necessary for accurate replication, future research and development, and knowledgeable discussion of the JPMS and its associated performance data. Development of this report was performed under Contract No. F41689-86-D-0052 awarded to UES, Inc (formerly called Universal Energy Systems, Inc.).

Many people contributed to the efforts described in this report -- Air Force and government scientists, contractor researchers, Major Command representatives, active duty specialists, base personnel, and so on. To comprehensively list these individuals here would be impossible, as would the accomplishment of the Air Force Job Performance Measurement Project have been without their involvement. The authors wish to acknowledge and thank these individuals for their contributions. The authors would, however, like to thank Dr. Mark Teachout and Maj Marty Pellum for their guidance in the formulation of this document.

JOB PERFORMANCE MEASUREMENT SYSTEM

DEVELOPMENT PROCESS

SUMMARY

This report documents the development process for each component of the Job Performance Measurement System (JPMS) for eight Air Force Specialties (AFSs). Procedures are described in general terms, with discussion included as needed to either explain deviations or highlight specific features. Input for this document was obtained through review of previous technical reports, technical papers, unpublished reports, and other informal documentation. Recommendations for future research and application of the JPMS methodology are discussed.

I. INTRODUCTION

In July, 1980, the Assistant Secretary of Defense (Manpower, Reserve Affairs, Logistics) directed the Military Services to establish a program of research on enlisted personnel job performance. The focus of the initiative was to determine the feasibility of linking enlistment standards directly to job performance rather than to other intermediate measures such as training performance. Each Service was instructed to develop hands-on performance measures in selected occupational specialties. In addition, each Service was responsible for developing specialized expertise in specific surrogate performance measures. Beyond the initial purpose of providing performance data for the validation of the Armed Services Vocational Aptitude Battery (ASVAB), these measures were also proposed for use in classification, training evaluation, and personnel research. This collective effort of R&D became known as the Joint-Service Job Performance Measurement (JPM)/Enlistment Standards Project.

The Air Force Human Resources Laboratory (AFHRL) responded to the directive by developing a technique called Walk-Through Performance Testing (WTPT).[1] The WTPT is intended to measure performance on tasks which are critical to a job. WTPT expands the range of tasks on which an individual is measured by combining hands-on task performance and Interview Testing to provide a high fidelity measure of job competence. The hands-on component resembles a traditional work sample test and is used to measure technical job proficiency. Interview Testing is used to measure proficiency on tasks that cannot be assessed by the hands-on method due to safety, time, or cost constraints.

In addition to WTPTs, Job Knowledge Tests (JKTs) and rating forms were developed to measure the same job content tested by the WTPT. Interview Testing,

---

[1]The use of the term "WTPT" may refer interchangeably to Walk-Through Performance Testing, the process, or to the Walk-Through Performance Test, the instrument.

surrogates for the more costly and time-consuming hands-on proficiency measures. Finally, questionnaires were developed to assess variables known to affect performance or measurement quality; these variables included job experience, motivation, prior training received, and acceptability. The WTPT, JKT, rating forms, and related measures are referred to collectively as the Job Performance Measurement System (JPMS).

Data were collected from first-term airmen (i.e., 1 - 48 months Total Active Federal Military Service (TAFMS)), in eight Air Force specialties (AFSs). The prototype development involved the Jet Engine Mechanic career field (AFS 426X2[2]). Following data collection with this initial set of JPMS instruments, work on three additional AFSs (AFS 272X0, Air Traffic Control Operator; AFS 328X0, Avionic Communications Specialist; and AFS 492X1, Information Systems Radio Operator) was simultaneously undertaken and the JPMS instruments were developed and administered. A final set of four AFSs (AFS 122X0, Aircrew Life Support Specialist; AFS 324X0, Precision Measurement Equipment Laboratory Specialist; AFS 423X5, Aerospace Ground Equipment Specialist; and AFS 732X0, Personnel Specialist) were then included for development and administration of the array of measures.[3] These eight sets of instruments satisfied the Air Force's commitment to the Joint-Service JPM Project in terms of JPMS development and data collection. A complete account of the data collection procedures for the JPMS can be found in a separate report (Laue, Bentley, Bierstedt, & Molina, 1992).

Purpose

The purpose of this report is to document the procedures that were followed in developing the Air Force JPMS. An accounting and recording of developmental procedures is necessary for accurate replication, future research and development, and knowledgeable discussion of the JPMS and its associated performance database. This report will provide a general overview of the development processes, with individual AFSs discussed as needed to explain deviations and highlight unique procedures. It should be noted that a strict adherence to the underlying philosophy and methodological approach to JPMS development was necessary. However, differences in AFS structure, recency and specificity of occupational survey information, equipment availability, and so on, required procedural flexibility to produce an accurate and reliable JPMS.

---

[2]Air Force enlisted occupational specialties are identified by a title and a corresponding five-digit code. The fourth digit designates a specific skill-level within the specialty and is replaced by an "X" when the code is used in reference to the entire specialty. All specialty codes used in this report were those used to identify subject specialties at the time the research was conducted. Reclassification of specialties in the interim may have resulted in changes to some AFS codes.

[3]JKTs were developed for these four AFSs only; data from this measure are not available for the other four AFSs in this project.

The methodologies described in this report were initially implemented with AFS 426X2; any deviations or significant evolutionary changes are noted.

Figure 1 displays the sequencing of the JPMS developmental efforts. This report begins by discussing steps taken prior to the development of the JPMS. Issues addressed include identification of an AFS for inclusion in the JPM Project, background research and information gathering regarding each AFS, coordination and communication with various Air Force personnel and offices, and selection of tasks to be considered for WTPT development. The chapters that follow detail the developmental processes for each JPMS component (i.e., WTPT, JKT, rating forms, related measures). Although each component is discussed separately, points of interdependence among measures and the development procedures are noted. Key stages of development, common to each measure, include: development of a task/item pool, task/item selection, iterative revision and validation by subject matter experts (SMEs) and researchers, development of instruments, and preliminary data collection in pilot testing and/or pretesting efforts. Conclusions and recommendations for future developmental efforts are discussed in the concluding chapter.

II. PRELIMINARY JPMS DEVELOPMENT

This chapter recounts the steps taken prior to development of all JPMS components. Selection of an AFS, coordination with major commands and functional managers, job domain research, and initial task selection were required for the development of all JPMS components. Therefore, these common processes are addressed before detailing the developmental procedures specific to each component.

Selection of an AFS

Early in the JPM Project, researchers selected AFSs for inclusion in the project based on scientific and practical factors that would permit cost-effective development and data collection of the JPMS. Although all of these factors could not be satisfied for any specialty, they served as guidelines for selection because they affected development efforts, data collection, and/or data analysis for an AFS. The ten factors considered during selection of specialties for JPMS development are discussed below.

1. Population: Specialties were sought which had a population of at least 1000 first-term airmen such that a sample of 300 airmen could be drawn for testing from no more than 15 bases. This would provide an adequate sample size and avoid excessive travel costs during data collection.

2. Current issue of the Occupational Survey Report (OSR): The Occupational Measurement Center (OMC) routinely and systematically surveys career field members to determine current job and task information. This material represents the most thorough data available for individual Air Force career fields and was the prime source for AFS task information. The majority of AFSs are studied approximately every three years, and a recent update of the OSR (i.e., less than

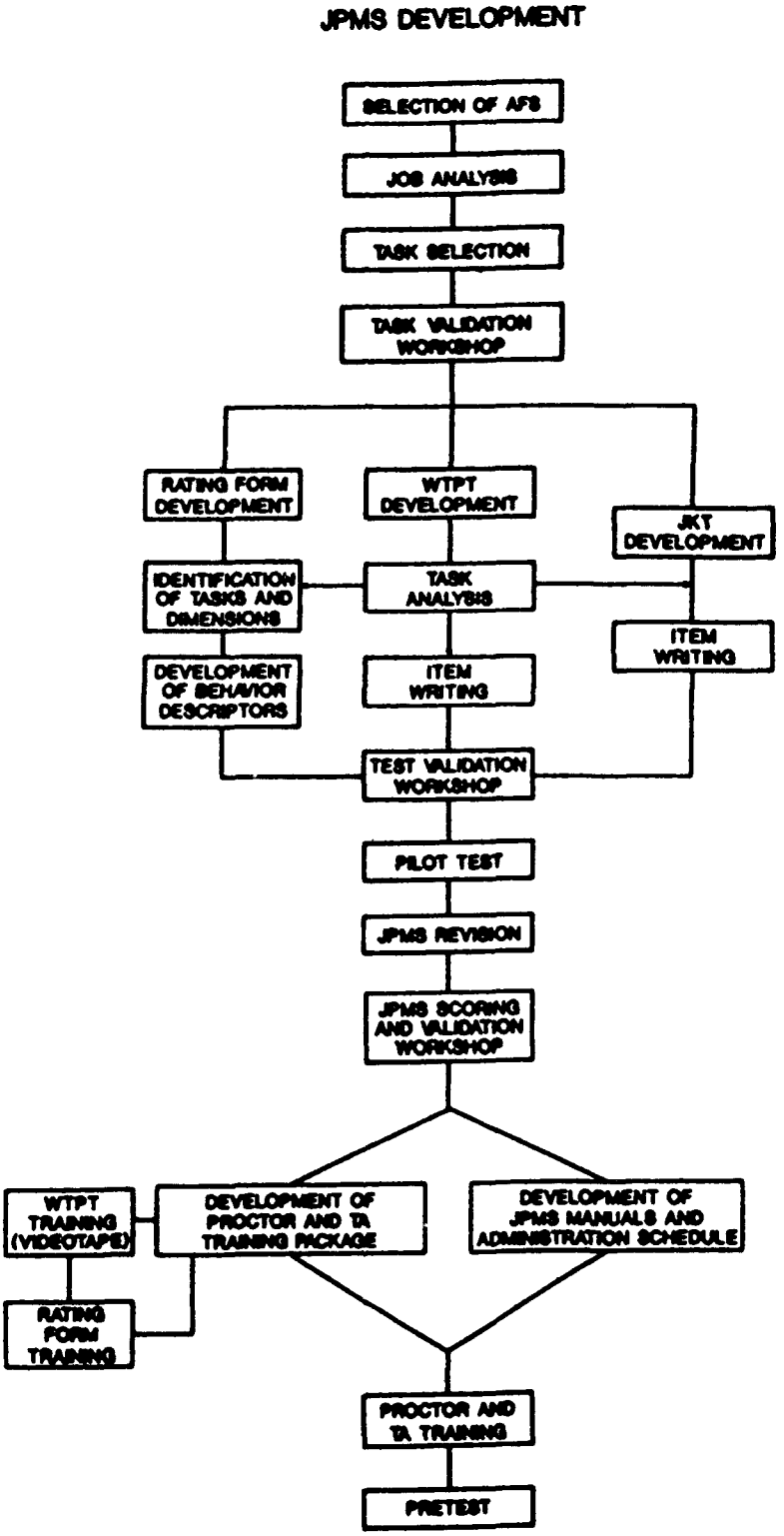

Figure 1. Diagram of JPMS Developmental Stages.

three years old) was desirable to insure that the tasks selected for testing were representative of the current performance domain. Regardless of OSR date, however, a specialty was examined in more detail to determine the accuracy of the most recent OSR. For example, some specialties were found to be relati ely stable over time, while in others, job requirements tended to fluctuate due to technological changes in the specialty. More detail on this topic is addressed in a following section in this chapter titled "Job Domain Research."

3. Armed Services Vocational Aptitude Battery (ASVAB): The selected AFSs were required by joint-Service research guidelines to represent both a cross-section of the four major aptitude index (AI) areas measured by the ASVAB (i.e., Mechanical, Electronic, Administrative, General) and a range of minimum aptitude cut-scores within each AI area. Table 1 lists the eight AFSs in the JPM Project, the appropriate AI, and the minimum cutoff for classification into the specialty.

4. Importance to the Air Force mission: The chosen specialties reflected those areas relatively more critical to the primary mission of the Air Force, that is, the protection of the USA in the air. More tangential areas (e.g., cook, band member) were excluded from consideration, focusing instead on key job areas throughout the Air Force. This ensured that the JPM measures would provide information that would be of value to the Air Force manpower, personnel, and training communities, and applicable to large numbers of enlisted personnel. (See Table 1 for assignment statistics.)

5. Measurability: Critical tasks should be observable and measurable such that JPM development efforts would be successful in producing a set of measures that focus on the key aspects of job performance. Specialties where the critical tasks were deemed neither observable nor measurable were considered poor choices for the JPM Project.

6. Documentation of special concerns: It was important that problems and special concerns (e.g., attrition, safety factors, security classifications) within the AFS were identified and documented. Security requirements were especially important to foresee due to their impact on data collection logistics.

7. Training: It was considered advantageous if the selected specialty was scheduled for an update in training so that task analysis would have immediate utility for the training community (e.g., development of task training plans).

8. Related research: AFSs which were the focus of other ongoing research projects were avoided. Simultaneous research activities could overwhelm a specialty, and potentially impact mission readiness and research efforts. However, if JPMS research could be integrated with other research, the imposition on the specialty would be lessened and was considered advantageous for selection.

9. Diversity of the task pool: Heterogeneity of tasks is represented by a wide variety of job activities, multiple types of equipment required, and various locations of task performance. Jobs that primarily consisted of tasks that reflected a great diversity of skill requirements and job demands were considered to increase the complexity of the project. AFSs consisting of highly

Table 1. Air Force JPMS Specialties and Descriptive Data

Specialty AFS Code	Aptitude Index/ Cutoff	OSR Date	Total Assigned	Percent of Total Air Force
Jet Engine Mechanic AFS 426X2	M/30	1982	9704	1.7%
Information Systems Radio Operator AFS 492X1	A/45	1981	1513	0.3%
Air Traffic Control Operator AFS 272X0	G/43	1980	5386	0.9%
Avionic Communications Specialist AFS 328X0	E/65	1981	1910	0.3%
Aerospace Ground Equipment Mechanic AFS 423X5	M/35 E/30	1983	7276	1.2%
Personnel Specialist AFS 732X0	A/50	1979	7546	1.3%
Aircrew Life Support Specialist AFS 122X0	G/50	1984	2297	0.4%
Precision Measurement Equipment Specialist AFS 324X0	E/65	1984	1995	0.3%

Note. M = Mechanical; A = Administrative; G = General; E = Electronic. Assignment figures are based on yearly statistics from the date of the corresponding OSR publication. ASVAB cutoff level information is dated 30 April 1984 for all AFSs, with the exception of AFS 492X1 (dated 30 April 1986).

diverse tasks were avoided in favor of those with a simpler or more homogeneous structure (e.g., fewer major job duties) since the diversity would impact development, administration, and analysis of the JPMS. However, some diversity was considered appropriate in that it would allow testing of JPMS applicability across a wider variety of career fields.

10. Minorities and women: Whenever possible, AFSs were selected which have minorities and women well-represented in their population. With sufficient sample sizes, the impact of the measurement techniques on these groups could then be evaluated.

Summary

Consideration of each of these factors helped to insure success of the JPMS development. Unfortunately, no specialty met each of these criteria, and the pros and cons of each factor had to be weighed in the final selection of AFSs for inclusion in the JPM Project. The impact of these factors on JPMS development will become more apparent as the development process is fully described.

Once the specialties were identified, the AFHRL began working with the focal career fields in preparation for the initial research and information-gathering required for development of a JPMS. This communication continued through each stage of development and administration.

Coordination with Major Commands

As a starting point in the research of each selected AFS, researchers[4] visited the headquarters of three major commands (MAJCOMs) for informational briefings on the JPM Project.[5] It was important to secure the support and cooperation of key personnel at the MAJCOM level in order to successfully carry out JPMS development and data collection efforts. Thus, it was vital that functional managers and other key personnel at the MAJCOMs were briefed on JPM Project history, goals, and potential payoffs. As part of the briefing, they were informed that their assistance was required to make the project successful.

Functional managers for each AFS at the MAJCOM level were requested to serve as points-of-contact (POCs) for the AFHRL in arranging assistance needed from personnel in the field. These POCs played an important role in coordinating development activities such as base visits for task analysis and travel by active

---

[4]Development of each JPMS was conducted by a contract research scientist under the guidance of an AFHRL scientist. This pairing of contractor and AFHRL researchers worked successfully through the development efforts and insured project continuity when changes in personnel occurred.

[5]The MAJCOMs involved in this stage of the JPMS development were the Strategic Air Command (SAC), Military Airlift Command (MAC), and Tactical Air Command (TAC).

duty personnel to Brooks AFB, TX for workshops at the AFHRL. Functional managers were specifically asked to:

1. Act as a POC for the duration of the development and data collection efforts.

2. Send messages to bases requesting support as required.

3. Arrange for active duty SMEs to participate in the Task Selection Workshop. SMEs were active duty Non-Commissioned Officers (NCOs) with a 5- or 7-skill level, held the rank of Technical Sergeant (E-6) or Master Sergeant (E-7), and had knowledge of current first-term airman job requirements for the specialty.

4. Indicate problems that might be encountered during development, and identify any upcoming changes in the specialty. Types of changes to be considered included equipment/technology innovations, staffing changes, and AFS structure modification.

5. Suggest bases to be visited for task analysis. Locations having large populations of first-term airmen were identified for later consideration as data collection sites, not task analysis. Of particular interest were bases with situations that might prove problematic for task analysis (e.g., undergoing equipment and/or weapon system conversions). Bases scheduled for major inspections (e.g., Operation Readiness Inspections) were also to be avoided.

Summary

Central to the accurate development of job performance measures was the gathering of pertinent data and background on each AFS. Visits to the MAJCOMs provided a good opportunity for researchers to gather information and support materials with regard to AFS structure, staffing issues, terminology, technological changes, and so on. Early communication and information-sharing was essential in establishing good working relationships between the AFHRL and MAJCOMs. This research continued through study of the job domain as described in the next section.

Job Domain Research

Job domain information served as the technical foundation for the development of many of the JPMS instruments. The job domain of interest was defined as *the universe of tasks within an AFS which were commonly performed by first-term airmen* and thorough research was critical for the accurate representation of performance requirements. Identification and description of the job domain provided a basis for structure of the WTPT. The resulting design and content of the WTPT then guided development of the job knowledge tests and rating forms.

Sources of Information

For most AFSs in the JPM Project, there were a number of information sources that were used to identify the job domain. Sources included: occupational survey data reported in the OSR; Air Force Regulation (AFR) 39-1, Airman Classification Regulation; Plan of Instruction (POI); and the Specialty Training Standard (STS). These varied in specificity and content regarding AFS information.

The OSR for a career field provided a variety of information, such as the type of work done in the career field, AFS structure, types of jobs performed by first-term airmen, the equipment used, and the types of facilities where work is performed. The STS is a contract between the technical schools and the Air Training Command (ATC) regarding what is to be taught in the technical schools through the specification of the levels of knowledge or skill attainment to be achieved. The POI details which tasks are taught in the technical school course and serves as a guideline for instructors. The POI describes the training objectives in behavioral terms for each task (e.g., minimum proficiency and knowledge required for task performance). Developers of OSRs typically use the POI as a guideline for survey development. AFR 39-1 provides only a general description of the AFS and outlines job duties and responsibilities.

Of these sources, occupational survey data provided the most detailed and comprehensive information, and were used to define the job domain for JPMS development. These other sources, however, helped to provide a better description of the career field and enhanced the interpretation of OSR data.

Following initial familiarization with the job domain, a visit to the ATC technical school was made by researchers. The purpose of this visit was to obtain the most current copies of training documents (e.g., STS, POI) and to talk with instructors and course developers. Discussions focused on the basic skills taught by the school, regulations and procedural guides used most frequently for task performance, and knowledges, skills, and abilities considered most important for a good performer in the specialty. This visit also provided an opportunity for familiarization with equipment and tools used in the AFS.

The following discusses identification of the structure for a WTPT. This description is used to illustrate how findings from the job domain research were analyzed and applied to provide an appropriate framework for a JPMS instrument.

WTPT Structure

An initial decision on WTPT structure was made by the AFS researcher based on information gathered from the occupational survey and job domain data. Each major component of a WTPT was referred to as a "phase." Some phases were common to all members of an AFS, whereas other phases were written for testing in a specific portion of the career field. The number of WTPT phases necessary to adequately represent an AFS was dependent upon the heterogeneity of the task pool. An AFS with a diverse task pool, with certain tasks performed by unique groups, required a multiple-phase WTPT; a career field described by a common set

of tasks was tested with a one-phase test. Figure 2 depicts a hypothetical job domain and a task sampling approach to represent that domain.

Good indicators of diversity in a specialty were the number of "job clusters" (duty areas) and job types described in the most recent OSR, and the percentage of AFS personnel in each area A relatively large number of duty areas and job types with small percentages of personnel indicated heterogeneity within the specialty task pool. For example, AFS 732X0 was found to be quite diverse, with 18 job clusters and 4 independent job types reflected in the OSR. In addition, most job clusters contained less than 10% of the AFS population. Conversely, AFS 423X5 was fairly homogeneous with seven job clusters, two of which combined contained 75% of specialty personnel.

Of the eight AFSs for which WTPTs were developed, six had two-phase tests, one (AFS 426X2) had a three-phase test, and one (AFS 423X5) had a one-phase WTPT. The three phases are defined as follows:

1. Phase I - Specialty-wide tasks. These are tasks performed by first-term airmen across the AFS and are not unique to any particular subgroup. The Phase I portion of the WTPT was administered to each first-term airman tested in an AFS. The structure of AFS 423X5 allowed all measurement to occur at this level.

2. Phase II - Duty-core tasks. These are tasks which are specific to major duty areas such as a weapon system or workcenter. Phase II of the WTPT was composed of two to five sections, each representing a selected duty area. For example, AFS 426X2 required three Phase II sections for different engine types, J-79, J-57, and TF-33. AFS 328X0 was structured around the types of aircraft across three MAJCOMs and, thus, had three Phase II tests. Each Phase II section of the WTPT was administered only to those airmen assigned to that particular duty area or MAJCOM.

3. Phase III - Site-specific tasks. These are tasks which are uniquely performed by workers in certain job types or functional areas. AFS 426X2 was the only AFS for which it was necessary to develop testing with this level of specificity. For each Phase II test, the Phase III component had two sections representing shop or flightline personnel. Each Phase III section was administered only to those airmen who were assigned to that particular functional area.

Graphic representations of a typical job domain and the WTPT structure necessary for testing are shown in Figure 2. In this example, the career field is depicted as being moderately diverse with three identified duty-core areas (labeled as A, B, and C). The WTPT is structured similarly with a Phase I consisting of tasks common to the career field and three Phase II portions covering the three duty-core areas.

Table 2 displays the types of WTPT structure used for the eight AFSs in the Air Force JPMS. Note that a Phase I section is indicated for each; seven tests contained a series of Phase II tests, and only one WTPT required a Phase III component. The basis for the structuring of the Phase II/III tests is also indicated (e.g., engine type, weapon system, workcenter).

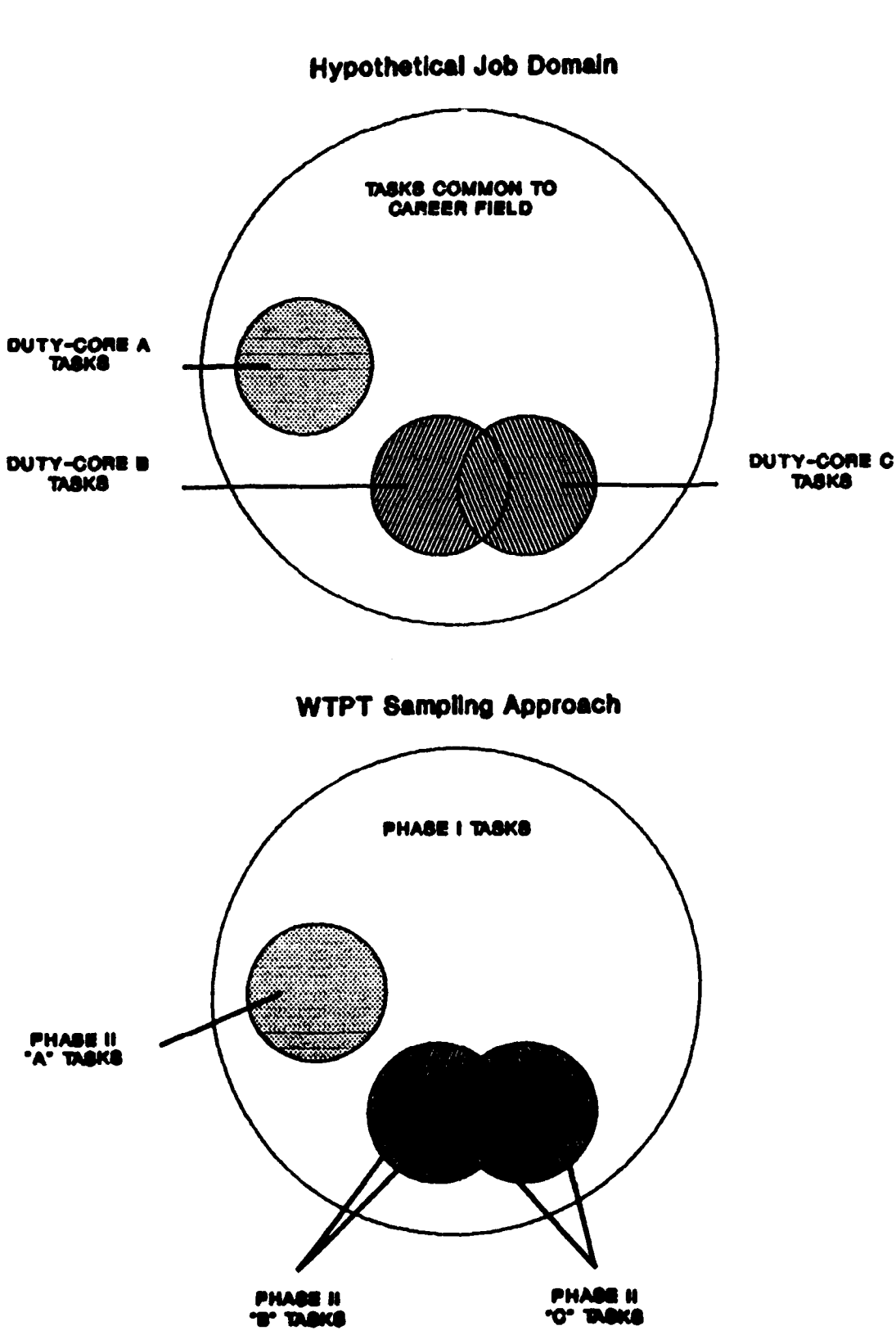

Figure 2. Representations of a Job Domain and the Corresponding WTPT Sampling Approach.

Table 2. Composition of WTPTs

SPECIALTY	PHASES I	II	III
AFS 426X2	X	Engine Type: J-57 J-79	Worksite: Shop Flightline TF-33
AFS 272X0 Radar	X	Worksite/Equipment: Tower	
AFS 328X0	X	Weapon System/MAJCOM: SAC MAC TAC	
AFS 492X1	X	Equipment: CISG GCCS GT	
AFS 122X0	X	Weapon System/MAJCOM: SAC MAC TAC	
AFS 324X0	X	Worksite/Equipment: K1/K2 K3 K4 K5/K6 K8	
AFS 423X5	X		
AFS 732X0	X	Workcenter: Classification & Training Manning Control Outbound Records Separations	

Summary

The hierarchical WTPT design depicting the groupings of job tasks served well to model the structure of various specialties. The utilization of phases provided a common basis for test development across the AFSs and proved to be a flexible method for accommodating the diversity of specialties. The decision process for identifying the appropriate WTPT structure and content is described in the remainder of this chapter.

Initial Task Selection

After becoming familiar with the job domain and identifying a preliminary WTPT structure, the next step was to select tasks to be included in the JPMS. Tasks were selected as potential test items to reflect the preliminary structure established for the WTPT. The task selection process yielded a set of job tasks that were thought to adequately cover the primary duties and demands of a first-term airman. An iterative process, with a selection criterion associated with each phase (i.e., percent performing levels), produced a listing of randomly selected tasks.

AFHRL scientists had identified a goal of approximately 20 to 30 tasks in a WTPT to insure coverage of the job domain while limiting the length of the test to a single work shift. A task sampling plan with this goal was used to generate a set of representative tasks to be considered for JPMS development. This plan, described here in general terms, is graphically displayed in Figure 3. Readers desiring a more detailed explanation of the task sampling plan should refer to Lipscomb (1987). It is important to note that the steps listed below are a generic strategy for task sampling, and minor variations were often necessary to accommodate the particular characteristics of an AFS.

Selection of Specialty-wide Tasks (Phase I Tasks)

1. The task pool for each specialty was comprised of task statements in the Occupational Survey Task Inventory (as reported in the OSR) or tasks included in the POI for initial AFS training. These two sources were considered to cover the universe of tasks either performed by the first-term airman on the job or taught in the technical school. All tasks included in the POI were selected for the Phase I task pool. In addition, tasks which were performed by at least 30% of the first-term airmen, but not in the included in the POI, were selected from the task inventory. This limited the task pool to those tasks deemed important enough to include in technical school training and those that were performed by a substantial number of first-term airmen across the AFS. In this manner, key tasks comprised the task pool, while those tasks either rarely performed or not requiring schoolhouse training were excluded.

2. Selected tasks were aligned according to the appropriate Occupational Inventory Duty Outline area. This step organized the pool of tasks into performance/knowledge clusters. Examples of these duty areas include "Planning and Organizing," "Inspecting and Evaluating," and "Training."

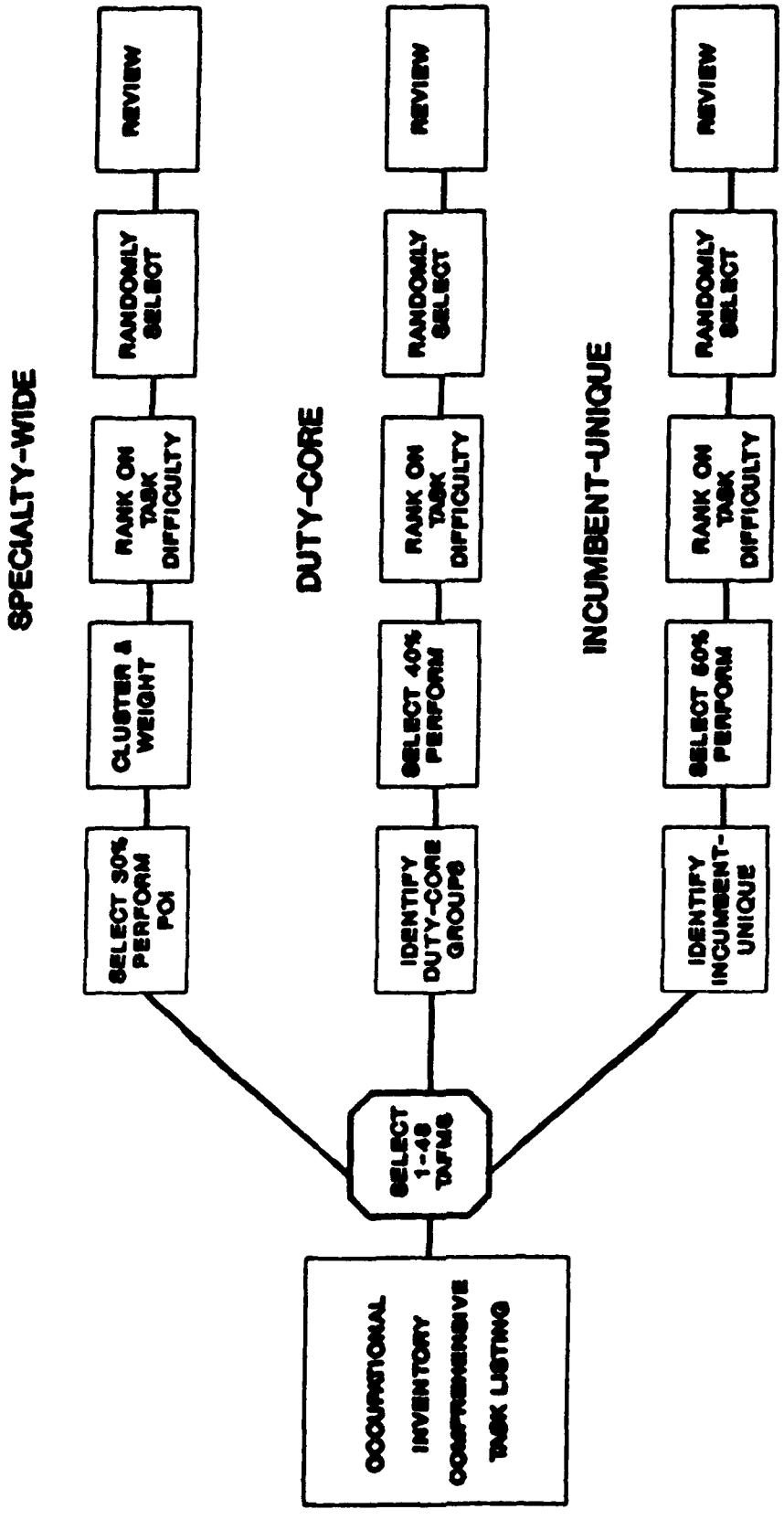

Figure 3. Task Selection Process.

3. Each task cluster was weighted to reflect its relative importance to the overall performance of first-term airmen within the specialty. Weighting was accomplished using the following OSR task factor data: (a) training emphasis ratings for each task and (b) the cumulative relative time spent performing the task. Weights were created by generating the product of the mean training emphasis of tasks within each cluster and the cumulative time spent performing tasks within each cluster. Both of these measures reflect the overall difficulty and importance of tasks and serve to help identify those most critical, or essential, to first-term airman performance. Each cluster of tasks were then ordered in terms of importance according to this weight.

4. The number of tasks to be selected from each cluster for inclusion in the final set of tasks was dictated by the relative importance of the cluster (i.e., cluster weights). Cluster weights were totaled, and each individual cluster weight was divided by the total to obtain a percentage reflecting its relative importance. The predetermined number of Phase I tasks (i.e., 20) was multiplied by each cluster percentage to determine the number of tasks to be randomly selected from that cluster. This process insured appropriate sampling within the task pool across the clusters to produce a task list that reflected the profile of duties of first-term airmen.

5. Within each cluster, tasks were randomly selected to reflect the range of learning/task difficulty. This was done by: (a) ranking the tasks according to task difficulty ratings derived from occupational survey data; (b) dividing the ranked list into quartiles of difficulty; (c) selecting 40% of the total tasks within a cluster from the fourth or most difficult quartile; (d) selecting 30% from the third quartile; (e) selecting 20% from the second quartile; (f) selecting 10% from the first or least difficult quartile; and (g) repeating the process for each cluster.

Tasks not selected during this process, but remaining in the Phase I pool, formed a list of alternate tasks for Phase I. These alternate tasks were used as replacements if tasks initially selected for Phase I were found to be unsuitable by the SMEs or researchers at a Task Selection Workshop. A detailed discussion of the Task Selection Workshop will follow in a later section in this chapter.

Selection of Duty-core Tasks (Phase II Tasks)

AFSs with a complex structure required representation of duty areas not common to the entire career field. As discussed earlier, diversity within the AFS necessitated a WTPT structure consisting of a generic, specialty-wide Phase I test and a series of Phase II tests related to specific duty areas. As shown in Table 2, two to five Phase II duty-core tests were required to represent most job domains. Diversity within the task pool was most commonly a result of equipment or weapon-system differences.

The performance domain for a duty area (e.g., a specific type of equipment or workcenter) was more narrowly defined than for the entire specialty. Fewer tasks were needed in the WTPT for an adequate representation of the duty area, although a higher cutoff of percent performing was mandated by the task sampling

plan. Tasks previously selected during the Phase I process were not eligible for duty-core selection. When duty areas "overlapped" or included the same tasks, tasks could be selected for more than one duty area test. The following steps were followed for selection of tasks for each duty area test.

1. Tasks were selected from among those not used in Phase I and performed by at least 40% of the first-term airmen in the duty area.

2. Tasks were ranked by task difficulty and divided into quartiles. Forty percent of the total tasks in the duty area were selected from the fourth or most difficult quartile, 30% from the third quartile, 20% from the second quartile, and 10% from the first quartile.

As with Phase I, tasks not selected during the Phase II selection process, but in the original pool of Phase II tasks, formed a list of alternates for each duty area.

Selection of Site-specific Tasks (Phase III Tasks)

The two-phase structure was inadequate for representation of the job domain for AFS 426X2. Consequently, a three-phase hierarchy was implemented. In this case, job incumbent[6] location, shop or flightline, required WTPT specialization that included this difference in its structure. The performance domain for each job-type was even more narrowly defined than for individual duty areas, making fewer tasks necessary for an adequate sample. Also, similar to Phase II, selection of tasks to more than one job-type was allowed. Tasks selected for Phase I or II were not used to develop the Phase III tests. The following steps were followed for each job-type.

1. Tasks were selected from the pool of those performed by at least 50% of the first-term airmen in the job-type and had not been identified for inclusion in either Phase I or Phase II.

2. Tasks were ranked by task difficulty and divided into quartiles. Forty percent of the total tasks in the duty area were selected from the fourth quartile, 30% from the third quartile, 20% from the second quartile, and 10% from the first quartile.

3. Tasks not selected during the Phase III selection process, but in the pool of Phase III tasks, formed a list of alternates for each job-type.

---

[6]The term "incumbent" is used in this report to designate those first-term airmen tested on the WTPT. The AFHRL identified a minimum of six months on the job as a requirement for participation as an incumbent in the JPM Project. This amount of job experience, following completion of initial technical training, allowed for some experience on most of the WTPT tasks.

Summary

The end result of task selection was a listing of 20 to 30 randomly selected tasks that covered the major job areas of a first-term airman with a focus on those tasks viewed as most important to job performance. These tasks crossed the range of duty clusters and reflected a range of difficulty. This listing included tasks that are: specialty-wide (Phase I), and, where appropriate, duty-core (Phase II) and site-specific (Phase III) tasks. After assemblage of the lists of primary and alternate tasks, a workshop was held for SME review and validation of each task designated for inclusion in the WTPT.

Task Validation Workshop

The Task Validation Workshop was conducted to judge the selected tasks to insure that the WTPT was a valid representation of the jobs performed by first-term airmen. The proposed structure of the WTPT was also evaluated. Two SMEs from each of three MAJCOMs attended the workshop, held at Brooks AFB, TX. These MAJCOMs (i.e., SAC, MAC, and TAC) were involved during development and data collection for all eight specialties.[7] The MAJCOMs were asked to send SMEs with a broad range of experience across the specialty. Additionally, an ATC representative from the specialty's technical school was invited to this workshop to provide guidance from a specialty-wide training perspective.

A schedule was constructed to enable the goals of the workshop to be met in the four or five days available. The workshop began with a briefing on the JPM Project, workshop goals, and the initial task selection process, with overhead slides serving as visual aids. Workshop attendees were then given handouts with the initial task selection lists arranged by difficulty quartile for each phase. Agenda and other Task Validation Workshop materials are provided in Appendix A.

Task Validation Process

In anticipation of replacement of deleted tasks, a list of alternate tasks for each phase had been prepared and placed in quartiles prior to the workshop. SMEs were not allowed access to the alternate task list since this knowledge may have biased consideration of the original tasks such that potentially suitable tasks from the primary list would be rejected in favor of tasks on the alternate list.

Task evaluation criteria were outlined and discussed with the group. Each task was reviewed individually, and SMEs were allowed to delete a task if there was agreement that:

---

[7]Additional MAJCOMs became involved during the data collection phase for some specialties when the number of airmen in these three commands were found to be insufficient.

1. The task statement was unclear, broad, complex, or trivial;

2. The task was obsolete;

3. The task was not routinely performed by first-term airmen;

4. The task statement was similar to or overlapping with a previously selected task;

5. The task was performed using equipment known to be scheduled for replacement within a year; or

6. There was no observable way to determine successful task completion.

When SMEs rejected a task from the original list, an alternate was randomly selected from the same phase and quartile as the rejected task. If tasks in that quartile were exhausted, the replacement was selected from a more difficult quartile or, finally, if necessary, from the next lower quartile. The SMEs were allowed to move a task from one phase to another if appropriate, keeping in mind that initial task selection criteria had to be considered (e.g., moving a Phase I task to Phase II changed the cutoff percentage of members performing from 30% to 40%). It was emphasized that a legitimate rationale had to exist for moving or deleting tasks. The AFS researcher carefully documented each task deletion, the reason for deletion, each replacement task selected, and the extent to which replacement occurred, along with any other changes to the initial task selection.

Time was set aside at the end of this workshop to identify additional task information that would later aid the task analysis process. SMEs provided information on identification of tasks that could be evaluated using Hands-on Testing and those suitable for Interview Testing, the AFRs used for task performance, and the estimated average time for task performance by a typical first-term airman.

The final product of this workshop, a listing of tasks appropriate for WTPT development, was then used as the basis for task analysis. Several components of the JPMS, including the WTPT, JKT, a task rating form, and a task experience measure, were based on this task listing. At this stage, the AFS researcher had a good understanding of the specialty, a tentative WTPT structure, and a list of tasks appropriate for testing. After identification of this common foundation (i.e., set of job tasks), development of the separate measures began. The following chapters document the procedures followed in the development of the various JPMS components.

III. WALK-THROUGH PERFORMANCE TESTING

WTPTs consist of two types of work sample tests, hands-on and interview. Hands-on Testing, designated as the benchmark or standard, was used to measure job proficiency whenever possible. These test items required airmen to *demonstrate* task performance using the appropriate equipment and technical reference materials. Interview Testing was used whenever time, cost, or safety considerations made actual hands-on performance of a task impractical. These

test items asked the airmen to *explain* how a task is performed without actually performing the task. Interview tasks are performed at the equipment so the airman can make use of environmental cues and reference the equipment in a "show and tell" manner. Interview Testing was also included in the JPMS for evaluation as a potential substitute or surrogate for Hands-on Testing (Hedge, Teachout, & Laue, 1990). Thus, parallel hands-on and interview WTPT items were developed for some tasks. A brief overview of the development process precedes a detailed accounting of each procedural stage. Refer to Figure 1 for a graphic depiction of the developmental process.

WTPT development relied heavily upon SME input at each step of the process. After the Task Validation Workshop, task analysis was performed by interviewing and observing SMEs at several Air Force bases. After test items were drafted, a Test Validation Workshop was conducted for SME review and revision. Information concerning specifics of test item administration and data collection logistics were also discussed with the SMEs. After this workshop, the WTPT was pilot tested in a field setting. A Scoring and Validation Workshop was then convened for additional SME review of test items and input regarding information necessary for test scoring. Finally, test administrators were selected and trained, and the WTPT was pretested in the field prior to full-scale data collection. Each of these steps is fully described in the following sections of this chapter.

Task Analysis

The purpose of task analysis was to gather detailed information on the performance requirements for each task previously validated at the Task Validation Workshop. Task analysis yielded a complete listing of the steps required for completion of each WTPT task. Additionally, each task was identified as a potential hands-on, interview, or overlap (i.e., parallel hands-on and interview versions) test item, and logistical and administration concerns were identified. The form used to record task analysis information is shown in Figure 4.

Methodology

Two basic approaches (i.e., SME workshop and site visitation) were used to gather task analysis information. The centralized workshop was attended by SMEs knowledgeable about task requirements; site visits required the AFS researcher to travel to several work sites to meet with SMEs and view task performance. While the former offered the advantages of reducing time-consuming base visits and allowing group discussion among SMEs, the latter allowed the researcher direct contact with equipment configuration and logistical requirements while observing task performance in a work setting. The centralized workshop was used only for analysis of AFS 426X2 WTPT Phase I tasks, while the site visit approach was used for all eight AFSs in the JPM Project.

TASK ANALYSIS WORKSHEET

TASK # _____ OBJECTIVE _____

PHASE ____ JOB GROUP _____ INTERVIEW __ OVERLAP __ HANDS-ON __

BASE _____ SME _____

OFFICE SYMBOL _____ TELEPHONE (___)_____ AUTOVON _____

ESTIMATED TIME TASK PERFORMED ON ____ _____

TASK DOCUMENTATION _____

CONFIGURATION _____

TOOLS/EQUIPMENT _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

REVIEWED AT: _____ _____ _____

Figure 4. Task Analysis Form.

Task Analysis Workshop participants included the MAJCOM functional manager, technical school representatives, and SMEs familiar with each job identified during task selection. The centralized workshop was used to generate detailed information about task performance. Its advantage was having SMEs from all concerned job areas present to review Phase I items and insure tasks were performed in a similar manner across job areas. The format of this workshop was the same as that described for the Task Selection Workshop. Site visitation was used for task analysis in all eight specialties. Prior to visiting a base for task analysis, a message was sent from the AFHRL through the MAJCOM to the individual bases, requesting the following:

1. Permission to visit the base for task analysis.

2. Assignment of a base-level POC for the duration of the visit (approximately four days).

3. Access to two or three SMEs at the 5- or 7-skill level for up to four days. Since individual SMEs often have differing knowledges and opinions concerning task performance, working with two or more SMEs simultaneously during task analysis revealed these differences. Resolution and clarification was promoted through discussion among participants, resulting in an accurate and thorough description of task performance.

4. Use of equipment, a conference room, technical orders (TOs), and procedural guides. Technical materials served as background resources for the AFS researcher and were used to settle disagreements among SMEs.

Analysis Criteria

Regardless of the task analysis method used, similar task analysis procedures were followed. A task analysis checklist guided the analysis of the following areas, ensuring that key issues were addressed.

Suitability for testing. Decisions were required on whether or not to attempt item development on each task. Decision rules, listed below, also guided the choice of work sample approach (hands-on or interview).

1. Are there objective, easily discernible standards for determining correct performance? If not, correct versus incorrect performance of the task probably cannot be readily determined by observation. Therefore, the task may not be measurable using the WTPT methodology.

2. Does a first-term airman complete the entire task alone or require assistance during the task? Tasks requiring helpers to assist the incumbent are difficult to coordinate logistically and standardize for testing, making them better candidates for interview rather than hands-on testing. Tasks requiring a team effort were deemed inappropriate for this study of individual performance.

3. Will the necessary equipment be available for testing purposes? Can alternative equipment be identified for usage? WTPT methodology for both Hands-on and Interview Testing required that the equipment be available for the

incumbent during task performance. Tasks requiring rarely or infrequently used equipment were viewed as poor choices given the logistical requirements for data collection. Use of very expensive or high demand equipment was avoided to reduce interference with the mission. Also, lack of available equipment for a particular task usually indicated that the task was either infrequently performed or outdated due to equipment changes.

4. Does the task discriminate between good and poor performances, or would all incumbents perform similarly? Tasks that do not allow for a range of performance and do not differentiate between good and poor performers are less desirable for testing. However, results of the task selection process often overrode decisions based on discriminability to produce a test based on highly important tasks and maintain high content validity.

5. The type of work sample approach (hands-on or interview) to be used for the task was assessed by considering the following:

a. Is it practical or logistically feasible to administer a hands-on evaluation of the task within the normal work environment?

b. Is there a risk of damage to equipment if the task is performed several times a day for a week or more?

c. Are there personal safety considerations involved in task performance?

d. Are there concerns about the length of testing required for completion of the task?

Test construction needs. Several questions guided the development of the WTPT by focusing on the step-level descriptions of performance. Consideration of these helped to assure these key criteria: representativeness, completeness, and clarity.

1. What TO or standard procedural guide was typically used during performance of the task? Were there general requirements needed for successful task completion that are not mentioned in available technical documentation?

2. Where, in the overall job, does a task begin and end? What is the sequence of steps performed between the beginning and ending of a task? Must the steps be performed in a particular order for successful task completion? Is it possible to omit certain steps and still accomplish the task objective? Consideration of these questions focused on the identification of the key steps within the task and the requirements for successful performance of each step.

3. What steps are frequently performed incorrectly by first-term personnel? These steps would help to discriminate between good and poor performers, and consequently should be evaluated. Steps which were seldom incorrectly performed were less likely to discriminate, making them candidates for omission if the number of steps in the task had to be decreased because of time concerns. This information was also important at a later point during the development of videotapes for use in the test administrator training.

4. If it was impractical to perform an entire task, subtasks were identified for use in a WTPT based on the following criteria:

a. Representativeness of the general task. Are the behavioral elements of the subtask representative of the general task such that, if the incumbent can perform this subtask, he or she can perform the entire task?

b. Discrimination. Does the subtask have the ability to discriminate between good and poor performers?

c. Assessability. Can the subtask be measured by the WTPT?

d. Economy. Can the subtask be evaluated in a relatively short time? If not, the task may be a good candidate for Interview Testing.

e. Frequency. Is the subtask frequently performed by first-term airmen? Infrequent performance would indicate that the task is not a critical component of the job and may not be representative of the job.

Phase considerations. Although a preliminary phase structure was reflected by the initial task lists, the final decision about WTPT structure depended on the results of task analysis. The following questions helped to either verify or modify the proposed phase structure:

1. Is the same equipment used across the specialty?

2. Is the task performed the same way in all functional areas? Significant differences in equipment or technique would indicate that a multiple-phase WTPT was needed to account for diversity within the AFS.

3. Do local operating procedures influence the way in which a task is performed across the specialty? A tasks that varied from site to site was inappropriate for inclusion since no common standard of performance could be established.

Administration Information

Certain details of administration were also addressed through task analysis. SMEs and technical references provided information regarding specifics of task performance which would later be incorporated into WTPT administration procedures. The following logistical requirements were recorded for each task (as shown in Figure 4):

1. Estimated time required to perform the task.

2. Tools and equipment needed.

3. Special equipment configuration required at the beginning of the task.

4. Technical references and other documentation necessary for task completion.

Procedure

Depending on the diversity of tasks within the specialty, four to six task analysis site visits were necessary to gather sufficient information. The visits were scheduled to allow at least one week between each task analysis trip. During the interim period, test items were written for tasks analyzed during the preceding visit. These items were then reviewed by SMEs during subsequent site visits to insure accuracy, clarity, and completeness. This iterative review process was especially critical in the case of Phase I test items to insure their applicability across the specialty (i.e., across MAJCOMs, in different geographic locations).

Summary

Task analysis allowed for in-depth investigation of the steps involved in performance, the conditions surrounding performance, and the equipment requirements for each task. Thorough analysis of each task was necessary to determine if it: (a) was appropriate for inclusion in the WTPT, (b) could be administered in an effective and efficient manner, and (c) contained all information necessary for a fair and objective evaluation of performance. The information gathered through task analysis was then organized and developed into a WTPT item.

Test Development

The development of WTPT items required that the task analysis findings be translated into a standard format, specially tailored to the JPMS. Each item contained all of the information necessary to administer the task and evaluate performance. To guide developmental efforts, a goal of 50% hands-on test items, 25% interview items, and 25% overlap items was attempted. This mix of items would allow a sufficient number of benchmark (i.e., hands-on) and surrogate (i.e., interview) items for later data analysis and research. Figure 5 contains a sample hands-on task and the corresponding interview overlap item.

Item Writing

WTPT items ordinarily were written by the AFS researcher; however, AFS 426X2 items for Phase I were written during an Item Writing Workshop. SMEs were brought together in a setting similar to that of the Task Selection Workshop and were instructed on how to write test items. They were provided with TOs and procedural guides and then constructed test items for these tasks. It was found in other JPMS development efforts that items were produced in a more timely and efficient manner when item writing was the responsibility of the AFS researcher. SME review was essential, however, and was achieved during task analysis site visits and workshops.

Phase I Hands-On Task 233

Objective: To evaluate the incumbent's ability to safety wire system components.

Estimated Time: 10M Start: Finish: Time Req:

Time Limit: 15M #Times Performed: Last Performed:

Tools and Equipment: .032 lockwire, lockwire trainer, lockwire pliers, T.O. 1-1A-8.

Background: A lockwire trainer was fabricated to standardize this task across MAJCOMs.

Configuration: Any existing lockwire should be removed prior to the start of the task. T.O. is available but need not be used by the incumbent.

Instructions to Administrator:

Administer at the interview table utilizing the lockwire trainer.

SAY TO THE INCUMBENT

I WANT YOU TO SAFETY WIRE THE TWO WING NUTS ON THIS BOARD IN ACCORDANCE WITH THE GENERAL LOCKWIRE PROCEDURES CONTAINED IN T.O. 1-1A-8.

Performed or Answered Correctly	Yes	No
Did the incumbent:		
1. Cut a length of lockwire approximately 18 inches long from the spool?	____	____
2. Select the hole in the uppermost wing of the left wing nut?	____	____
3. Feed one end of the safety wire through the hole in the left wing nut and pull approximately halfway through?	____	____
4. Measure the double strand of wire over the top of the left wing nut and under the right wing nut to the hole in the lower wing (tightening direction)?	____	____

Figure 5. Sample WTPT Items (AFS 328X0).

Phase I — Hands-On Task 233

	Performed or Answered Correctly	Yes	No
5.	Apply the pliers to the measured point on the double strand of lockwire and twist at a rate of 8 to 10 turns per inch?	____	____
6.	Feed one end of the untwisted strand of wire through the selected hole in the right wing nut?	____	____
7.	Check the twisted wire for proper tension?	____	____
8.	Apply pliers 1 to 2 inches beyond the right wing nut and twist the double strand of wire?	____	____
9.	Dike off the twisted wire 4 to 6 turns beyond the right wing nut?	____	____
10.	Turn the pigtail into the wing nut so as to eliminate any hazard?	____	____
11.	Test final assembly for proper tension and direction?	____	____

STOP TIME: _____

TURN PAGE FOR RATING SCALE

Figure 5. (Continued).

OVERALL PERFORMANCE

5 Far exceeded the acceptable level of proficiency

4 Somewhat exceeded the acceptable level of proficiency

3 Met the acceptable level of proficiency

2 Somewhat below the acceptable level of proficiency

1 Far below the acceptable level of proficiency

Figure 5. (Continued).

Phase I Interview Task 233

Objective: To evaluate the incumbent's knowledge of procedures required to safety wire system components.

Estimated Time: 5M Start: Finish: Time Req:

Time Limit: 10M #Times Performed: Last Performed:

Tools and Equipment: .032 lockwire, lockwire trainer, lockwire pliers.

Background: A lockwire trainer was fabricated to provide standardization across MAJCOMs.

Configuration: Existing lockwire should be removed from the trainer.

Instructions to Administrator:

Administer at the interview table allowing the incumbent to look at the lockwire trainer.

SAY TO THE INCUMBENT

TELL ME THE STEP BY STEP PROCEDURES YOU WOULD FOLLOW TO SAFETY WIRE THE TWO WING NUTS IN ACCORDANCE WITH THE GENERAL LOCKWIRE PROCEDURES. REMEMBER TO DESCRIBE THIS TASK IN AS MUCH DETAIL AS POSSIBLE.

Performed or Answered Correctly	Yes	No
Did the incumbent say he/she would:		
1. Cut a length of lockwire approximately 18 inches long from the spool?	____	____
2. Select the hole in the uppermost wing of the left wing nut?	____	____
3. Feed one end of the safety wire through the hole in the left wing nut and pull approximately halfway through?	____	____
4. Measure the double strand of wire over the top of the left wing nut and under the right wing nut to the hole in the lower wing (tightening direction)?	____	____

Figure 5. (Continued).

Phase I — Interview Task 233

	Performed or Answered Correctly	Yes	No
5.	Apply the pliers to the measured point on the double strand of lockwire and twist at a rate of 8 to 10 turns per inch?	____	____
6.	Feed one end of the untwisted strand of wire through the selected hole in the right wing nut?	____	____
7.	Check the twisted wire for proper tension?	____	____
8.	Apply pliers 1 to 2 inches beyond the right wing nut and twist the double strand of wire?	____	____
9.	Dike off the twisted wire 4 to 5 turns beyond the right wing nut?	____	____
10.	Turn the pigtail into the wing nut so as to eliminate any hazard?	____	____
11.	Test final assembly for proper tension and direction?	____	____

STOP TIME: _____

TURN PAGE FOR RATING SCALE

Figure 5. (Continued).

OVERALL PERFORMANCE

5 Far exceeded the acceptable level of proficiency

4 Somewhat exceeded the acceptable level of proficiency

3 Met the acceptable level of proficiency

2 Somewhat below the acceptable level of proficiency

1 Far below the acceptable level of proficiency

Figure 5. (Concluded).

Item Content

The examples of WTPT items shown in Figure 5 illustrate the content and format of a WTPT item. This figure also allows comparison of the two parallel testing approaches, hands-on and interview, included in the WTPT. As shown in Figure 5, WTPT items were written to be uniform in format and contained the following basic information:

1. The corresponding OSR task number for reference purposes.

2. The objective of the item, usually the original OSR task statement. A revised, more specific version of the objective was required for many tasks in order to focus on the application of the task within the testing environment, usually with reference to a specific piece of equipment.

3. The estimated time required to perform the task. This was defined as the time necessary for an average first-term airman to complete the task (as estimated by researchers and SMEs).

4. The maximum time allowed to perform the task; that is, the time necessary for the least competent performer to complete the task (as estimated by the AFS researcher and SMEs).

5. A complete list of the tools and equipment required to perform the task, including any references (e.g., TOs, procedural guides).

6. A description of the correct equipment configuration required at the onset of task performance. The test administrator used this information to prepare the equipment in a standard manner for each examination.

7. Background information associated with the task, such as local procedures, which might impact performance.

8. Instructions to the test administrator on where to administer the test item and under what conditions it should be administered (e.g., equipment configuration, TOs required).

9. Instructions to the first-term airman "incumbent" regarding test administration.

10. Steps comprising the task and required for successful task completion. Next to each step were blanks marked "Yes" and "No." The test administrator placed a check mark in the appropriate blank to indicate whether or not the step was successfully performed by the incumbent.

11. The final portion of each WTPT item contained a summary evaluation, the Overall Performance Rating (OPR) scale. Overall performance for each task was rated on a five-point scale, ranging from 5 ("Far exceeded the acceptable level of proficiency") to 1 ("Far below the acceptable level of proficiency"). These ratings were based on an evaluation of percent of steps performed correctly, the criticality of the steps completed or missed, time taken to complete the task, and general technique and safety procedures used by the

incumbent. Factors such as the incumbent's experience, appearance, verbal facility, and performance on other tasks were excluded from this rating.

Summary

The test development effort resulted in a series of highly specific items in a standardized format. These items were written to provide the detail necessary for administering and evaluating task performance. When all items had been drafted, they were reviewed by a panel of SMEs attending an Item Validation Workshop. This review and critique process is described next.

Test Validation Workshop

The primary purpose of this workshop was to validate WTPT items developed for the AFS. The workshop was held following task analysis and initial test item development. The workshop, arranged in a manner similar to the previous workshops, was attended by two SMEs from the focal MAJCOMs. Validation involved verifying the appropriateness of items included in the WTPT and correctness of the item content. A secondary objective of the workshop was to develop behavioral anchors for the Dimensional and Global Rating Forms. (This process is described in later in this report.) Logistics for pilot test administration were also discussed as time allowed. An agenda for Item Validation Workshops is provided in Appendix B.

Workshop Procedures

Each WTPT item was reviewed and revised by the SMEs to ensure that steps were valid and instructions were clear and complete. Items were examined for job domain representativeness and a range of task difficulty. The SMEs also validated time estimates for performance of each task. All SME revisions were documented by the AFS researcher.

Additionally, WTPT items were sequenced for test administration. The first step in this process was to combine Phase I items with items for each Phase II section, thus creating a separate test for each duty area. As previously mentioned, AFS 426X2 had Phase III test items in addition to Phases I and II. There were three Phase II sections (engine types J-79, J-57, and TF-33) and two Phase III sections (work areas in the shop and on the flightline). Therefore, six separate tests were created:

1. Phase I, Phase II - J-79, Phase III - Shop
2. Phase I, Phase II - J-79, Phase III - Flightline
3. Phase I, Phase II - J-57, Phase III - Shop
4. Phase I, Phase II - J-57, Phase III - Flightline
5. Phase I, Phase II - TF-33, Phase III - Shop
6. Phase I, Phase II - TF-33, Phase III - Flightline

WTPTs with a two-phase structure were similarly constructed into unique tests consisting of Phase I items and the appropriate Phase II items. The AFS 122X0 WTPT, for example, consisted of three tests:

1. Phase I, Phase II - SAC
2. Phase I, Phase II - MAC
3. Phase I, Phase II - TAC

After forming the separate tests, WTPT items were organized into a testing sequence based on equipment usage so that testing could be accomplished in the most efficient manner. Tasks were clustered according to the focal equipment required to allow all tasks using a specific piece of equipment to be administered in sequence. Although the need to sequence tasks according to equipment is more evident in mechanically-oriented AFSs (e.g., heaters and low pressure air compressors in AFS 423X5; solder station and low frequency test bench in AFS 328X0), the WTPT for AFS 732X0 (Personnel Specialist) was also organized around equipment. In this case, tasks requiring a typewriter or computer terminal were clustered to facilitate administration logistics. Tasks concerned with forms completion, common to all AFSs, were sequenced for administration away from the work area.

Pilot test planning was the final topic of workshop discussion. SMEs were asked to recommend possible pilot test locations with consideration of a variety of factors (e.g., proximity to Brooks AFB to minimize travel costs, representativeness of the base to the specialty, base population). In this stage of data collection, a fairly small population of first-term airmen was desirable in order to reserve bases with large populations for data collection. It was seen as desirable to avoid repeated testing at a single site to prevent test security concerns, test-retest confounds, and straining of resources at the site. SMEs were also asked for suggestions concerning data collection procedures.

Summary

The test validation process yielded a revised set of test items approved by SMEs and researchers. The WTPT had begun to take shape with the test structure and sequencing of items identified. Following initial development of the WTPT, the instrument was taken out into the operational environment for testing.

Pilot Test

The purpose of pilot test was to verify the validity of WTPT instruments and to reveal potential administration problems prior to data collection. Ideally, the WTPT was administered to a small sample of five to ten airmen, with each test item administered at least once. Active duty NCOs at the pilot test site were trained as WTPT administrators.

Procedures, directions, task steps, and time limits for each task were carefully examined, and different testing schedules were evaluated for efficiency and smoothness of transition between tasks. Most of the jobs represented by a

multi-phase WTPT were not located at a single base and required more than one site for pilot testing of each duty area.

During pilot test, rating forms and related questionnaires were administered to a small sample of the three groups rating each WTPT incumbent (i.e., self, peer, and supervisor). Raters were briefly instructed by the AFS researcher on how to complete the forms, and each rater was asked to review the forms for appropriateness and clarity.

Testing Procedures

Prior to pilot test, messages were sent through the MAJCOMs to the selected locations requesting:

1. Permission to visit bases for pilot test.

2. First-term airmen to serve as WTPT incumbents.

3. SMEs (5- or 7-level NCOs) to serve as WTPT administrators for the duration of pilot test.

4. Use of equipment and TOs/procedural guides for testing, and a conference room for training and debriefing SMEs.

5. Assignment of a base POC, usually a senior NCO in the AFS.

Upon arrival at the pilot test base, decisions on testing schedules and participants, previously determined by phone discussions with the POC, were finalized. The AFS researcher briefed key personnel on the JPM Project, JPMS development for the specialty, and pilot test goals. SME test administrators were provided with abbreviated training which focused on standardized WTPT administration procedures and WTPT item familiarization.[8] Each test item was reviewed by the SMEs prior to administration to insure familiarity with instructions, task steps, and necessary tools and equipment.

Each WTPT administration was carefully observed by researchers to assess individual item validity and administration logistics. When testing was complete, the SMEs were asked to provide a written critique of the WTPT. The critique focused on the following issues:

1. Does the WTPT provide overall coverage of the job domain?

2. Does the test allow for discrimination between good and poor performers?

---

[8]SMEs were familiarized with WTPT administration procedures using the WTPT Administrator's Manual. Role-playing was also used during pilot test training for practice of administration techniques. Time limitations prevented an in-depth training period.

3. Do the test items capture the essential elements of the tasks?

4. Are interview items adequate substitutes for the hands-on items?

5. Is the item wording clear and easily understood?

6. Which testing schedule do you think is best (if more than one schedule was pilot tested)?

7. Are there any test items you would delete because most first-term airmen do not perform the task?

8. Should additional instructions be included for the test administrator?

9. Are equipment configuration guidelines typical of what is encountered in the field?

10. What additional comments do you have to improve testing procedures?

Responses to these questions were used by researchers to revise items, reorganize test sequence, and refine administration guidelines.

Summary

The pilot test of the WTPT provided researchers with first-hand knowledge of the feasibility of administration of test items. It also allowed another review of test content and format by AFS personnel. After pilot test, WTPT items and administration procedures were revised as needed. The instruments were then in nearly final form, allowing for the development of scoring procedures which were identified in a final SME workshop.

Scoring and Validation Workshop

The Scoring and Validation Workshop was conducted to: (a) provide for a final SME review of WTPT items, rating forms, and related questionnaires prior to pretest; and (b) obtain SME ratings of the relative importance and criticality of each task and its steps. These SME ratings were later used as a component of WTPT scoring procedures conducted by the AFHRL. The workshop was conducted in a manner similar to the Test Validation Workshop, with two SMEs requested to participate from each of the three MAJCOMs. The Scoring and Validation Workshop agenda used for eight AFSs are provided in Appendix C.

Procedure

Each WTPT item was reviewed in detail by the SMEs to be certain that every aspect was clear, accurate, and complete. Test instruments were revised, if necessary. Rating forms and related questionnaires were reviewed in the same manner.

When the SMEs were satisfied that each WTPT item was in its best possible form, they were asked to make criticality and importance judgments of each step. First, SMEs judged each item individually and made criticality assessments for each task step. A "critical" step was defined as one which, if performed incorrectly, would prevent successful task completion. For example, to successfully perform the task "starting a car," one must place the key in the ignition. Therefore, this would be considered a critical step of the task. Criticality ratings were dichotomous (i.e., "Critical" or "Non-critical") and were determined by group vote.

After identification of critical steps in a task, the SMEs reviewed all steps and assigned each an importance rating. The importance rating was defined as "relative importance to overall task performance" on a scale from 1 ("Not Important") to 9 ("Extremely Important") (see Figure 6). By definition, critical steps were rated as "Extremely Important."

---

1 - Not important
2
3 - Somewhat important
4
5 - Moderately important
6
7 - Very important
8
9 - Extremely important

Figure 6. Step Importance Scale.

Each task was also given an overall importance rating relative to other WTPT tasks for the specialty using the same scale shown above. Importance ratings were determined by having each SME provide a rating for each task and then computing the mean of all SME ratings.

Pretest logistics and locations were also discussed at this workshop. Pretest bases which had a large population of first-term airmen representative of the specialty were preferred.

Summary

Researchers carefully recorded all information gathered during this workshop. In the days immediately after the workshop researchers made revisions as necessary. After this information was incorporated into the WTPT, the test items were in final form and ready for pretesting.

Selection and Training of Pretest Administrators

Prior to pretest data collection, it was necessary to select and train test administrators. Technical expertise was the most important qualification for potential test administrators. Familiarity with the JPMS was another consideration.

Test Administrator Qualifications

Technical experience was viewed as a key requirement for WTPT administrators. All test administrators were either active duty SMEs or prior-service contractor personnel. These criteria helped ensure technical expertise and familiarity with the Air Force. AFS 426X2 and AFS 272X0 employed civilian administrators hired by the contractor for pretest. These individuals were recently separated or retired AF personnel with experience in the duty areas covered by the WTPT. The remaining six specialties used active duty SMEs from each AFS as administrators. Active duty personnel were used initially because qualified civilians with technical knowledge of the specialties could not be located and employed as administrators. Later, in the last four AFSs for which data were collected, active duty members of the specialties were used primarily because they were readily available and cost-effective.[9]

Approximately one month prior to the scheduled training, active-duty SME administrators were sought. Personnel with previous experience in the project (e.g., participation in task analysis or workshops) were identified by the AFS researcher and requested by name in messages to the MAJCOMs. Prior association allowed the researcher to assess the interest, motivation, interpersonal skills, and technical ability of SMEs, and request those who would be most competent in the role of test administrator. Previous experience with the JPM Project was also desirable because familiarity with WTPT instruments and procedures would decrease orientation time and simplify the training process. If the personnel requested by name were not available for the training and pretest time period (2 - 3 weeks), the MAJCOMs were asked to substitute other 5- or 7-level SMEs with experience in the duty areas covered by the WTPT.

Test Administrator Training

Active-duty test administrators used for six of the eight AFSs were assembled for training one week prior to pretest. Administrators were trained to administer only the WTPT portion of the JPMS, while a contractor researcher served as "proctor" and briefed base personnel and administered the rating forms

[9]A discussion of the advantages of civilian versus active duty test administrators is provided in a report on JPMS data collections procedures (Laue, Bentley, Bierstedt, & Molina, 1992).

and questionnaires during pretest.[10] WTPT training procedures employed for all specialties is described in general terms. A pre-test training workshop outline is located in Appendix D.

WTPT training focused on test familiarization and practice in evaluating and scoring both hands-on and interview task performance. Training materials consisted of the WTPT items, the WTPT Administrator's Manual, and videotapes of task performance. The Administrator's Manual, which provides general guidelines for WTPT evaluation, was reviewed during training and served as a reference for the duration of pretest. Videotapes were used to give administrators an opportunity to practice observation and scoring of task performance. After administrators reviewed task steps, a videotape of the task was shown. Videotapes displayed both correct and incorrect performance. Task performance was scored, and the scores were discussed step by step to resolve discrepancies among administrators. The overall performance ratings for each task were also discussed to achieve agreement on criteria used to arrive at this judgment.

Interview techniques, such as appropriate probing techniques for interview items and how to open and close a testing session, were demonstrated using a videotaped modeling exercise. Administrators practiced these techniques by role playing interview item administration in pairs, alternating the roles of incumbent and administrator. Finally, logistical requirements and pretest schedules were reviewed.

Pretest

Pretest was designed to be a "dress rehearsal" for full-scale data collection. A small sample of approximately ten first-term airmen were to be tested for each duty area covered by the WTPT. Under conditions closely approximating those of data collection, logistics and administrator training were assessed to determine if revisions were required.

Pretest Procedures

Prior to departure for pretest, all necessary materials were assembled for transport. These materials included:

1. WTPT Manuals consisting of the WTPT items. Each page in the test manual was enclosed in a document protector. Non-permanent transparency markers were used to mark the performance evaluation of task steps (indicated by a check mark in the "Yes" or "No" column) and overall performance ratings on the document protectors. After the results of a testing session were transferred to an answer sheet, marks were wiped off and the cycle was repeated for the next incumbent.

---

[10]The civilian test administrators for AFS 426X2 and AFS 272X0 were hired by the contractor several weeks before pretest, allowing more time for their training. Administrators for these two AFSs were trained to brief base personnel and administer the WTPT, rating forms, and related questionnaires.

2. WTPT Administrator's Manual contained background information on testing procedures.

3. Incumbent Manuals containing the objective and instructions for each WTPT item. The incumbent used this manual during testing to read along as information was provided orally by the test administrator. Additional standardized materials (e.g., scenarios, diagrams) needed for task performance were also included.

4. Rating forms and related questionnaires.

5. JKT booklets (in four AFSs only).

6. Three types of computer-scan answer sheets: WTPT, rating forms and related questionnaires, and JKT (in four AFSs only).

7. Pencils for answer sheet completion.

As with pilot test, a message was sent through the MAJCOMs requesting permission to visit the bases selected for pretest. The sample of WTPT incumbents at each base was selected randomly from a list of qualified first-term airmen prior to arrival at the pretest site.

The first day of pretest was reserved for project briefings to base personnel, checks of equipment for availability and configuration, verification of WTPT incumbents and raters, rater training, and completion of rating forms. The remaining days were occupied with JKT and WTPT administration. On the final day on site, base personnel were outbriefed and thanked for their cooperation and support.

After pretest, any necessary revisions to test items and procedures were made. For example, editing of items was occasionally necessary to incorporate new information on equipment configuration or task requirements, or instructions were reworded for clarity. The WTPT was then in final form and ready to be used for data collection. Test administrator training associated with data collection is described in Laue, Bentley, Bierstedt, and Molina (1992).

IV. JOB KNOWLEDGE TEST DEVELOPMENT

Paper-and-pencil objective knowledge tests were developed as potential surrogates or supplements for the WTPT for four specialties (AFSs 122X0, 324X0, 423X5, and 732X0).[11] JKT development began after completion of the Task Validation Workshop and preliminary task analyses. The knowledge tests were developed simultaneously with the WTPTs. In order to make most efficient use of SME time, testing schedules, and other resources, many of the WTPT and JKT

[11]Bentley, Ringenbach, and Augustin (1989) should be referenced for details concerning the replication of Army JKT procedures and evaluation of the transfer of technology effort for three of the specialties (AFSs 122X0, 423X5, and 732X0).

activities were scheduled to coincide. Joint pretesting of the instruments was planned to simulate final data collection procedures.

The content of each JKT was designed to closely correspond to the content of the WTPT for the AFS. Consequently, each JKT was comprised of a series of "task tests," sets of job knowledge test items that covered individual WTPT tasks. Conventional paper-and-pencil items were developed to cover the steps included in each WTPT task. For pairs of WTPT overlap tasks, the hands-on version served as the basis for JKT task test development. This correspondence was intended to maximize the "surrogate" potential of the JKTs. The following documents the general procedures followed in developing JKTs.

Item Development

Development of a JKT required a detailed review of tasks comprising the WTPT. Therefore, WTPT review involved a thorough examination of each step, including referencing WTPT task analysis information and appropriate TOs, AFRs, and Career Development Courses (CDCs). The WTPT and these other documents were important sources of information concerning the performance of tasks included in the WTPT.

Item development was initiated during a workshop attended by SMEs from each area of specialization covered by the WTPT. SMEs reviewed each WTPT task and identified key elements (i.e., steps) within a task having serious repercussions if not performed or if performed incorrectly. In addition, SMEs identified plausible incorrect procedures for performing the steps identified as key elements. These were later used in the development of alternative distractors.

Following this initial information-gathering stage, the AFS researcher first constructed items for all steps identified as key elements within a task. Next, additional task steps were selected for item development until task coverage was achieved. Adequate coverage of each task was determined through SME review.

The AFS researcher was guided by a set of test construction criteria consisting of the following:

1. Items were written to tap knowledge needed to perform a task. Whenever possible the items required the examinee to actually perform some step in a task before identifying the correct alternative.

2. Item stems were usually limited to two lines and were worded so they could be answered without reference to the alternatives.

3. The number of alternatives developed for an item represented the number of plausible alternatives for performing the step. Thus, it was not necessary that every item have the same number of alternatives.

4. Only one alternative was correct. Items which required selecting the best response from a group of correct alternatives were not permitted.

5. Illustrations, sample forms, and reference material were included whenever possible to convey information in an efficient, effective format that reflected the requirements of the job.

6. Items were written in terminology commonly used on the job, avoiding complicated technical terms. The intention was to test job performance-based knowledge, not the examinees' reading level.

7. Inter-item cuing was avoided to insure that questions or alternatives in one item did not facilitate answering other items.

Test Validation and Revision

Test Validation Workshop

Review of the draft items began with an SME workshop. Initially, the SMEs were divided into small review groups to encourage discussion of the items. For each item, SMEs confirmed the correct answer, determined whether the written alternatives were plausible, and generated additional plausible alternatives. SMEs also decided whether the illustrations, sample forms, and reference materials were accurate, and if similar materials should be included in any other items. A final review of each revised item was made in a large group session.

The SME Test Validation Workshop was also used to compare JKT items with WTPT tasks. Comparisons were made to determine whether the tasks were being sufficiently covered by the JKT items or if additional items were needed. When additional items were needed, the SMEs assisted the AFS researcher in constructing new items.

Item Revision

Following this workshop, JKT items were revised by the AFS researcher and additional items were developed based on information gathered from the SMEs. Items were then reviewed by at least two other test developers who were familiar with general item writing guidelines but not necessarily familiar with the selected AFS. This review was aimed at ensuring proper spelling, grammar, readability, and standard formatting. Following this review, the task tests were assembled into JKTs and prepared for administration to AFS members in a pilot test.

Pilot Test

A preliminary field test was conducted for each JKT to gather data from active duty job experts. The focus was on qualitative comments rather than statistical item analyses. The resulting information was used to guide final revision of items and the overall test structure.

Testing Procedures

JKTs were administered to several groups of five or more AFS members meeting the same criterion as the intended test examinee sample (i.e., first-term airmen). This "incumbent pilot test" involved a group administration of the JKT, one task test at a time. Task test completion times were recorded to obtain an estimate of the amount of time required to complete the entire test. After completion of each task test, incumbents were asked to identify any items, illustrations, or particular terms which were difficult to understand. The correct alternative for each item was also identified and confirmed by the group of incumbents.

The pilot test also included administration of the JKT to a group of senior-level SMEs. This "SME pilot test" focused on appropriateness and technical accuracy of the task tests. SMEs also assessed keying of responses, vocabulary, plausibility of incorrect alternatives, appropriateness and clarity of illustrations, and adequacy of task coverage.

Test Preparation and Assemblage

JKT items were revised based on the input received from incumbents and senior-level SMEs during pilot test. Item revision included: (a) further simplification of technical terms, (b) clarification of sentence structure or wording, (c) improvement of distractor plausibility, and (d) deletion of poor or implausible distractors.

Final JKT items were grouped by task and compiled into booklets for pretest administration. Within each task test, items were arranged in the sequence they would be performed on the job. When possible, task tests were arranged in a logical order. For example, AFS 423X5 task tests were grouped by equipment usage. In addition, where possible, task tests were ordered within a booklet from least to most difficult. This ordering prevented the presence of extremely difficult items at the beginning of a test from causing unnecessary anxiety among examinees. A sample task test from is shown in Figure 7.

Task tests were placed into booklets corresponding to the various phases of the WTPT. Each examinee received a Phase I test booklet and appropriate Phase II booklet. Because the WTPT for AFS 423X5 was not divided into phases, the task tests were divided into two booklets based on the time required for completion. Each booklet was planned to take no more than about one hour to complete.

When job knowledge testing required the use of additional reference materials (e.g., TOs, AF forms, data sheets), these were assembled in a separate booklet. This supplemental booklet allowed easy access to the reference materials during testing.

Task Number: 555

Task Title: Prepare AGE for mobility and training exercises

You are preparing an NF-2 light cart for air shipment under mobility conditions.

18. How much fuel should the fuel tank contain?

A. None; the tank should be drained and purged.
B. None; the tank should be drained.
C. No more than 3/4 full.
D. A full tank of fuel.

19. How do you prevent the loss of any loose fittings or hardware?

A. Secure on the unit with tape or cord.
B. Remove and box; store inside unit.
C. There are not loose fittings or hardware on and NF-2 light cart.

20. What should be done to the tires to prepare them for air shipment?

A. Deflate tires by 20% of rated value to allow for expansion at high altitude.
B. Inflate tires by 20% to compensate for altitude.
C. Visually inspect for deflation, weather cracking, and other defects.
D Inspect and gage tires for proper inflation and for serviceability.

21. What are you required to ensure is properly marked on the unit before it is shipped?

A. Weight of unit.
B. Center of balance.
C. Date and time unit is prepared for shipment.

22. What documentation must be shipped with the unit?

A. AFTO form 95
B. AFTO form 244
C. AFTO form 349
D. AFTO form 350

Figure 7. Sample JKT Task Test (AFS 423X5).

Pretest

JKT pretests were conducted in conjunction with WTPT pretests to closely approximate the actual data collection process. The primary objective of pretest data analysis was to gather information for making final revisions to the JKT. A sample of examinees large enough to allow for statistical analyses of the results and representative of the population of interest (i.e., first-term airmen) was desired.

Administration Procedures

Prior to base arrival, JKT administration facilities with adequate lighting, privacy, ventilation, and working space were requested. Only one test administrator was required to test all examinees in a group setting.

Time required to complete the entire JKT was recorded to establish time requirements of JKT administration for data collection. This administration time included a rest period between the first and second booklets. Administration of the paired test booklets was counterbalanced in an effort to control fatigue effects. Optical scan sheets were used for recording responses.

JKT administration preceded WTPT administration in all cases. The WTPT provided a thorough review of the material covered in the JKT and it increased the potential for seriously inflated JKT scores (and inflated WTPT-JKT correlations) if it preceded the JKT. The rationale was that performance more readily elicits knowledge recall than does knowledge recall aid performance.

Data Analysis

For individual JKT items, the percentage of incumbents selecting each alternative was computed to determine item difficulty. Any item with an item difficulty of less than 10% correct (too difficult) or more than 90% (too easy) was deleted from the test unless its inclusion was required to maintain face validity (i.e., it covered an important aspect of the task that incumbents are required to know). Item-task correlations were computed, and an internal consistency estimate (coefficient alpha) was computed for the group of items comprising each task test. These data were used to eliminate weak items and reduce test length due to time constraints. After deletion of items, the JKTs were ready for full-scale data collection.

Summary

Inclusion of the multiple choice JKTs fulfilled one requirement for the Air Force contribution to the JPM Project, that is, the development of a specific surrogates for the hands-on testing. Each JKT was comprised of items parallel in content to the WTPT thus allowing for detailed examination of the two testing formats, written tests and work-sample test.

V. RATING FORMS

Preliminary Decisions

One goal of the JPM Project was to measure job proficiency using different measurement techniques. This research requirement prompted the inclusion of a variety of rating forms into the JPMS. Decisions concerning development of the rating forms were made by researchers during the early phases of the JPM Project and applied to AFS 426X2. As such, the developmental procedures described are those used for this first specialty. Development of rating forms for the seven other specialties followed closely the methodology applied during the development of the AFS 426X2 rating forms.

Rating Forms and Sources

In order to assess the effectiveness of different rating forms, four measures with varying levels of specificity were conceptualized for development. Three of these forms, Task, Dimensional, and Global, were to tap specialty-unique *job proficiency* across a specificity continuum from micro to macro measurement. A fourth instrument, Air Force-wide, was to measure overall *job performance* across all Air Force specialties.

Task Rating Form. The most specific of the four rating forms, the Task Rating Form was conceived of as covering a broad range of task-level job requirements for a particular AFS. The final set of task statements that make up this rating form came directly from tasks selected in the Task Validation Workshop. Consequently, little developmental work was required for this rating form. This rating form contained all final hands-on and interview task statements, plus additional tasks that were eliminated during the WTPT developmental process because of time or logistical constraints. The Task Rating Form reflected the job domain for a particular AFS, and different rating forms were required based on the phases of the WTPT. Each Task Rating Form contained a Phase I set of tasks plus the appropriate Phase II and Phase III sets of tasks, just as the WTPT did.

Dimensional Rating Form. The Dimensional Rating Form provided the second-most specific rating data. Again, supervisors, peers, and incumbents rated the technical proficiency of first-term airmen across important areas of the job. The number of dimensions rated for each AFS ranged from four to nine. Possible dimensions were identified through cluster/factor analysis of tasks in the Occupational Research Data Base (ORDB) performed by first-term airmen. This analysis, based on task co-performance, provided initial information useful for eliciting input from SMEs in preliminary workshops.

Global Rating Form. A desire to assess overall technical proficiency suggested the development of a global technical rating. It was decided that an interpersonal proficiency item should be included to help raters remove impressions of inter-personal proficiency from their technical ratings. Thus, two items were generated to cover the job domain (i.e., technical proficiency and interpersonal proficiency). AFS-specific examples of technical and interpersonal

proficiency were generated during SME workshops to be discussed later in this chapter.

Air Force-wide Rating Form. Finally, it was decided that a rating form should be developed to cover job performance of airmen across all AFSs, and that its primary focus should be general success in the Air Force. In combination, these two ideas suggested a rating form whose primary emphasis would be a number of interpersonal performance factors. A review of existing literature generated a list of approximately 15 Air Force-wide factors that was to be used to initiate discussions in SME workshops.

Rating Sources. Three sources of ratings were included in the research design. Supervisors, peers, and incumbents (i.e., self) would be asked to complete these four rating forms. These sources were included to assess whether: (a) each yielded unique information, an indication that the data from separate raters should be combined; or (b) sources were overlapping, or similar, in their ratings.

Scale Characteristics

Several decisions were made by project personnel concerning characteristics of the rating scales based on the current rating form research literature and purposes of the project. A five-point scale was chosen for use across all rating forms. This format avoided potential biases encountered in the rater/ratee enlisted population as a result of the nine-point Airman Performance Report rating system currently in use.

Second, adjectival anchors were included to clearly define and differentiate the five scalar points. These adjectival anchors were constructed within a competency framework. Thus, raters were required to make distinctions between airmen who meet or fail to meet a specified level of job proficiency.

Finally, behavioral descriptors were included, whenever feasible, to assist raters in making consistent distinctions across ratees and between levels of performance within ratees. As recommended by Borman (1979), a behavioral summary statement approach was selected as the format for these descriptors. Thus, rather than anchoring each scalar point with a single example, multiple behavioral descriptors were included develop a frame-of-reference for each scale value.

Rating Form Development

Given these preliminary decisions about rating form content and structure, the AFS researcher began to generate preliminary stimulus materials to guide and foster SME input. SMEs were the primary architects of the rating forms, both in the establishment of the prototype forms and, later, in the creation of additional AFS-specific forms.

Rating Form Instructions

Instructions were developed for each rating form to clearly explain the purpose of the ratings and procedural requirements for completion of ratings. The AFS researcher generated several versions of a general instruction form that explained the purpose of the JPM Project, the requirements of the rating task, and the confidential nature of their ratings. In addition, specific instructions explained how each rating form was to be used, detailed the components of the form (e.g., the Global form has technical and interpersonal proficiency scales), and emphasized the orientation of the ratings (i.e., proficiency or performance). Different versions of all instructions were varied in terms of organization and detail, and readied for presentation during SME workshops.

Rating Scale Characteristics

The AFS researcher also generated several versions of rating scale anchors that varied in terms of adjectives used for the anchors (e.g., frequency, amount, level) to allow SMEs to judge clarity and preference. In addition, several versions of scale layout (i.e., 1 to 5 versus 5 to 1) were generated as stimulus materials for SME reactions. Finally, examples of preliminary behavioral descriptors for the Dimensional and Global Rating Forms were produced to assist SME generation of AFS-unique descriptors.

Initial Workshops

Two four-hour workshops were held with AFS 426X2 SMEs to gather initial reactions and input concerning rating instructions, and proficiency rating forms (i.e., Task, Dimensional, Global) layout and content. In addition, two four-hour workshops were held with SMEs from a variety of AFSs to gather information for construction of the performance rating form (i.e., Air Force-wide). These proficiency and performance workshops will be discussed separately below.

Proficiency Workshops

At the two proficiency workshops, 5-, 7-, and 9-level AFS 426X2 SMEs discussed proficiency rating form structure and content. Rating form instructions and input for the three rating forms are detailed separately below.

Rating form instructions. After a brief explanation of the purpose of the JPM Project and this workshop, SMEs reviewed three different versions that varied in terms of level of detail. SMEs agreed that a moderate level of detail should be used for both general instructions and rating form instructions. In addition, content and wording changes were discussed, and changes made to improve clarity. The resulting draft general rating instructions included a very brief overview of the JPM Project and explanation of the various rating forms, emphasizing the differences between proficiency and performance. In addition, instructions were to emphasize the anonymity of the ratings. The rating form-unique instructions were to include a description of rating requirements, an example of the rating scale, and (for the Dimensional, Global, and Air Force-wide forms) the items,

dimensions, or factors that were to be rated. Also, these SMEs agreed on the adjectives that most clearly depicted the five performance/proficiency levels on the rating scales, and that the ordering of the numerical scale vertically from five to one was superior to other layouts. Finally, agreement was reached that the adjectival anchors would include a competency orientation, with a minimal competency cutoff to be established between numerical anchors two and three.

Dimensional rating form. Preliminary dimensions generated from the factor analyses were presented, and discussions focused on selecting a set of dimensions that best reflected the work requirements of the specialty. Once agreement among workshop participants was reached, discussions focused on generating behavioral descriptors for each level of the rating form. This work was aided by several examples previously generated by the AFS researcher. Workshop participants decided that behavioral examples across the five levels should reflect differences in proficiency due to: (a) difficulty of the task, (b) amount of supervision required, (c) reliance on TOs, and (d) time required to perform the task.

Global rating form. The focus of discussion was generation of the behavioral descriptors across five levels of proficiency for the two previously identified items. It was decided that technical proficiency descriptors should use the same factors discussed with the Dimensional Rating Form. Factors judged to be relevant for the Interpersonal Proficiency items included cooperation with coworkers, receptiveness to supervision, and job motivation.

Performance Workshops

Eight SMEs attended the first workshop and seven SMEs attended a second workshop. These SMEs were resource managers at the Air Force Military Personnel Center (Assignments Section), and represented a broad cross-section of AFSs, grade, and experience. After a brief presentation about the purpose of the workshop and JPM Project, the 15 potential performance factors identified from the literature were presented to the group. These factors were discussed and the participants agreed on eight performance factors as representing performance requirements common to all enlisted specialties in the Air Force. Factors comprising the Air Force-wide ratings are listed below. Note that one factor is technically-based, while the other seven relate to various interpersonal factors.

1. Technical Knowledge/Skill
2. Initiative/Effort
3. Knowledge of and Adherence to Regulations/Orders
4. Integrity
5. Leadership
6. Military Appearance
7. Self Development
8. Self Control

Each of these factors was then subjected to the same detailed analysis that occurred in the Dimension and Global workshops, with participants generating behavioral descriptors for the five scale anchors.

Rating Form Revisions

Based on information generated in the initial workshops, the AFS researcher developed draft rating forms and instructions for use in subsequent workshops. These workshops were held in conjunction with the Test Validation Workshops and Scoring and Validation Workshops, utilizing SMEs present for JPMS development. SMEs reviewed all rating form materials, making suggestions to clarify and improve content.

Developmental Efforts for Seven Specialties

As noted previously, many of the decisions that guided the format and content of rating forms and instructions were finalized with the AFS 426X2 work just described. Consequently, structure of the forms remained constant across the remaining seven AFSs. All instructions and the Air Force-wide form were used word-for-word in all specialties. An example of an Air Force-wide Rating Form factor is found in Figure 8. In addition, tasks identified in the Task Validation Workshop were inserted into the Task Rating Form, thus finalizing that instrument. An example of this form can be seen in Figure 9. SME input to the Dimensional and Global rating forms occurred for the seven AFSs at the Test Validation and Scoring and Validation Workshops. Potential dimensions had been identified through of clustering of co-performance data in the ORDB, and workshop participants identified representative dimensions and generated behavioral descriptors for each dimension. An example of a dimension from the Dimensional Rating Form can be found in Figure 10. Similarly, behavioral descriptors were generated as needed for the Global Rating Form by workshop participants. Because of the general nature of the Global Rating Form, workshop participants made few changes across AFSs. An example of the technical item of this form can be found in Figure 11.

Rating Form Presentation

The set of four rating forms represented only a portion of the material required for the data collection session in which the ratings were to be made. Prior to completing the forms, each rater received a standard training program designed to explain the JPM effort and instruct the raters on proper execution of the rating forms. Also, a series of three questionnaires was developed to supplement the rating information and survey attitudes concerning the JPM process. The questionnaires included: a Task Experience Questionnaire to obtain data on the frequency of incumbent performance on the tasks contained in the Task Rating Form; a General Background Questionnaire to document general Air Force and job experience and job satisfaction of the rater; and a Rating Form Questionnaire that measured the acceptability of the rating process. These measures will be discussed in Chapter VI.

Performance Factor 3: Knowledge of and Adherence to Regulations/Orders

Displaying knowledge of and adhering to Air Force (AF)/unit rule, regulations, and orders and displaying respect for authority.

Level	Rating	Behavioral Examples
Always exceeds acceptable level of performance	5	Demonstrates an exceptional knowledge and understanding of AF/unit rules and regulations. Follows the spirit as well as the letter of rules and regulations; obeys orders quickly; always reports promptly for duty, formations, appointments, etc.; remains alert while on duty even when it is inconvenient to do so.
Frequently exceeds acceptable level of performance	4	Demonstrates an excellent knowledge and understanding of AF/unit rules and regulations; follows rules and regulations without fail; always obeys orders; can be counted on to be at appointed area on time; displays appropriate respect of authority.
Meets acceptable level of performance	3	Follows AF/unit rules and regulations almost without fail; is knowledgeable of those rules and regulations that concern safety or security; rarely late for duty or formation; never leaves assigned duty section; always obeys orders.
Occasionally meets acceptable level of performance	2	Occasionally may fail to follow AF rules or regulations; occasionally late for duty formations; usually obeys orders but may question them.
Never meets acceptable level of performance	1	Ignores or fails to follow AF/unit rules, regulations, or orders; often displays of performance lack of respect toward superiors; may leave assigned work area.

Figure 8. Example of an Air Force-wide Rating Form (AFS 732X0).

TASK PROFICIENCY

5	Always exceeds acceptable level of proficiency
4	Frequently exceeds acceptable level of proficiency
3	Meets acceptable level of proficiency
2	Occasionally meets acceptable level of proficiency
1	Never meets acceptable level of proficiency

1. Troubleshoots AC/DC analog voltmeters.
2. Troubleshoots AC/DC analog ammeters.
3. Troubleshoots ohmmeters.
4. Aligns AC/DC analog multimeters.
5. Troubleshoots general purpose oscilloscopes.
6. Aligns general purpose oscilloscopes.
7. Calibrates general purpose oscilloscopes.
8. Solders or desolders discrete circuit components or single layer circuit boards using PACE kits.
9. Reconstructs lands, runs, or soldering pads.
10. Replaces electronic equipment pins, connectors, or plugs.
11. Replaces electronic equipment pins, connectors, or plugs.
12. Researches manuals for parts numbers.
13. Researches microfiche documents for parts information.
14. Performs digital integrated circuit analysis.
15. Interprets calibration correction charts for reference and working standards.
16. Troubleshoots electronic counters.
17. Troubleshoots test oscillators.
18. Calibrates distortion analyzers.
19. Aligns electronic counters.

Figure 9. Example of a Task Rating Form (AFS 324X0).

Dimension 1: General AGE Maintenance

This refers to performing tasks using common hand tools, special tools, test equipment, and shop support equipment in the isolation and correction of malfunctions by removing, repairing and replacing components. This includes general maintenance tasks such a lockwire installation, corrosion treatment, and minor structural repair.

Level	Rating	Behavioral Examples
Always exceeds acceptable level of proficiency	5	Accurately completes even complex maintenance assignments such as removing and replacing the engine of a gas turbine compressor without supervision and without errors.
Frequently exceeds acceptable level of proficiency	4	Accurately completes even complex maintenance assignments such as removing and replacing the engine of a gas turbine compressor with minimum supervision and infrequent minor errors.
Meets acceptable level of proficiency	3	Acceptably completes even complex maintenance assignments such as removing and replacing the engine of a gas turbine compressor with some direct supervision and occasional minor errors.
Occasionally meets acceptable level of performance	2	Completes routine maintenance assignments such as building up a bleed air hose with considerable supervision and numerous errors.
Never meets acceptable level of proficiency	1	Unable to complete even routine maintenance assignments such as building up a bleed air hose without constant supervision and assistance while making excessive errors.

Figure 10. Example of a Dimension Rating Form (AFS 423X5).

TECHNICAL PROFICIENCY

This refers to how skilled a person is at performing various tasks on the job, ignoring interpersonal factors (willingness to work, cooperation with others), or situational factors (lack of tools, parts, or equipment).

Level	Rating	Behavioral Examples
Always exceeds acceptable level of proficiency	5	Successfully completes all tasks with minimal supervision. Completes all tasks rapidly, always using proper maintenance procedures.
Frequently exceeds acceptable level of proficiency	4	Successfully completes all simple tasks and most complex tasks with minimal supervision. Completes most tasks rapidly while consistently using proper maintenance procedures.
Meets acceptable level of proficiency	3	Successfully completes most tasks with some supervision. Occasionally requires excessive time to complete complex tasks. Usually uses proper maintenance procedures.
Occasionally meets acceptable level of proficiency	2	Successfully completes most simple tasks with some supervision, but requires constant supervision to successfully complete some complex tasks. Requires excessive time to complete some complex tasks. Occasionally uses improper maintenance procedures.
Never meets acceptable level of proficiency	1	Unable to successfully complete simple tasks without constant supervision. Requires excessive time to complete the most simple tasks. Frequently used poor maintenance procedures.

Figure 11. Example of Global Rating Form (AFS 122X0).

Assemblage of Materials

The format for assembling the various materials underwent evolution through the various phases of the JPM Project. The changes were due to experience gained through data collection efforts as well as differences in instrument design. Initially, training materials for the rater training session were assembled in a separate booklet. These materials included an introductory narrative, an explanation of each rating form, and three practice exercises designed to provide understanding and experience in completing the forms. After the training session, the booklets were collected and a different booklet of rating forms was distributed. The rater made his/her ratings in the booklet which was then retained for data entry. Since three to five booklets were required for each incumbent, the sheer volume and inconvenience of material to be transported and stored became overbearing and an alternate system was designed for the second and third data collection efforts. An optical scan answer sheet was designed that contained sections for each rating form and questionnaire, as well as identification data, that allowed reusable booklets for the entire training/rating process.

Composition of the booklets was similar across the latter seven specialties. They contained a brief introduction to the rater training session, an explanation of the forms, training exercises on the Air Force-wide and Dimensional Rating Forms, a rating error exercise, and a conclusion in the training portion of the booklet. The ordering of forms in the booklets was Global, Dimensional, Task, and Air Force-wide, preceded by an instruction page. The General Background Questionnaire, Rating Form Questionnaire, and Task Experience Questionnaire followed the rating forms. The format proved both utilitarian and convenient and was viewed as superior to the multiple booklet format.

Pretest

The rating forms were included with the other JPMS instruments for field testing prior to the data collection phase of the various AFS studies. The instruments were administered, under full test conditions, to as many incumbents, supervisors, and peers as could be scheduled under the pretest conditions. With the number of incumbents tested ranging from 21 to 41 across the AFSs, the representation was sufficient to reveal any flaws in rating form design. In the case of the Global Rating Form and the Air Force-wide Rating Form, pretest served as a revalidation tool since these forms remained virtually unchanged across all AFSs once they were found acceptable for the AFS 426X2 study. Pretest for the Dimensional and Task Rating Forms was, essentially, the initial validation since neither the workshops nor pilot testing provided sufficient numbers of administrations or full test conditions with which to evaluate the forms. No major deficiencies were uncovered during the pretest, but minor adjustments were made to several of the Dimensional Rating Form behavioral descriptors as well as the rater training.

Summary

The procedures for developing the rating forms worked well across each of the specialties in the JPMS. This careful development of the four forms resulted in a mechanism for systematically collecting performance ratings from three different sources. Measurement varied in specificity from the very narrow (i.e., task level) to the very broad (i.e., global) providing data to be used for the evaluation of the ratings as surrogates for the work-sample testing.

VI. JPMS QUESTIONNAIRES

Additional instruments were designed to gather a wide variety of information from incumbents, supervisors, and peers to provide a more complete picture of the work environment and perceptions of the JPMS. Several questionnaires were included in the JPMS to provide data on background information (General Background Questionnaire), prior experience on specific tasks (Task Experience Questionnaire), and opinions related to the various components of the JPMS (Rating Form Questionnaire and WTPT Questionnaire).

Each questionnaire will be described in the following sections of this chapter. Included in the discussion will be the rationale behind the development, the construction process, the target group for administration, degree of uniqueness required across specialties, and administration procedures.

General Background Questionnaire

The General Background Questionnaire (GBQ) was administered to all raters as part of the rating form administration session. Although many raters used the forms to rate multiple first-term incumbents, each person completed the GBQ only once. Incumbents were instructed to complete this section on the "Self" rating form while others were to complete this questionnaire after completing the ratings on the first set of rating forms that they received.

The GBQ was divided into two sections; the first gathered data on work history and the latter contained questions related to morale, job satisfaction, job constraints, and so on. The actual content of questions was specialty-specific and tailored to the uniqueness aspects of each AFS. The type of background information requested included time in unit, current and previous work assignments, training history, and MAJCOM. The format of the questionnaires included a mix of open-ended and multiple-choice items. The open-ended items (e.g., "Months in current unit") were later hand-entered in the JPMS database; responses to multiple-choice items for the final seven AFSs were recorded on the optical scan answer sheet.

The intent of the background questions was to gather pertinent incumbent information that would help to interpret or explain either the performance on the WTPT or the ratings given. These data were also collected from all raters to identify biographical data that may help describe the characteristics potentially associated with accuracy in rating behavior. If a "good rater" profile could be

established, this information would help guide decisions about the selection and training of raters for both research and work situations.

Sixteen items comprised the set of multiple-choice attitude measures. These items were general in nature and were used for all of the specialties. The content of the items was related to: morale; skill utilization; sufficiency of supplies, technical information, and training; job satisfaction; motivation; andsupervisory support. These concepts are generally considered by theorists to be relevant to the study of job performance and work behavior. They were included in this study to investigate their relationships to other variables such as ASVAB scores, WTPT performance, and opinions about the JPMS. Additionally, these items may tap individual difference variables. Kavanagh et al. (1987) discussed the research evidence on individual traits and rater characteristics as they relate to measurement quality. Many of the variables measured by the GBQ may help to further clarify the impact of individual differences on job performance ratings.

Task Experience Questionnaire

The Task Experience Questionnaire (TEQ) paralleled the content of the Task Rating Form discussed in a previous chapter. In this instance, however, the incumbents were requested to assess "the amount of relevant on-the-job experience you've had on that task, excluding technical school training" for the series of job-related tasks. The scale ranged from "No or almost none" to "A very great amount." The incumbents completed the TEQ during the rating session, after rating themselves on proficiency/performance; all others were instructed to skip this section.

This task information was collected to enhance interpretation of work sample test performance, since experience is thought to be predictive of job performance. This type of data can also be used as a cross-check or a contrast to other types of task experience measures such as the "Last Performed" and "Times Performed" information collected during the WTPT. As mentioned in the discussion of the Task Rating Form, the list of tasks included those from the Phase I WTPT, all appropriate Phase II/Phase III tasks, and other relevant tasks from the task sampling plan. This specificity of tasks required a separate TEQ for each phase of the WTPT.

Rating Form Questionnaire

The Rating Form Questionnaire (RFQ) measures perceptions and attitudes related to the administration and completion of the four rating forms. The prototype JPMS for the AFS 426X2 data collection included a RFQ focused on motivation to complete the ratings, accuracy of ratings, discrimination, and acceptability. These last three dimensions were evaluated for each of the four forms independently on a 5-point, adjectivally anchored rating scale ("Not at all" to "To a very great extent"). Finally, each of the forms was ranked according to these same three dimensions.

The RFQ was expanded for subsequent data collection to more fully cover the concepts mentioned above and to expand the domain of rating behavior. Three major concepts, hypothesized to be related to rater attitudes, comprised the revised RFQ. These concepts were motivation to rate (7 items), trust in the rating process[12] (7 items), and acceptability/usefulness of the ratings (24 items). These acceptability items required the raters to evaluate each rating form on six dimensions (e.g., fairness, ease of use, confidence in ratings). Item responses on the RFQ were made using the five-point scale described above. Raters were also required to rank the four forms with regard to ease of use, discrimination, and acceptability.

The RFQ was administered to all participants following completion of the four rating forms and the GBQ. As with the other questionnaires, raters were instructed to complete this questionnaire only once, with incumbents answering on the "Self" answer sheets, and other raters responding to the RFQ upon completion of their first rating form.

This form was administered in conjunction with the GBQ to collect information hypothesized to represent a user acceptability construct or factors related to acceptability (Hedge et al., 1987). They proposed using these data to develop a method of using acceptability as a criterion for comparing and evaluating rating forms. The concept of motivation to rate, and its impact on the quality of ratings, was addressed by Kavanagh et al. (1987). Data resulting from the RFQ were designed to address this issue which, may in turn, be useful for selecting and/or training raters.

WTPT Questionnaire

A final questionnaire was designed to measure incumbents' attitudes and perceptions of the entire JPM Project following completion of WTPT. Originally titled "General Utility/ Acceptability Questionnaire" for the AFS 426X2 data collection, the form consisted of six items addressing issues such as acceptability, motivation, concerns about the purpose of testing, etc. An open-ended question asked for suggestions on improving the WTPT instructions. A final item required the incumbent to rank order the rating forms, hands-on test, and interview test on their ability to provide accurate and useful information about an individual's performance.

This form was later revised and additional items were added to more completely measure these same concepts. For the second data collection, the form was retitled "WTPT Questionnaire." Twelve items, on a five-point, adjectivally-anchored graphic rating scale, focused on test performance motivation and trust in the testing process. Seven items required the incumbents to evaluate the two WTPT components, Hands-On and Interview, with regard to

---

[12]These trust items were adapted from the Trust in the Appraisal Process Survey [Bernardin, H. J., Orban, J. A., & Carlyle, J. J. (1981). Performance ratings as a function of trust in appraisal, purpose of appraisal, and rater individual differences. Proceedings of the Academy of Management, 311-315.]

acceptability, usefulness, and quality of test instructions. As before, the incumbents were asked for input on improving the instructions for the WTPT. A final question required the incumbent to rank order the rating forms, hands-on test, and interview test on their ability to provide accurate and useful information about an individual's performance. In the data collection requiring job knowledge testing, the JKT was also evaluated on these dimensions, and the form was renamed "JPMS Questionnaire."

Pretest

These four questionnaires were administered with the other instruments during the pretest phase of the JPMS. As these questionnaires are very straightforward and not dependent on specialty-unique factors, it was unlikely that changes would be needed after pretest. It was important, however, to administer them during the pretest to get the best simulation of the full-scale data collection, including accurate estimates of time requirements. Following the pretest, all instruments were finalized in preparation for actual data collection.

VII. SUMMARY AND RECOMMENDATIONS

This report documents the development process for each component of the JPMS: the WTPT, JKT, rating forms, and questionnaires. These procedures were followed for eight AFSs as part of the Joint-Service Job Performance MeasurementProject. While strict adherence to the methodological approaches outlined here was mandated by research requirements, the importance of procedural flexibility became evident during successive developmental efforts. Differences among specialties, such as job structure, recency and specificity of task information, and equipment availability, required flexibility during JPMS development to produce accurate and reliable instruments. Necessary deviations from general developmental procedures have been noted in this document and should be considered in any future efforts of this type.

Repeated application of these developmental procedures resulted in many "lessons learned." The most critical of these are summarized here to provide additional guidance for those wishing to employ and/or modify the methods described in this report.

First, the selection of an AFS is a major point of concern because of the far-reaching impact on the development of JPMS instruments. While the procedures for development were implemented across all eight specialties in the JPM Project to an acceptable level of success, there are certain criteria that seem to be crucial to the overall development process and quality of resulting products. Primary concerns focus on: (a) the structure of the AFS (i.e., the diversity and number of job-types); (b) the availability of current and relevant occupational information (e.g., OSR data); and (c) contributions of technical experts.

The structure of the AFS impacts on the complexity of a WTPT and related measures (i.e., JKT, rating forms). A complex test demands extensive developmental time, travel, and administrative costs (e.g., word processing,

copying, binding), compared to a test with a more simple design. Costs and efforts associated with data collection are also impacted, since multi-phase tests usually require additional travel and a greater number of administrators. Experience has shown that the more complex the AFS, the more complex the corresponding WTPT, and the more costly the development and data collection efforts. It is also possible that the overall quality of the measures and data collection standards could be diluted when resources (e.g., personnel) need to be spread thinly over a cumbersome project.

Part of the research mission of the Air Force's JPM Project was to apply the prescribed development procedures across a wide variety of specialty structures. Eight projects have demonstrated that this can be done. It is suggested, however, that future selection among candidate AFSs focus on preliminarily identifying the structure of the specialty to ensure selection of an homogeneous career field. In this manner, efforts could be spent on development of job performance measures where they are most likely to be highly successful. Key to this success would be a simple AFS structure, ideally requiring a one-phase WTPT.

Additionally, development of measures would be greatly facilitated by the availability of up-to-date occupational data as reported in the OSR. Many of the development projects were negatively impacted by obsolete information contained in the "current" OSR, actually prepared several years prior to JPMS development efforts. Outdated source information necessitates additional time for researchers to obtain identification and verification of the current state of the career field. If the OSR is chosen as the prime source document for future efforts, development should be planned to coincide with the issuing of a new OSR to ensure recency of data.

A final suggestion concerns the involvement of SMEs in the development process. Their contributions to the success of the project should not be underestimated, as they are the central sources of detailed task information and technical expertise. Their inclusion is mandatory at virtually every stage of development for the WTPT, JKT, and rating forms.

It is important to request SMEs with a mix of grade and experience that reflects the work place environment. SMEs of lower grade (i.e., senior airmen, sergeants) often have the best feel for how tasks are currently being performed. The inclusion of these SMEs is most appropriate at the early stages of instrument development (i.e., task selection, task validation) and is vital during task analysis. SMEs from the workcenter supervisor level (i.e., staff sergeants, technical sergeants) can contribute a broader perspective of career field issues. For example, they can identify dimensions of job performance for rating form development and importance factors related to the scoring of the WTPT.

The "ideal" group of SMEs involved at any single stage might include those with prior JPMS experience (e.g., prior attendees of workshops) and those naive about the project. This continual mixing of "old and new blood" would keep the project moving steadily on an appropriate path with input from diverse groups of experts. Establishment of a static panel of experts at the onset of a project would increase the likelihood that the resulting measures would not be representative of the career field. Instead, they would reflect the collective

experiences of the panel members. At the other extreme, turnover with no continuity of technical experts would tend to slow the development process and preclude any "corporate memory" or history related to the project. Previous efforts were greatly facilitated by the continued involvement of SMEs and the influx of ideas from new participants.

The findings related to the development efforts described here and in other JPMS reports (e.g., Bentley et al., 1989; Hedge & Teachout, 1986; Hedge et al., 1990; Lipscomb, 1987) provide strong evidence of the psychometric quality, flexibility, and applicability of the procedures used to create a JPMS. Future research and development of the JPMS or operational implementation of its component measures should use the guidelines, processes, and recommendations detailed in this report to maximize efficiency of the development efforts and enhance the quality of products.

REFERENCES

Bentley, B. A., Ringenbach, K., & Augustin, J. W. (1989, May). *Development of Army job knowledge tests for three Air Force specialties* (AFHRL-TP-88-11, AD-A208 245). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Borman, W. C. (1979). Format and training effects on rating accuracy and rating errors. *Journal of Applied Psychology, 64*, 410-421.

Hedge, J. W., & Teachout, M. S. (1986, Nov.). *Job performance measurement: A systematic program of research and development* (AFHRL-TP-86-37, AD-A174 175). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Hedge, J. W., Teachout, M. S., & Dickinson, T. L. (1987, Aug.). User acceptance as a criterion for choosing performance measures. In M. Secunda (Chair), *Beyond current performance appraisal research: Acceptability as a new paradigm.* Symposium conducted at the annual meeting of the American Psychological Association, New York.

Hedge, J. W., Teachout, M. S., & Laue, F. J. (1990, Nov.). *Interview testing as a work sample measure of job proficiency* (AFHRL-TP-90-61, AD-A228 054). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

Kavanagh, M. J., Borman, W. C., Hedge, J. W., & Gould, R. B. (1987, Sept.). *Job performance measurement in the military: A classification scheme, literature review, and directions for research* (AFHRL-TP-87-15, AD-A164 837). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Laue, F. J., Bentley, B. A., Bierstedt, S. A., & Molina, R. (1992, XXXX.). *Job Performance Measurement System data collection procedures* (AL-TR-1992-XXXX, AD-XXXX XX). Brooks AFB, TX: Armstrong Laboratory, Human Resources Directorate, Technical Training Research Division.

Lipscomb, M. S. (1987, Sept.) A task-level domain sampling strategy: A content valid approach. In J. W. Hedge & M. S. Lipscomb, (Eds.), *Walk-Through Performance Testing: An innovative approach to work sample testing* (AFHRL-TP-87-8, AD-A185 479). Brooks AFB, TX: Training Systems Division, Air Force Human Resources Laboratory.

APPENDIX A: TASK VALIDATION WORKSHOP AGENDA

I. Purpose

- Validate proposed test structure
- Validate tasks
- Identify equipment
- Identify potential problem areas
- Identify bases for task analysis and data collection

II. Sequence of Events

- Introductions and Administrative Announcements
- Project orientation
  - Research objectives and focus
  - Air Force Job Performance Measurement System
    - Walk-Through Performance Testing
    - Rating forms
    - Experience measures
    - Job knowledge testing
- Task Selection/Validation Considerations
  - Review of Task Domain
    - Occupational Survey Report
    - Plan-of-Instruction
  - Task Selection/Validation Plan
    - Selection process
    - Explanation of phases
  - Test Content and Structure
    - Phase I
    - Phase II (as needed)
    - Phase III (as needed)
  - Task Validation
    - Deletion criteria
    - Overview of Phase I tasks
    - Discuss Phase I tasks and alternatives
    - Overview of Phase II task
    - Discuss Phase II tasks and alternatives
    - Identify tasks to be developed as overlap items
  - Discuss Problem Areas
  - Discuss Bases to be Visited
  - Close Workshop

Task Selection Evaluation Checklist

I. Task Statement Evaluation

A. Is the task statement too broad?
B. Is the task statement too specific?
C. Is the task statement too obsolete?
D. Is the task statement vague?

II. Task Performance

A. First-termer Performance
1. Is this task routinely performed by first-termer?
2. Does a first-termer routinely perform this task in its entirety?
If a first-termer does not perform the task in its entirety, what part would be routinely performed?
3. On the average, how often during a week would this task be performed by a first-termer?
4. Is more than one individual involved in the task?
5. After graduation from tech school, on the average, how long would it be before a first-termer could perform the task?

B. Equipment and Tool Issues
1. Are there tools or equipment involved in task performance?
2. Does the equipment vary from location to location? If the equipment varies by location, would this cause differences in how the task is performed? Would the differences be significant?

C. Work Environment
1. Is the task performed differently depending on where it is done?
2. Are these differences significant? In what way?

D. Reliance on Directive Procedures
1. Are there directives that cover the details of task performance?
2. Is there a requirement for these directives to be used in task performance?

E. Command or Local Management Procedures
1. Do command or local management procedures impact on task performance?
2. Is it possible to develop a standardized performance evaluation by not including these local or command procedures?

F. Time-related Factors
   1. How long does it take to complete this task?
   2. If the task is lengthy, is it possible to reduce the length and still capture the essence of performance?

III. Task Evaluation Method

A. Is it feasible to evaluate this task?
   1. Hands-on
   2. Interview
   3. Combination hands-on/interview

B. Are there safety considerations involved in task performance?

C. Is there a risk of damage to equipment if the task is performed several times over a few days?

D. Are there any security classification issues related to this task?

E. Can the task be objectively evaluated?

IV. Overall Evaluation

A. How representative is the task in terms of skills, knowledges, and abilities required for this specialty?

APPENDIX B: TEST VALIDATION WORKSHOP AGENDA

I. Purpose

- Validate test items
- Revise test items as required
- Review Rating Forms
- Discuss testing locations

II. Sequence of Events

- Introductions and administrative announcements
- Overview of JPM Project and workshop goals
- Test structure overview
- Task selection/validation process
- Test item development process
- Test item validation
  - Process
  - Overview of Phase I test items
  - Validation of Phase I test items
  - Overview of Phase II/III test items
  - Validation of Phase II/III test items
- Development of Rating Forms
  - Global Rating Form
    - Review and discussion
  - Dimensional Rating Form
    - Review and discussion
    - Discussion of cluster diagrams
    - Generation of clusters of tasks
  - Task Rating Form
    - Review and discussion
    - Rewriting of task statements
  - Air Force-wide Rating Form
    - Review and discussion
    - Writing of behavioral descriptors
- Logistics discussion
  - Discuss pilot-test and pretest locations
- Close workshop

APPENDIX C: SCORING AND VALIDATION WORKSHOP AGENDA

I. Purpose

- Step Criticality Ratings
- Step Importance Ratings
- Rating Form Development and Review

II. Sequence of Events

- Introductions and administrative announcements
- Overview of workshop goals
- Review of WTPT structure and content
- Criticality ratings
  - Explanation of step criticality scoring process
  - Assignment of step criticality ratings
- Importance ratings
  - Explanation of step importance rating process
  - Assignment of step importance ratings
- Identification of sample interview item
  - Select a task for inclusion in Incumbent's Manual
  - Perform task analysis and list steps for the task
- Review Rating Forms
- Discuss videotaping procedures
- Discuss pretest logistics
- Discuss data collection logistics
- Close workshop

APPENDIX D: WTPT ADMINISTRATOR TRAINING WORKSHOP AGENDA

I. Purpose

- Training of Administrators for Pretest Data Collection

II. Sequence of Events

- Introductions and administrative announcements
- Overview of JPMS
- Materials familiarization
  - Test Administrator's Manual
  - WTPT Manual
  - WTPT answer sheet
  - JPMS Questionnaire
  - Equipment requirements
  - Code sheet information
  - Definition of terms
- Interview techniques (videotape and discussion)
  - Practice interview item
  - Review of interview items and videotapes
  - Role-playing of item administration
- Hands-On administration (videotape and discussion)
  - Review of hands-on items and videotapes
  - Role-playing of item administration
- Answer sheet completion training and exercise
- Administration of JPMS Questionnaire
- Shadow scoring procedures
- Discussion of Overall Performance Ratings
- Discussion of logistical procedures
  - Scheduling and team coordination
  - Procedures for handling of testing materials
  - Quality assurance for answer sheet completion
  - Communication with AFHRL and contractors
  - Related aspects of JPMS pretest data collection
    - Job knowledge testing
    - Rater training
    - Rating forms administration

*U.S. GOVERNMENT PRINTING OFFICE 1992-0-701-052/60047