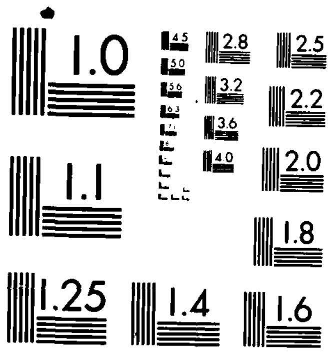
AD-A161 797 BENCHMARK LEARNING DIFFICULTY TECHNOLOGY: FEASIBILITY OF OPERATIONAL IMPLEMENTATION(U) AIR FORCE HUMAN RESOURCES LAB BROOKS AFB TX S K GARCIA ET AL NOV 85 1/1

UNCLASSIFIED AFHRL-TP-85-33 F/G 5/9 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AFHRL-TP-85-33

(11)

AIR FORCE

HUMAN RESOURCES

AD-A161 797

BENCHMARK LEARNING DIFFICULTY TECHNOLOGY: FEASIBILITY OF OPERATIONAL IMPLEMENTATION

Sharon K. Garcia
Hendrick W. Ruck
Joe Weeks

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

November 1985

Interim Paper for Period March 1983 - August 1984

Approved for public release; distribution unlimited.

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

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

This paper has been reviewed and is approved for publication.

NANCY GUINN VITOLA, Technical Director
Manpower and Personnel Division

RONALD L. KERCHNER, Lt Col, USAF
Chief, Manpower and Personnel Division

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

A161797

REPORT DOCUMENTATION PAGE

1a REPORT SECURITY CLASSIFICATION Unclassified	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TP-85-33	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Manpower and Personnel Division 6b OFFICE SYMBOL (If applicable) AFHRL/MOD	7a. NAME OF MONITORING ORGANIZATION
6c ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Brooks Air Force Base, Texas 78235-5601	7b. ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory 8b. OFFICE SYMBOL (If applicable) HQ AFHRL	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10. SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. 62703F PROJECT NO. 7719 TASK NO. 19 WORK UNIT ACCESSION NO. 11

11 TITLE (Include Security Classification)

Benchmark Learning Difficulty Technology: Feasibility of Operational Implementation

12 PERSONAL AUTHOR(S)

Garcia, Sharon K.; Ruck, Hendrick W.; Weeks, Joe

13a TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Interim	FROM Mar 83 TO Nov 84	November 1985	100

16 SUPPLEMENTARY NOTATION

Work Performed Under TS Study Number 8388

17 COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	aptitude requirements; CODAP; benchmark scales; occupational learning difficulty

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

In 1973, the Air Force Manpower and Personnel Center requested that the Air Force Human Resources Laboratory develop an objective procedure for establishing the relative aptitude requirements for enlistee occupations. After 10 years of extensive research, the Laboratory developed a state-of-the-art technology for this purpose. The technology produces measures of occupational learning difficulty that can be used as a job-centered frame of reference for establishing aptitude requirements. These measures were used as an empirical basis for establishing aptitude requirements in the April 1982 revision of Air Force Regulation 39-1, Airman Classification Regulation. ~~Further,~~ occupational learning difficulty is applied within Air Force person-job match algorithms for determining enlistee job assignments. Because occupational learning difficulty will be used for these purposes in the future, it is important that the associated measurement procedure be transferred to an operational organization for routine implementation. To transfer the technology it was first necessary to investigate the feasibility of having Air Force personnel, as opposed to research personnel, collect benchmark ratings for use in deriving the learning difficulty measures. For this (Continued)

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT [X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT [ ] DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION Unclassified
22a NAME OF RESPONSIBLE INDIVIDUAL Nancy A. Perrigo, Chief, STINFO Office	22b. TELEPHONE (Include Area Code) (512) 536-3877 22c OFFICE SYMBOL AFHRL/TSR

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

Item 19 (Concluded):

purpose, eight members of the Air Force Occupational Measurement Center (OMC) collected benchmark ratings for nine enlisted specialties that had been previously evaluated by research personnel. For each specialty, a measure of occupational learning difficulty was derived. Ratings produced by the OMC team were then compared to ratings collected by the research personnel to assess the reliability and validity. Results of the analyses indicated that the reliability and validity of the OMC ratings were equivalent to the reliability and validity of the ratings previously collected by the research personnel. Hence, this study supports the feasibility of having OMC personnel routinely collect benchmark ratings of learning difficulty.

AFHRL Technical Paper 85-33 November 1985

BENCHMARK LEARNING DIFFICULTY TECHNOLOGY: FEASIBILITY OF OPERATIONAL IMPLEMENTATION

Sharon K. Garcia
Hendrick W. Ruck
Joe Weeks

MANPOWER AND PERSONNEL DIVISION
Brooks Air Force Base, Texas 78235-5601

Reviewed and submitted for publication by

R. Bruce Gould
Chief, Force Utilization Branch
Manpower and Personnel Division

This publication is primarily a working paper.
It is published solely to document work performed.

SUMMARY

At the request of the Air Force Manpower and Personnel Center, the Air Force Human Resources Laboratory (AFHRL) conducted research to develop a methodology for establishing aptitude requirements for enlistee occupations. The resulting methodology produces job-centered measures of occupational learning difficulty representing the time required to learn to perform an occupation satisfactorily. These measures were used as an empirical basis for establishing aptitude requirements published in the Airman Classification Regulation. It has also been applied to the Air Force person-job match system for determining enlistee job assignments. In order to transfer this technology from a research to an operational setting, it was necessary to investigate the feasibility of having Air Force personnel, rather than research personnel, collect the data for use in deriving learning difficulty. This paper summarizes a study conducted by AFHRL to assess the feasibility of having Air Force personnel from the Air Force Occupational Measurement Center collect these data to support the operational implementation of the benchmark learning difficulty technology.

PREFACE

This study was completed under Work Unit Number 77191911, "Measurement and Analysis of Job and Mission Requirements," and fulfills the requirements of Training Research Agreement 93, Feasibility Study of the Operational Implementation of Learning Difficulty Technology.

The authors wish to express appreciation to Mr. Dick Rogers, Mrs. Janice Buchhorn, and Mrs. Connie Villareal of the Air Force Human Resources Laboratory (AFHRL) for their computer analysis and programming support. The authors also wish to thank the following Air Force Occupational Measurement Center inventory developers and job analysts who participated in this research effort: Mr. Bob Alton, Mr. Hank DuBois, Mr. Joe Bergmann, Mr. Jim Slovak, Capt Gary Patterson, Capt Robert Worley, 1Lt Carl Middleton, and CMSgt Ted Wilcox.

Special appreciation is extended to Dr. William E. Alley of AFHRL for his technical guidance.

TABLE OF CONTENTS

LIST OF TABLES

BENCHMARK LEARNING DIFFICULTY TECHNOLOGY:
FEASIBILITY OF OPERATIONAL IMPLEMENTATION

I. INTRODUCTION

Occupational research is an important component of sound organizational management. Although the value of such research for training management has been well recognized, the use of occupational data in the selection and classification of personnel is a relatively recent development.

In 1973, the Air Force Manpower and Personnel Center (AFMPC) requested that the Air Force Human Resources Laboratory (AFHRL) conduct research to develop an objective procedure for establishing relative aptitude requirements for enlistee occupations. After 10 years of extensive research, the Laboratory developed a state-of-the-art technology for this purpose. The technology produces job-centered measures of occupational learning difficulty. Occupational learning difficulty is operationally defined as the time it takes to learn to perform an occupation satisfactorily (Mead, 1970a, 1970b; Mead & Christal, 1970; & Lecznar, 1971).

The use of learning difficulty data for establishing relative aptitude requirements is based on the assumption that aptitude (i.e., aptitude minimums) and learning time (i.e., task learning difficulty) are related. Considerable research has demonstrated a high correlation between task aptitude requirement ratings from behavioral scientists and task difficulty ratings from senior Air Force supervisors (Fugill, 1972). Additional research by Block and Anderson (1975), Cronbach and Snow (1977), Gettinger and White (1979) and Krumboltz (1965) lend further support to the notion that aptitude is related to learning time.

In deriving measures of occupational learning difficulty, three types of occupational information were employed: (a) task time-spent ratings provided by incumbents, (b) supervisory ratings of task difficulty--both available from the Air Force Occupational Measurement Center (OMC), and (c) benchmark ratings of task learning difficulty obtained through evaluations by contract personnel. Benchmark ratings were necessary because supervisory ratings of task difficulty only provided information concerning the relative order of tasks within occupations. Consequently supervisory ratings were not comparable across occupations. On the other hand, benchmark ratings of task learning difficulty which are based on task-anchored benchmark rating scales (Burtch, Lipscomb, & Wissman, 1982) are comparable across occupations. The benchmark rating scales were designed to capture the range of learning difficulty characteristic of all tasks within occupations in a given aptitude area.

Benchmark Procedure

Benchmark ratings were collected for the ultimate purpose of adjusting supervisors' task ratings so they would be comparable across occupations. To date, occupational learning difficulty measures have been derived for more than 200 enlisted Air Force Specialties (AFSs), representing approximately 100,000 tasks and 170,000 enlisted positions.

The procedure followed by the contractor to collect benchmark learning difficulty ratings for a particular specialty consisted of the following steps. First, 60 tasks from the associated job/task inventory were selected based on the following criteria:

1. Non-supervisory: tasks performed solely by supervisors were eliminated.

2. Captured the range of difficulty: tasks were selected to represent the range of supervisor relative ratings of learning difficulty.

3. High rater agreement on the difficulty of the task: tasks were selected for which there was high agreement among raters when judging learning difficulty.

4. Performed by first-term (1 to 48 months) personnel: tasks were selected that were performed by persons in their first term of enlistment.

5. Frequently performed: tasks that were most frequently performed were selected.

6. Easily observed: tasks that could be easily observed were selected.

7. High face validity: tasks were selected that appeared valid for representing the range of learning difficulty associated with the occupation.

Th content validity of each of these tasks was then confirmed with technical school instructors.

Second, a group of 14 raters, each having expert knowledge of the tasks to be rated, visited operational sites to interview job incumbents to gather information regarding performance of each of the 60 tasks. The raters were divided into two teams of seven members each, and each team visited a different operational site to obtain information for the same set of tasks. Third, after becoming familiar with the tasks to be rated, each team member independently provided ratings of task learning difficulty using a benchmark rating scale. Briefly, a benchmark rating scale is a 25-point scale with each point anchored or defined by two tasks of equivalent learning difficulty (Burtch, Lipscomb, & Wissman, 1982). Fourth, the ratings were averaged across all raters. Fifth, the average benchmark difficulty ratings were used to adjust average task ratings of relative learning difficulty supplied by occupational supervisors. These adjusted task ratings are referred to in this paper as benchmark learning difficulty estimates. Unlike supervisory ratings, the difficulty estimates are comparable across occupations. These benchmark learning difficulty estimates for each task were then multiplied by the time spent performing each task for each occupational incumbent, and the products were summed across tasks and divided by a factor of 10 to produce an index referred to as an Average Task Difficulty Per Unit Time Spent (ATDPUTS). Although the ATDPUTS are computed on a position-by-position basis, they are averaged to represent occupational learning difficulty for specified incumbent groups.

Once the estimates of occupational learning difficulty were available, they were used for two separate organizational purposes. First, they were used in combination with training and recruiting information for establishing aptitude requirement minimums stated in AFR 39-1, Airman Classification Regulation. Second, as described by Weeks (1984), these measures were used in the Air Force person-job match algorithms for determining individual assignments to specialties. Since measures of occupational learning difficulty will be used for these purposes in the future, it is important that the associated measurement procedure be transferred to an operational organization for routine implementation.

Objective of the Present Study

To transfer the learning difficulty measurement technology from a research to an operational setting, it was necessary to investigate the possibility of having benchmark ratings routinely collected by Air Force personnel rather than by contract job analysts. Hence, AFHRL, in coordination with the OMC, conducted a feasibility study. The study was designed to assess the feasibility of having Air Force personnel collect benchmark ratings for use in deriving measures of occupational learning difficulty. This paper describes the approach taken and results obtained in addressing this research issue.

II. METHOD

Selection of Specialties

Nine Air Force enlisted personnel specialties were selected for assessment:

Air Force Specialty Code (AFSC)	Specialty
251X0	Weather
272X0	Air Traffic Control
316X3	Instrumentation
325X0	Automatic Flight Control Systems
325X1	Avionics Instrument Systems
426X2	Jet Engine Mechanic
426X3	Turboprop-Propulsion Mechanic
431X1	Tactical Aircraft Maintenance
431X2	Airlift/Bombardment Aircraft Maintenance

The selection of these AFSs was based on the following criteria: first, benchmark learning difficulty ratings collected by contractor personnel were available, and second, the specialties were representative of the Mechanical, Electronic, and General aptitude areas.

Development of Task Lists

For each specialty, the representative sample of tasks originally used by the contractor was obtained. Each sample consisted of 60 tasks selected from job inventories containing between 518 and 1124 tasks. The 60 tasks were those originally selected during the initial contract effort.

Selection of Raters

Once the selection of specialties and development of task lists had been accomplished, the next step was to select Air Force personnel to collect benchmark task difficulty ratings for each of the nine specialties. For this purpose, OMC provided eight inventory developers and job analysts to serve as raters. OMC personnel were especially suited to participate as raters because of their expertise in collecting and analyzing job and task data. The OMC raters consisted of four military and four civilian personnel. The selected personnel had an average of 5 years experience in the areas of Air Force inventory development and occupational analysis.

Training of Raters

To familiarize the OMC personnel with the specialized procedures used in collecting benchmark data, a 3-day training workshop was conducted at AFHRL. Training consisted of instruction and exercises on procedures for analyzing tasks to determine learning difficulty. An important aspect of the training dealt with the use of the benchmark rating scales and how ratings are assigned. Training also focused on the conduct of interviews, the principal method for gathering information on the tasks for which ratings of learning difficulty are produced. Procedural guides for using the benchmark scales were prime training materials. The guides included definitions of assessment criteria and descriptions of each task anchor on each benchmark scale. Attention was placed on the use of eight task assessment criteria as guidelines in determining exactly what each task is, how it is performed, and what skills or knowledge are required to perform it adequately. The eight assessment criteria are (a) task definition, (b) number of steps in a task, (c) tools and equipment unique to the task, (d) regulations, manuals, and standard operating procedures, (e) memorization, (f) standards of performance, (g) time criticality, and (h) basic skills or knowledge. The course objectives for the training are provided in Appendix A. Definitions of each of the criteria are provided in Appendix B.

Upon completion of the 3-day training workshop, the eight OMC raters were divided into two teams of four members each. The two teams then conducted practice sessions in the field, accompanied by an AFHRL instructor, to apply the various techniques that had been learned for use in deriving benchmark task learning difficulty ratings. In addition, these practice sessions were conducted to assess the effectiveness of the training and to determine any problem areas where more training might be warranted.

The specialties studied during the practice sessions were Air Force Specialty Code (AFSC) 811X0, Security Police, and AFSC 341X5, Analog Navigation/Tactics Training Devices. These specialties were selected due to the availability of personnel in the local area. Both practice sessions were conducted at Randolph AFB.

Each team independently interviewed a 3-skill-level and a 5-skill-level airman from AFSC 811X0 and AFSC 341X5. Each interview session lasted about 4 hours. During the sessions, OMC team members queried the airmen regarding the performance of each of the 60 tasks. Questions were directed toward task performance relative to the eight assessment criteria previously listed.

After each task on the list had been addressed and team members were confident that they had acquired enough information about the performance of each task, the session was terminated. Each team member independently rated the learning difficulty of each task using the appropriate benchmark task difficulty rating scale. Individual ratings for each task were then summed and divided by four to derive an average team rating for each task.

For ratings collected during the practice sessions, acceptable levels of interrater agreement were reached for both team 1 and team 2 (R_{11}=.72 for team 1 and .80 for team 2). In addition, both teams agreed that training had been effective and that the two practice sessions were helpful in developing an understanding of the rating procedures.

Field Study

Once OMC personnel had been trained in the benchmark rating procedure, each team visited different operational sites to interview occupational incumbents from each of nine enlisted specialties. On the basis of information gathered during these interviews, each team member independently rated each of the 60 tasks for each specialty in terms of learning difficulty. Data collected during this phase were to be used in the assessment of the feasibility of operational implementation of the learning difficulty technology. Operational site visits were made to Randolph, Kirtland, Holloman, Davis-Monthan, and Barksdale AFBs during a 3-week period. A listing of the specialties studied by base and by team is provided in Appendix C.

At each site, two or more airmen from each specialty were interviewed as a group. Each interview session lasted approximately 4 hours. After interviews, each team member independently assigned a benchmark rating to each task. Independent ratings were then averaged across raters within each team to yield the average benchmark difficulty rating for each task rated. In addition, ratings of task difficulty were averaged across team 1 and team 2 members to provide average task ratings for the total team. This was done for each of the nine specialties studied.

Upon completion of the operational site visits, OMC teams forwarded their ratings to AFHRL for data processing and analyses. In analyzing the data from the OMC field study, data originally collected by the contractor were used for comparative purposes.

III. ANALYSIS

There were three major goals of the analyses. The first goal was to determine the internal consistency of the 25-point benchmark task difficulty ratings collected by OMC team members. Descriptive statistics on the ratings were obtained for OMC teams 1 and 2 and for contractor teams 1 and 2. Reliability of ratings was assessed using the CODAP program REXALL (Christal & Weissmuller, 1976), which computes an index of interrater agreement. Reliability indices were derived for OMC team 1, team 2, and total team and for contractor team 1, team 2, and total team. In addition, intercorrelations for OMC team 1 vs. team 2 and contractor team 1 vs. team 2 were computed.

The second goal of the analyses was to determine the validity of the OMC benchmark ratings. Validity was assessed first by comparing the characteristic response patterns of the OMC total team and contractor total team by specialty and second by comparing the relationship between OMC teams' and supervisors' ratings of task difficulty and contractor teams' and supervisors' ratings of task difficulty. Relationships among benchmark ratings were analyzed using TRICOR, a general purpose correlation program (Black, 1978).

The third goal of the analysis was to compare measures of occupational learning difficulty (ATDPUTS) generated from OMC benchmark data with those generated by the contractor. ATDPUTS were generated for each of the nine specialties based on OMC total team benchmark ratings. Specialties were then rank ordered (from highest to lowest) in terms of their associated ATDPUTS. The ranking of specialties was then compared to the ranking of specialties based on contractor-derived ATDPUTS, to determine the consistency of rankings between the two teams.

IV. RESULTS

Reliability

A comparison of OMC teams 1 and 2 in terms of average ratings of task learning difficulty is presented in Table 1, together with the associated standard deviations by specialty. Seven of the nine specialties were found to be roughly equivalent, falling around the mid-point of the scale. Significant differences were found, however, between teams for AFSC 251X0, Weather, (t (102) = 4.69, $p \leq .05$) and AFSC 426X2, Jet Engine Mechanic, (t (94) = 2.42, $p \leq .05$). Differences were attributable to the fact that team 1 tended to rate tasks higher in learning difficulty than did team 2 for AFSC 251X0; whereas, team 1 tended to rate tasks lower than did team 2 for AFSC 426X2.

Table 1. Average Task Difficulty Ratings for OMC Team 1 and Team 2 Personnel

	OMC Team 1 (N = 4)		OMC Team 2 (N = 4)	
AFSC	$\bar{X}$	SD	$\bar{X}$	SD
251X0*	15.0	3.23	11.9	3.77
272X0	14.4	2.87	14.1	2.84
316X3	13.8	3.26	13.9	3.86
325X0	13.9	2.79	13.1	3.19
325X1	13.2	2.08	13.5	3.42
426X2*	11.9	2.34	13.5	3.98
426X3	13.6	2.39	13.3	3.75
431X1	12.6	1.91	13.4	2.50
431X2	12.8	1.94	12.6	2.54

*An asterisk denotes a significant difference ($p \leq .05$).

Descriptive statistics for contractor teams 1 and 2 are provided in Appendix D. For eight of the nine specialties, no significant differences were found between teams, in the average ratings of difficulty. A significant difference was found for AFSC 251X0, Weather. For this specialty, contractor team 1, on the average, rated tasks lower than did team 2.

Estimates of interrater reliability (R_{kk}) for rating groups were obtained for OMC total team (k = 8), team 1 (k = 4), and team 2 (k = 4) and for contractor total team (k = 14), team 1 (k = 7), and team 2 (k = 7). Reported R_{kk} values for the contractor teams were based on an *N* of 8 for the contractor total team and an *N* of 4 for each of the two contractor teams. This adjustment was performed using the Spearman-Brown formula to render contractor team data more comparable to OMC team data. Across the nine specialties studied, the range of R_{kk} values for the OMC total team was .81 to .95, and the median (Mdn) was .90. The ranges of R_{kk} values for OMC team 1 and team 2 were .62 to .89 (Mdn = .81) and .74 to .94 (Mdn = .93), respectively. For the contractor total team, R_{kk} values ranged from .87 to .95 (Mdn = .94) across the nine specialties while values ranged from .62 to .89 (Mdn = .81) for contractor team 1 and .89 to .94 (Mdn = .93) for contractor team 2.

Intercorrelations among ratings for OMC team 1 and team 2 and contractor team 1 and team 2 were computed as a further check of the internal consistency of team ratings. Correlation coefficients for OMC team 1 vs. team 2 and contractor team 1 vs. team 2 for each of the nine specialties are presented in Table 2. The significance of each correlation coefficient was evaluated using a critical-ratio *z* test. All correlations indicated significant agreement ($p \leq .05$) within rating groups.

Table 2. Correlations Between OMC Team 1 and Team 2 and Contractor Team 1 and Team 2

AFSC	OMC Team 1 vs. Team 2	Contractor Team 1 vs. Team 2
251X0	.81	.79
272X0	.62	.72
316X3	.45	.84
325X0	.86	.81
325X1	.74	.78
426X2	.78	.80
426X3	.80	.84
431X1	.43	.81
431X2	.62	.61

Note. All correlations are significant ($p \leq .05$).

Validity

The means and standard deviations for OMC and contractor total teams across each of the nine specialties are presented in Table 3. Contractor teams originally rated 60 tasks for each specialty. For purposes of analysis, however, only those tasks rated by OMC teams were considered in analyzing contractor data. These task subsets included from 43 to 55 of the original 60 tasks. In general, the average task difficulty ratings by OMC and contractor personnel were roughly equivalent, falling around the mid-point of the 25-point benchmark scales. Significant differences in mean values between teams were found, however, for AFSC 251X0, Weather, (t (102) = 3.28, $p \leq .05$); 272X0, Air Traffic Control, (t (108) = 5.62, $p \leq .05$); and 431X2, Airlift/Bombardment Aircraft Maintenance, (t (106) = 2.31, $p \leq .05$).

Table 3. **Average Task Difficulty Ratings by OMC and Contractor Personnel**

AFSC	OMC (N = 8)		Contractor (N = 14)	
	$\overline{X}$	SD	$\overline{X}$	SD
251X0*	13.5	3.34	11.3	3.55
272X0*	14.2	2.56	11.5	2.50
316X3	13.8	3.03	13.1	2.94
325X0	13.5	2.88	13.9	2.94
325X1	13.3	2.58	13.0	2.58
426X2	12.7	2.98	13.3	3.19
426X3	13.5	2.88	13.3	2.93
431X1	13.0	1.87	13.1	2.26
431X2*	12.7	2.03	13.6	1.90

*An asterisk denotes a significant difference ($p \leq .05$) between OMC and contractors in terms of their average ratings.

Intercorrelations among ratings were computed for each specialty for OMC total team, team 1, and team 2; contractor total team, team 1, and team 2; and supervisors. Of primary interest were the correlations between OMC and supervisors, contractors and supervisors, and OMC and contractors. Supervisory ratings were used as the criterion measure. Correlation coefficients for each of these three groups are provided in Table 4. Correlations for seven of the nine specialties were sufficiently high to suggest a strong relationship among teams' ratings of learning difficulty. Differences ($p \leq .05$) were found for OMC vs. supervisors and for contractors vs. supervisors for AFSC 426X2, Jet Engine Mechanic; AFSC 431X1, Tactical Aircraft Maintenance; and AFSC 431X2, Airlift/Bombardment Aircraft Maintenance. In the case of 426X2, the correlation between contractors and supervisors was somewhat lower than between OMC and supervisors. For 431X1 and 431X2, the OMC team correlated substantially lower with supervisors than did the contractors with supervisors. Full correlation matrices for each of the nine specialties for all seven rating groups are provided in Appendix D.

Table 4. **Correlations Between OMC and Contractor Ratings with Supervisors**

AFSC	OMC vs. Supervisors	Contractors vs. Supervisors	OMC vs. Contractors
251X0	.75	.83	.79
272X0	.74	.77	.86
316X3	.71	.77	.89
325X0	.85	.89	.88
325X1	.81	.78	.88
426X2*	.82	.73	.91
426X3	.78	.79	.90
431X1*	.59	.88	.75
431X2*	.54	.79	.72

*An asterisk denotes a significant difference ($p \leq .05$) between rating groups based on a Hotteling-t.

Correlations between OMC team 1 vs. supervisors and OMC team 2 vs. supervisors are presented in Table 5. For OMC team 1, significant agreement ($p \leq .05$) was found for eight of the nine specialties. Significance of r was again tested using the critical-ratio z test. For AFSC 431X1, Tactical Aircraft Maintenance, the obtained z fell below the .05 level of significance to indicate a lack of agreement between OMC team 1 and supervisors. The correlation coefficients obtained for team 2 vs. supervisors were all found to be significant.

Table 5. Correlations Between OMC Team 1 and Team 2 with Supervisors

AFSC	OMC Team 1 vs. Supervisors	OMC Team 2 vs. Supervisors
251X0	.74	.70
272X0	.60	.74
316X3	.66	.56
325X0	.84	.80
325X1	.71	.79
426X2	.82	.75
426X3	.75	.74
431X1	.29	.68
431X2	.36	.59

Note. All correlations are significant ($p \leq .05$) with the exception of OMC team 1 vs. supervisors for 431X1, based on a critical-ratio z test.

Correlation coefficients for contractor teams 1 and 2 vs. supervisors are provided in Appendix D. For both teams, significant agreement ($p \leq .05$) was found with supervisors.

Comparability of ATDPUTS

Average Task Difficulty Per Unit Time Spent (ATDPUTS) values were computed for first-term airmen for each specialty based on OMC team 25-point benchmark ratings. Previously derived ATDPUTS based on contractor total team benchmark ratings were examined for comparison purposes. ATDPUTS based on OMC total team and contractor total team are illustrated in Table 6 for first-term airmen across the nine specialties. For each team, ATDPUTS are ranked in descending order, from highest to lowest. The ATDPUTS derived for OMC total team was compared with that for the contractor total team. A Spearman rho correlation coefficient was computed to determine the relationship between the two rankings. The value of the Spearman rho was +.3 (ns, $p \leq .05$). For seven of the nine specialties (AFSCs 316X3, 325X0, 325X1, 431X2, 251X0, 426X3, and 431X1), the difference in rank was within one or two places, and in some cases, specialty ranks were identical. However, for AFSCs 272X0 and 426X2, rank differences were five and seven ranked places, respectively, between OMC and contractor ATDPUTS.

Table 6. OMC and Contractor ATDPUTS for First-Term Airmen

OMC			CONTRACTOR		
AFSC	$\bar{X}$	SD	AFSC	$\bar{X}$	SD
272X0*	132	6.73	426X2*	144	8.92
426X3	132	6.80	426X3	131	6.51
316X3	131	9.24	325X0	130	9.25
325X0	129	8.89	325X1	128	7.41
325X1	129	6.77	316X3	123	10.01
426X2*	128	10.50	431X2	122	8.44
431X1	120	5.83	431X1	112	10.18
431X2	118	6.54	272X0*	106	7.05
251X0	114	13.49	251X0	89	17.7

*An asterisk denotes those AFSCs where significant differences in rankings were found between the two teams.

V. DISCUSSION and CONCLUSIONS

The analysis reported herein assessed the feasibility of having Air Force personnel collect benchmark task learning difficulty ratings for use in deriving measures of occupational learning difficulty. This assessment was necessary to evaluate the transferability of the benchmark learning difficulty technology from a research to an operational setting.

Overall, results of the analysis were positive. That is, they supported the feasibility of having Air Force personnel, in particular OMC personnel, collect benchmark ratings. The interrater reliabilities based on a median R_{kk} of .81 to .93 across OMC rating groups showed high agreement for each of the specialties studied. Reliability of OMC ratings was consistent with reliabilities obtained for contractor ratings. Hence, the reliability of OMC ratings was acceptable.

For the majority of specialties, OMC average ratings were consistent with contractor average ratings. Significant differences were noted, however, for AFSC 251X0, Weather; 272X0, Air Traffic Control; and 431X2, Airlift/Bombardment Aircraft Maintenance. These differences may be attributed to any number of sources; for example, variations in procedures for collecting ratings (i.e., site differences and equipment differences). More study would be helpful in answering some of the questions regarding these differences. However, a more practical approach would be to assure that prescribed data collection procedures are explicitly and consistently followed in the future. If differences are noted, then it may be necessary for rating teams to return to the field to reaccomplish the data collection phase in an attempt to provide more consistent ratings across teams. Such was the procedure used by the contractor teams. This study did not allow for reaccomplishment of data collection given its status as a feasibility study.

Correlation coefficients between benchmark scale ratings and the criterion (supervisory ratings) across OMC rater groups and Air Force specialties were generally high and positive. Exceptions, however, were found for the AFSC 431X1, Tactical Aircraft Maintenance, and AFSC 431X2, Airlift/Bombardment Aircraft Maintenance, specialties. For AFSC 431X1, OMC total team and team 1 ratings correlated substantially lower with supervisory ratings in comparison to contractor teams. The reasons for this were not readily evident. For the Airlift/Bombardment Aircraft Maintenance specialty (AFSC 431X2), OMC team ratings again correlated lower than did the contractor's with supervisory ratings. Lower correlations could be attributable (a) to individual differences in the perception of task difficulty, (b) to the wide variation in aircraft being maintained by the incumbents interviewed (e.g., C-130s at Kirtland and B-52s at Barksdale), and (c) to the level of experience of airmen interviewed. Differences in contractor team ratings may be attributable to these factors as well.

In general, ATDPUTS derived on the basis of the OMC total team and contractor total team benchmark ratings were fairly consistent. Exceptions, however, were AFSCs 272X0, 426X2, and 251X0. ATDPUTS based on OMC-collected ratings for AFSC 272X0 were markedly higher than contractor ATDPUTS. In fact, AFSC 272X0 was ranked highest among the nine specialties studied by OMC; this difference can be attributed to the fact that OMC teams, on the average, rated tasks higher in learning difficulty than did the contractor teams for this specialty. ATDPUTS differences for AFSC 426X2, Jet Engine Mechanic, result from lower ratings being applied by OMC team 1 in comparison to OMC team 2 and contractor teams. This may be explained by the differences in equipment maintained at the operational sites visited. For example, at Davis-Monthan AFB, jet engine mechanics perform maintenance primarily on the A-10 aircraft; while at Kirtland AFB, a number of varied aircraft are maintained because it is a transient aircraft facility. Given more equipment being maintained at Kirtland, learning difficulty of tasks may be perceived as being more difficult as opposed to the difficulty associated with maintaining only one type of aircraft. Thus, it is theorized that equipment differences serve to account for ATDPUTS differences. For AFSC 251X0, a higher ATDPUTS value for OMC in comparison to the contractors can again be attributed to OMC team 1 ratings. OMC was significantly higher in their ratings of learning difficulty when compared to the other teams.

The data reported herein support the hypothesis that Air Force personnel can collect benchmark learning difficulty ratings. The evidence indicates that these ratings are both reliable and valid. These findings support the transferability of the benchmark technology. In transferring this technology, however, attention must be focused on some of the problems and issues associated with the collection of benchmark task difficulty ratings.

The primary issues to be considered in implementation are (a) the effects of task variability, (b) the selection of occupational incumbents to interview, and (c) the adequacy of task anchor definitions. In addition, several secondary issues need to be examined to assure proper utilization of the procedure for collecting benchmark ratings.

Task Variability

A key issue that has been noted during application of the benchmark learning difficulty scales is that of task variability. In discussing task variability, it is useful to consider how a task is defined. According to Morsh and Archer (1967), a task is defined as "a unit of work activity which forms a significant part of a duty." In writing a task statement for inclusion in a job inventory, it should (a) be understood by job incumbents, (b) be time ratable, (c) have a beginning and an end, and (d) be unambiguous. A major constraint on task specificity is that task lists (or job inventories) should take an airman no more than 2 hours to complete; hence, the number of tasks should generally be less than 1,000. Given these constraints, task statements can become broad. In particular, task statements become broad (a) when several specialties are described in a single task inventory (e.g., AFSC 325X0, Automatic Flight Control Systems and AFSC 325X1, Instrumentation), (b) when many aircraft are covered in a single task inventory (e.g., AFSC 431X1A-Z, Tactical Aircraft Maintenance, and AFSC 431X2A-E, Z, Airlift/Bombardment Aircraft Maintenance), (c) when specialties have a large amount of nonstandard equipment (e.g., AFSC 316X3, Instrumentation; AFSC 918X0, Biomedical Equipment Maintenance; or AFSC 324X0, Precision Measuring Equipment), and (d) when specialties have large amounts of equipment (e.g., AFSC 304X4, Ground Radio Communications).

Tasks that are too broad pose serious problems in terms of getting accurate estimates of learning difficulty. To exemplify, in the present data collection effort, incumbents from AFSC 325X0, Automatic Flight Control Systems, and AFSC 325X1, Avionics Instrument Systems, were asked to explain the following task: "Adjust attitude heading reference systems components." This task appeared in a single task inventory for both AFSs. When this task was explained by the AFSC 325X0 personnel, it was performed in-shop only, required special equipment (e.g., oscilloscope, directional gyro, three-axis table), required technical orders, required special field detachment training, and averaged 6 to 8 hours to perform. When described by AFSC 325X1 personnel, the only requirements for task performance were the technical orders and a screwdriver. It is possible that since the attitude heading reference system is an instrumentation system, the procedures used in the performance of this task are less complex for the AFSC 325X1, Avionics Instrument Systems specialist, than for the AFSC 325X0, Automatic Flight Control Systems specialist. The information provided by the AFSCs 325X0 and X1 personnel for the same task statement can be viewed as a reflection of an actual difference in the tasks and jobs performed by personnel in these two specialties. This could present a problem if not treated carefully because it could result in wide variations in learning difficulty estimates for the same task in the same job inventory. While these variations may point to a problem with the benchmark methodology, it appears to be more reflective of the way inventories are constructed and analyzed.

The decision to study two or more specialties using the same job inventory is based on the perceived similarity in the jobs performed by personnel in the specialties under study. This was the case with the AFSCs 325X0 and 325X1 specialties for example. However, the jobs and tasks were very different, hence the differences in perceived learning difficulty estimates. One solution may be to develop job inventories to study each Air Force specialty separately. An alternative would be to better specify tasks on job inventories when specialties are studied together, or job analysts could analyze supervisory ratings separately for each specialty.

Interviewing of Occupational Personnel

Another key issue pertains to the recommended personnel for providing task information about specialties. For collecting task information 3-level (apprentice) or junior 5-level (journeymen) airmen were interviewed to gather information for evaluating tasks in terms of their learning difficulty. Based on the experiences of the OMC rating teams, however, these personnel appeared to be limited in knowledge and experience on the performance of key tasks, resulting in the rating teams' inability to rate those unaddressed tasks. Across specialties, between 5 and 17 tasks could not be rated because little or no information was provided by the interviewees. To assure that a greater number of tasks could be rated, it is recommended that senior 5-level and 7-level personnel be used during interviews since their range of knowledge and experience would be greater than that of junior personnel.

Anchor Task Definitions

Focusing on the procedural guides (Hart, 1976) used for applying benchmark ratings, it has been noted by team members that many of the anchor tasks on the benchmark scales are not adequately defined in the available procedural guides. According to the procedure for benchmarking, a task should be evaluated relative to eight assessment criteria (see Appendix B). However, some of the anchor tasks on the benchmark scale are poorly defined and are not evaluated for all eight criteria. This has resulted in the raters having difficulty in comparing tasks to be rated against those in the benchmark scales. Therefore, it is recommended that the definition of the anchor tasks be refined, more specific evaluations be made on the eight criteria, and new procedural guides developed. It may also be necessary to evaluate each anchor task on the benchmark scales to determine whether (a) the task is still performed in the AFS; (b) the task is still performed by first-term airmen; (c) the task is representative in terms of task type, equipment, system or tools used; and (d) the task is difficult to comprehend or is worded in a manner that might confuse the rater. If it is determined that an anchor task does not meet these criteria, it should be replaced by a task that does.

Secondary Issues

The issues of task variability, recommended personnel to be interviewed, and definition of anchor tasks are viewed as the primary issues regarding the transfer of this technology. In addition to these issues, a number of secondary issues need to be adressed.

How many and which tasks should be benchmarked?

In this study, as well as the previous learning difficulty research, 60 tasks were rated for each specialty. This seems to be a reasonable number of tasks to use for future data collection efforts as well. Given the fact that it takes an average of 4 hours to interview personnel on the full set of 60 benchmark tasks, increasing the number of tasks to be rated would in turn increase the time required to conduct interviews. In addition, confounding variables (e.g., fatigue, motivation, job requirements, and demands) may interfere such that tasks rated toward the end of the interview would not be considered as extensively by the airmen as the first, thus yielding invalid results. Selecting fewer tasks to benchmark may adversely affect the stability of the learning difficulty estimates for the specialty studied. Hence, 60 tasks is considered most practical for use in benchmarking. Tasks selected to be benchmarked should be based on the task selection criteria outlined in Section II of this paper.

What specialty data are necessary for benchmarking?

Specialty data necessary for benchmarking include (a) a recent job inventory, (b) task-time-spent data from incumbents, and (c) supervisory task difficulty ratings.

When should a specialty be benchmarked?

The decision to benchmark a specialty is a policy decision rather than a scientific decision. An obvious and practical time is whenever a new job inventory is developed reflecting major revisions from the previous job inventory or when major changes occur in a specialty (e.g., restructuring).

Who should collect benchmark ratings?

In this study, inventory developers and job analysts from OMC collected benchmark task learning difficulty ratings for each of nine specialties. It is recommended that inventory developers and job analysts be trained in the use of the benchmark scales and associated procedural guides and be used in future data collection efforts because they possess a broad knowledge of Air Force enlisted specialties and have the expertise in routinely collecting and analyzing task factor data. The results of this research support the use of these personnel for collecting benchmark ratings in future data collection efforts. The experience and knowledge of these personnel contributed substantially in collecting reliable and valid benchmark difficulty data.

How many raters should there be per team?

In the present study, two teams of four raters each were used to collect benchmark ratings, whereas the original contractor study utilized two teams of seven raters each. To address the issue of how many raters per team are needed, interrater reliability analyses were conducted. In performing this analysis, the median R_{11} (.61) for both OMC teams combined was used to calculate the number of persons required per rating team to achieve an R_{kk} of approximately .90. The results of the analysis indicated that a team size of five proved more practical based on an obtained R_{kk} of .89. Thus, for future data collection efforts, an N of 5 members per team is recommended.

How many sites should be visited?

At a minimum, two sites should be visited; one site per team, per specialty. However, if between-site differences are known for a specialty, such as the case with AFSC 426X2, Jet Engine Mechanic, it would be advisable to visit additional bases to assure a more representative sampling of task performance. In consideration of which sites to visit, Air National Guard, Air Force Reserve, and research sites should be avoided.

VI. SUMMARY

This study established the feasibility of having Air Force personnel collect benchmark learning difficulty ratings supporting the transfer of this technology from a research to an operational setting. It is suggested that to ensure the success and effectiveness of this methodology, attention be focused on the implementation issues addressed in this paper. In so doing, the quality of the collected data and resulting data base will be assured.

REFERENCES

Air Force Regulation 39-1. (1982). Airman classification regulation. Washington, DC: Department of the Air Force.

Black, D.E. (1978). Mathematical and statistical software index (AFHRL-TR-78-66, AD-A062 991). Brooks AFB, TX: Computational Sciences Division, Air Force Human Resources Laboratory.

Block, J.R., & Anderson, L.W. (1975). Mastery learning in classroom instruction. New York: Macmillan.

Burtch, L.D., Lipscomb, M.S., & Wissman, D.J. (1982). Aptitude requirements based on task difficulty: Methodology for evaluation (AFHRL-TR-81-34, AD-110 568). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

Christal, R.E., & Weissmuller, J.J. (1976). New CODAP programs for analyzing task factor information (AFHRL-TR-76-3, AD-A026 121). Lackland AFB, TX: Occupational and Manpower Research Division, Air Force Human Resources Laboratory.

Cronbach, L.J., & Snow, R.E. (1977). Aptitudes and instructional methods. New York: Irvington.

Fugill, J.W.K. (1972). Task difficulty and task aptitude benchmark scales for the mechanical and electronics career fields (AFHRL-TR-72-40, AD-754 848). Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

Gettinger, M., & White, M.D. (1979). Which is the stronger correlate of school learning? Time to learn or measured intelligence? Journal of Educational Psychology, 71(4), 405-412.

Hart, F. (1976). Procedural guides for use of the benchmark scales. Unpublished manuscript.

Krumboltz, J.D. (Ed). (1965). Learning and the educational process. Chicago: Rand McNally.

Lecznar, W.B. (1971). Three methods for estimating difficulty of job tasks (AFHRL-TR-71-30, AD-730 594). Lackland AFB, TX: Personnel Division, Air Force Human Resources Laboratory.

Mead, D.F. (1970a). Continuation study on development of a method for evaluating job difficulty (AFHRL-TR-70-43, AD-720 254). Lackland AFB, TX: Personnel Division, Air Force Human Resources Laboratory.

Mead, D.F. (1970b). Development of an equation for evaluating job difficulty (AFHRL-TR-70-42, AD-720 253). Lackland AFB, TX: Personnel Division, Air Force Human Resources Laboratory.

Mead, D.F., & Christal, R.E. (1970). Development of a constant standard weight equation for evaluating job difficulty (AFHRL-TR-70-44, AD-720 255). Lackland AFB, TX: Personnel Division, Air Force Human Resources Laboratory.

Morsh, J.E., & Archer, W.B. (1967). Procedural guide for conducting occupational surveys in the United States Air Force (PRL-TR-67-11, AD-664 036). Lackland AFB, TX: Personnel Research Laboratory, Aerospace Medical Division.

Weeks, J.L. (1984). Occupational learning difficulty: A standard for determining the order of aptitude requirement minimums (AFHRL-SR-84-26, AD-147 410). Brooks AFB, TX: Manpower and Personnel Division, Air Force Human Resources Laboratory.

APPENDIX A: TRAINING COURSE OBJECTIVES

Table A-1. Training Course Objectives

COURSE TITLE: Application of Benchmark Rating Scales for Task Learning Difficulty

AIM: The aim of this course is to prepare occupational analysts to apply, in the field, each of the three benchmark rating scales for obtaining task learning difficulty data.

ELIGIBILITY: Participants of the course will be Inventory Developers and/or Job Analysts from the USAF Occupational Measurement Center (OMC). This implies that they will have familiarity with the occupational structure of the Air Force, and that they are experienced at interviewing personnel and observing tasks in order to generate occupational task inventories.

LOCATION: The course venue is the Air Force Human Resources Laboratory, Brooks AFB, Texas.

COURSE TERMINAL OBJECTIVES: Students are expected to meet the following objectives:

1. Objective. State the definition of the task factor term "learning difficulty."

   Condition. State verbally.

   Standard. Without error.

2. Objective. List and define the seven assessment criteria for task learning difficulty.

   Condition. State verbally.

   Standard. Without error.

3. Objective. Assign benchmark ratings to target tasks.

   Condition. Both in a group situation and independently.

   Standard. Instructor satisfaction.

4. Objective: Question selected AFS incumbents on the assessment criteria of task learning.

   Condition: As a group, question individual incumbents.

   Standard: Instructor satisfaction that sufficient information on the assessment criteria for learning difficulty is obtained.

5. Objective: Use the procedural guide associated with each benchmark scale so as to become familiar with the anchor tasks which define each point on the scale.

Condition: Individually to the satisfaction of the instructor.

Standard: To use the procedural guide effectively to rank the Mechanical benchmark anchor tasks without prior familiarization (3rd day exercise).

6. Objective: Describe the procedure used to obtain benchmark ratings of task learning difficulty.

Condition: Describe verbally.

Standard: Without error.

7. Objective: Select a representative subset of target tasks for benchmarking via DIFBEN.

Condition: Group situation.

Standard: Instructor satisfaction.

APPENDIX B: TASK ASSESSMENT CRITERIA DEFINITIONS

Table B-1. **Task Assessment Criteria Definitions**

1. Task definition: What is the task? What is and is not included in task performance? For example, if the task is changing spark plugs, must other components (e.g., air filter, compressor) be removed first, or is this a separate task?

2. The number of work steps in a task: Tasks that have many different steps are obviously more difficult to learn than those which have only a few steps. Tasks that contain many repetitions of the same step, however, may be relatively easy to learn.

3. Tools and equipment unique to the task: The learning time required for tools and equipment unique to a task adds to learning difficulty.

4. Regulations, manuals and standard operating procedures: How detailed is the documentation? The more detailed it is, the less has to be learned. Some tasks do not have to be learned, because they can be performed by simply following written instructions.

5. Memorization: Does the task or any portion of the task have to be memorized in order to be performed? This adds to the learning difficulty.

6. Standards of performance: Tasks differ in what level of quality or reliability is required for "satisfactory performance." For example, packing a parachute requires a higher standard of product reliability than does changing a faucet washer. In the latter case, if the faucet leaks, you can do it again.

7. Time criticality: A task that must be performed within a time limit is more difficult to learn than the same task with no time limit for performance.

8. Basic skills or knowledge: For many career fields, there are required basic skills or knowledges (e.g., typing, mathematics). In some cases these are taught in the USAF Technical School. These skills and knowledges add to the learning difficulty of individual tasks only to the extent that they are used in the performance of that task.

Note. Task Assessment Criteria Definitions were extracted from the procedural guide for the use of the benchmark scale.

APPENDIX C: LISTING OF SPECIALTIES STUDIED BY BASE AND TEAM

Table C-1. Listing of Specialties Studied by Base and Team

TEAM 1: Kirtland AFB, NM

426X3 - Turboprop-Propulsion Mechanic

431X2 - Airlift/Bombardment Aircraft Maintenance

325X1 - Avionics Instrument Systems

325X0 - Automatic Flight Control Systems

Holloman AFB, NM

426X2 - Jet Engine Mechanic

431X1 - Tactical Aircraft Maintenance

316X3 - Instrumentation

272X0 - Air Traffic Control

251X0 - Weather

TEAM 2: Davis-Monthan AFB, AZ

431X1 - Tactical Aircraft Maintenance

426X2 - Jet Engine Mechanic

426X3 - Turboprop-Propulsion Mechanic

325X1 - Avionics Instrument Systems

Barksdale AFB, LA

316X3 - Instrumentation

325X0 - Automatic Flight Control Systems

431X2 - Airlift/Bombardment Aircraft Maintenance

251X0 - Weather

Randolph AFB, TX

272X0 - Air Traffic Control

APPENDIX D: COMPARISONS BETWEEN RATING TEAMS ON BENCHMARK LEARNING DIFFICULTY

Table D-1. **Average Task Difficulty Ratings for Contractor Team 1 and Team 2**

AFSC	Contractor Team 1 (N = 7) X̄	Contractor Team 1 (N = 7) SD	Contractor Team 2 (N = 7) X̄	Contractor Team 2 (N = 7) SD
251X0*	10.0	4.03	12.5	3.46
272X0	11.2	3.05	11.8	2.34
316X3	12.9	3.04	13.2	3.52
325X0	13.7	2.80	14.1	3.38
325X1	13.4	2.73	12.7	2.74
426X2	13.2	3.15	13.3	3.56
426X3	12.9	2.96	13.6	3.39
431X1	13.0	2.36	13.0	2.64
431X2	13.7	2.09	13.4	2.18

*An asterisk indicates a significant difference ($p \leq .05$) between teams on average ratings of difficulty.

Table D-2. **Correlations Between Task Difficulty Ratings for Contractor Team 1 and Team 2 with Supervisors**

AFSC	Contractor Team 1 vs. Supervisors	Contractor Team 2 vs. Supervisors
251X0	.77	.80
272X0	.71	.72
316X3	.74	.74
325X0	.78	.90
325X1	.70	.77
426X2	.67	.72
426X3	.73	.78
431X1	.84	.83
431X2	.85	.58

Note. All correlations are significant ($p < .05$).

Table D-3. **Intercorrelations Among Ratings/Teams-AFSC 251X0**

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.95	.96	.79	.79	.69	.75
2. OMC Team 1		1.00	.81	.76	.76	.66	.74
3. OMC Team 2			1.00	.74	.75	.65	.70
4. Contractor Total Team				1.00	.96	.94	.83
5. Contractor Team 1					1.00	.79	.77
6. Contractor Team 2						1.00	.80
7. Supervisors							1.00

Note. All correlations are significant ($p < .05$).

Table D-4. Intercorrelations Among Ratings/Teams-AFSC 272X0

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.90	.89	.86	.80	.79	.74
2. OMC Team 1		1.00	.62	.77	.73	.70	.60
3. OMC Team 2			1.00	.77	.71	.72	.74
4. Contractor Total Team				1.00	.95	.91	.77
5. Contractor Team 1					1.00	.72	.71
6. Contractor Team 2						1.00	.72
7. Supervisors							1.00

Note. All correlations are significant ($P \leq .05$).

Table D-5. Intercorrelations Among Ratings/Teams-AFSC 316X3

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.82	.88	.89	.84	.86	.71
2. OMC Team 1		1.00	.45	.88	.85	.83	.66
3. OMC Team 2			1.00	.66	.61	.65	.56
4. Contractor Total Team				1.00	.95	.97	.77
5. Contractor Team 1					1.00	.84	.74
6. Contractor Team 2						1.00	.74
7. Supervisors							1.00

Note. All correlations are significant (p .05).

Table D-6. Intercorrelations Am[illegible]ings/Teams-AFSC 325X0

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.96	.97	.88	.79	.86	.85
2. OMC Team 1		1.00	.86	.82	.74	.82	.84
3. OMC Team 2			1.00	.86	.79	.84	.80
4. Contractor Total Team				1.00	.94	.96	.89
5. Contractor Team 1					1.00	.81	.78
6. Contractor Team 2						1.00	.90
7. Supervisors							1.00

Note. All correlations are significant ($p \leq .05$).

Table D-7. Intercorrelations Among Ratings/Teams-AFSC 325X1

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.90	.96	.88	.84	.82	.81
2. OMC Team 1		1.00	.74	.79	.75	.74	.71
3. OMC Team 2			1.00	.84	.81	.78	.79
4. Contractor Total Team				1.00	.94	.94	.78
5. Contractor Team 1					1.00	.78	.70
6. Contractor Team 2						1.00	.77
7. Supervisors							1.00

Note. All correlations are significant ($p \leq .05$).

Table D-8. **Intercorrelations Among Ratings/Teams-AFSC 426X2**

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.91	.97	.91	.88	.85	.82
2. OMC Team 1		1.00	.78	.84	.84	.77	.82
3. OMC Team 2			1.00	.87	.82	.82	.75
4. Contractor Total Team				1.00	.94	.96	.73
5. Contractor Team 1					1.00	.80	.67
6. Contractor Team 2						1.00	.72
7. Supervisors							1.00

Note. All correlations are significant ($p \leq .05$).

Table D-9. **Intercorrelations Among Ratings/Teams-AFSC 426X3**

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.92	.97	.90	.88	.86	.78
2. OMC Team 1		1.00	.80	.86	.85	.80	.75
3. OMC Team 2			1.00	.85	.82	.82	.74
4. Contractor Total Team				1.00	.95	.96	.79
5. Contractor Team 1					1.00	.84	.73
6. Contractor Team 2						1.00	.78
7. Supervisors							1.00

Note. All correlations are significant ($p \leq .05$).

Table D-10. **Intercorrelations Among Ratings/Teams-AFSC 431X1**

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.80	.89	.75	.78	.65	.59
2. OMC Team 1		1.00	.43	.42	.49	.32	.29
3. OMC Team 2			1.00	.80	.80	.73	.68
4. Contractor Total Team				1.00	.95	.96	.88
5. Contractor Team 1					1.00	.81	.84
6. Contractor Team 2						1.00	.83
7. Supervisors							1.00

Note. All correlations are significant ($p \leq .05$), with the exception of OMC team 1 vs. supervisors where the r was not significant.

Table D-11. Intercorrelations Among Ratings/Teams-AFSC 431X2

Variable	1	2	3	4	5	6	7
1. OMC Total Team	1.00	.87	.93	.72	.66	.63	.54
2. OMC Team 1		1.00	.62	.61	.53	.55	.36
3. OMC Team 2			1.00	.68	.64	.59	.59
4. Contractor Total Team				1.00	.89	.90	.79
5. Contractor Team 1					1.00	.61	.85
6. Contractor Team 2						1.00	.58
7. Supervisors							1.00

Note. All correlations were significant ($p \leq .05$).

Table D-12. OMC and Contractor Based ATDPUTS for 49 to 96 and Total Active Federal Military Service (TAFMS) Groups

	OMC				Contractor			
	49-96	TAFMS	Total	TAFMS	49-96	TAFMS	Total	TAFMS
AFSC	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD
272X0	136	8.25	138	9.85	110	8.59	113	11.23
426X3	134	5.95	134	7.42	133	5.70	133	7.11
316X3	132	9.97	135	10.59	124	10.82	127	11.35
325X0	129	8.94	130	8.95	131	9.38	132	9.44
325X1	130	7.28	131	7.20	128	8.07	129	7.99
426X2	132	8.21	132	11.05	148	6.24	148	9.67
431X1	123	6.10	125	7.92	116	10.66	119	13.82
431X2	121	6.94	123	8.52	122	8.44	126	8.96
251X0	124	16.23	126	17.79	101	20.90	104	23.00

END

FILMED

1-86

DTIC