PREDICTION OF OPERATOR EFFECTIVENESS IN DYNAMIC AIR TRAFFIC CONTROL SIMULATION

TECHNICAL REPORT

REPORT #30

PROJECT N

1 NOVEMBER 1959

COURTNEY AND COMPANY

PREDICTION OF OPERATOR EFFECTIVENESS IN DYNAMIC AIR TRAFFIC CONTROL SIMULATION

(Technical Report)

Report #30
Project N

1 November 1959

James W Danaher
John W Eberhard
Kenneth W Colman

PREPARED

FOR

Bureau of Research and Development
Federal Aviation Agency
Washington 25, D.C.
Contract FAA/BRD-27

BY

Courtney and Company
1711 Walnut Street
Philadelphia 3, Pennsylvania

ACKNOWLEDGMENTS

Special appreciation is hereby expressed to the following personnel of the Bureau of Research and Development at the Federal Aviation Agency

Major R L Webb, USAF, Systems Analysis Division, as Project Manager.

Mr Jack Grewell, Systems Analysis Division, who was instrumental in making this study possible

Mr Lloyd Van Rawlings, Chief, Simulator Operations Branch

Mr Alfred Castellano, Mr Justin Cervi, and Mr Robert Lucas, Simulator Operations Branch, for their cooperation and assistance in arranging for collection of the NAFEC data and in providing operator evaluation data

Mr John Hill, Systems Analysis Division, for his helpful suggestions in the early planning stages of the study

Mr F D Munroe, Personnel Branch at NAFEC, for obtaining the Civil Service Examination results on all operators, and for providing information regarding the nature of the tests and the Civil Service selection system

Mr E H Wright, Mr William Olson, Mr Ralph Johnson, and Mr George Scott, Simulation Team Members at NAFEC, for providing operator evaluation data

Special appreciation is also expressed to:

Mr Clair Anderson, and his fellow simulation supervisors at the Technical Development Center, Indianapolis, for their valuable suggestions regarding traits pertinent to an operator rating scale, and for providing TDC operator evaluation data

Dr. J S Kidd of the Aviation Psychology Laboratory, The Ohio State University, for his helpful suggestions regarding the development of rating instruments

All the simulator pilot operators, both at TDC and at NAFEC, for their interest and cooperation, with special thanks to the eight "potential supervisors" for providing the peer ratings

Dr William Angoff, for his valuable suggestions and assistance in the planning of the study, development of rating scales, analysis of the data, and critical review of the manuscript

Miss Rosa Bernstein, Courtney and Company research librarian, for her assistance in providing background material

Miss Elizabeth Bowman, Mrs Eleanor Wright, Miss Rhoda Drucker, Mrs Achsah Coursen, Miss Kay Trout and Mr John B Freeman of Courtney and Company, for the technical production of this report

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

BRIEF OF THE STUDY

FOREWORD

The research described in this report has been carried out for the Federal Aviation Agency under Contract FAA/BRD-27 Because of the importance attached to the operational implications of the results, a non-technical treatment is presented in Section One The remainder of the report provides an account of the origin and nature of the problem, followed by a technical presentation of the research methodology and findings

PURPOSE

This study was undertaken to develop a program for selecting simulator pilot operators to staff the dynamic air traffic control simulators at the Federal Aviation Agency's National Aviation Facilities Experimental Center (NAFEC), Atlantic City, New Jersey

RESULTS

The results of activity analyses on two such systems (the former at the CAA Technical Development Center, and the latter a simulator being constructed for FAA by Aircraft Armaments, Inc ,) were used in forming a composite Job Specifications list for simulator pilot operators. These analyses, combined with simulation supervisor opinions regarding job

requirements, served as a basis for the development of operator performance rating scales (an eight-trait rating scale and a stanine ranking scale) The data collected using these instruments were combined to form a composite criterion for use in the validation of 13 predictor variables The predictors were. a Civil Service Examination, ten tests comprising the Employee Aptitude Survey (EAS), Age, and Years of Education Validation studies were conducted on a dynamic air traffic control simulation system in two operational settings, i e , Technical Development Center, Indianapolis, and National Aviation Facilities Experimental Center, Atlantic City

The first (TDC) study, based on a population of experienced simulator operators (n =49), indicated that the Civil Service Examination currently in use as a selection device, was no better than chance (r = - 05) as a predictor of operator performance Only two of the predictor variables under study were significantly related to job performance ratings These were (1) the Verbal Reasoning test (38) of the EAS, and (2) Years of Education (31)

The second (NAFEC) study, based on the same equipment but at a different site and with a less experienced operator population (n = 51), indicated the Civil Service Examination as the single most effective prediction instrument (r = 51) In addition, five tests of the EAS (with validity coefficients ranging from 31 to 47) were significantly related to performance ratings

Optimum prediction of operator performance in the NAFEC results was determined using the Wherry-Doolittle Test Selection Method The results showed that best prediction was obtained by using four of the predictor variables Their order of contribution was Civil Service Examination, Age, Manual Speed and Accuracy, and Word Fluency (shrunken R = 67)

A comparison of results of the two studies indicated that there were no significant validity coefficients common to both The diverse results obtained were attributed to differences in the two operational situations, i e , type and length of job indoctrination and training, length of job experience, testing conditions, and rater qualifications

IMPLICATIONS

Ineffective prediction of operator performance with the Civil Service Examination in the first study, and effective prediction of performance with it in the second, suggests

1 A follow-up study should be executed on the NAFEC operator population to determine predictor effectiveness on a long-term basis at this facility

2 Caution should be used in extrapolation from the NAFEC results to the staffing of other simulator systems under development If practical, independent validation studies should be conducted for the staffing of the newly-developed systems

3 The relationship between Civil Service Examination scores and job performance ratings in the NAFEC results suggests that short term gains in operator effectiveness could be realized if greater selectivity were exercised In view of this, it is recommended that more emphasis be placed on recruiting in order to assure an adequate supply of applicants from which to choose If the long term ineffectiveness of the Civil Service Examination found in the TDC data is due to voluntary attrition of the more able selectees, the selection of only the better qualified applicants could be expected to result in a higher turnover rate with a consequent increased need for the selection and training of replacement operators

SECTION ONE

OPERATIONAL IMPLICATIONS OF THE STUDY

As an aid to management-level decision-makers, the operational implications of the results are presented in this section immediately following an introductory treatment of the nature and scope of the selection process as it relates to this investigation These implications are based on an analysis and interpretation of the mass of data collected during this study The complete data analysis is presented in detail in Section Four

I. IMPORTANCE OF THE PILOT OPERATOR IN THE SIMULATION PROCESS

Dynamic simulation techniques have become an important tool in the study of air traffic control problems As a part of an extensive long-range program for improving the nation's air traffic control system, the Bureau of Research and Development of FAA (at NAFEC, Atlantic City) has one dynamic air traffic control simulation system in operation, and has three others in various stages of completion Aside from the controller personnel required for staffing these systems, and the research personnel needed to conduct simulation studies, each of the systems requires approximately fifty trained simulator "pilots" to provide an approximation of the "outside-world" portion of the controllers' working environment [1]

[1] A complete description of these simulators and role of the operators who staff them is presented in Section Two

Since these "pilots" comprise the largest portion of the human element directly involved in the simulation process, their performance is a major determinant of the effectiveness of the dynamic simulators In view of this, the development of effective operator selection and training programs is considered a vitally important goal Effective operator selection is the essence of this study

II. A SYSTEMS APPROACH TO SELECTION

The general purpose of a personnel selection program is to seek and then identify those individuals with the proper combinations of abilities, interests, and motives who will learn the new job skills most quickly and who will reach a high level of proficiency on the job In addition, the employer may be interested in selecting individuals who will stay on the job longer, will have fewer accidents, or will be promotable to higher-level jobs In any case, the problem becomes one of predicting which individuals will eventually succeed in a given job

The determination of what constitutes "success" on the job necessarily hinges upon employer judgments regarding the relative importance of the various characteristics which are pertinent to job performance However, the ideal combination of these traits rarely exists in one individual Consequently, the selection process consists in choosing those persons possessing a combination of characteristics leading to the greatest likelihood of success

In some situations, all of the pertinent characteristics may not be compatible For example, an attempt at selecting to insure rapidity of learning may result in the hiring of a caliber of trainees who very shortly master the job duties, but who become disinterested in the work because of its lack of challenge or lack of opportunity for advancement This, in turn, could lead to worker dissatisfaction, poorer performance and a resultant increase in personnel turnover Thus, unsuitable persons may result from overselection with regard to some trait(s), as well as underselection or no selection at all

The scope of a selection program, as treated in this discussion and later in this study, extends before and beyond the point in time where the actual screening of certain individuals from among the field of applicants is effected In addition, because of varying employer needs, as well as a certain amount of unavoidable attrition, the systems approach to personnel selection consists of four distinguishable phases (1) recruiting, (2) screening, (3) training evaluation, and (4) job performance evaluation

These constitute chronological steps in the employment process, however, they also represent an ordering of the importance or desirability for eliminating unsuitable applicants That is, the earlier in the sequence an unsuitable candidate is eliminated, the more effective is the selection process Specifically, the more successful the recruiting phase is in obtaining applicants with a high probability of satisfactory job performance, the less extensive must be the screening, training, and job evaluation phase Obviously those workers who are finally judged as unsuitable as a result of job

performance ratings represent a much greater loss than similar applicants who were eliminated before the hiring or screening or training phases Thus, selection is a multiple filtering process wherein each successive filter (phase) results in finer (and more costly) discrimination The selection process and its relative relationship to cost is graphically illustrated in Figure 1 below

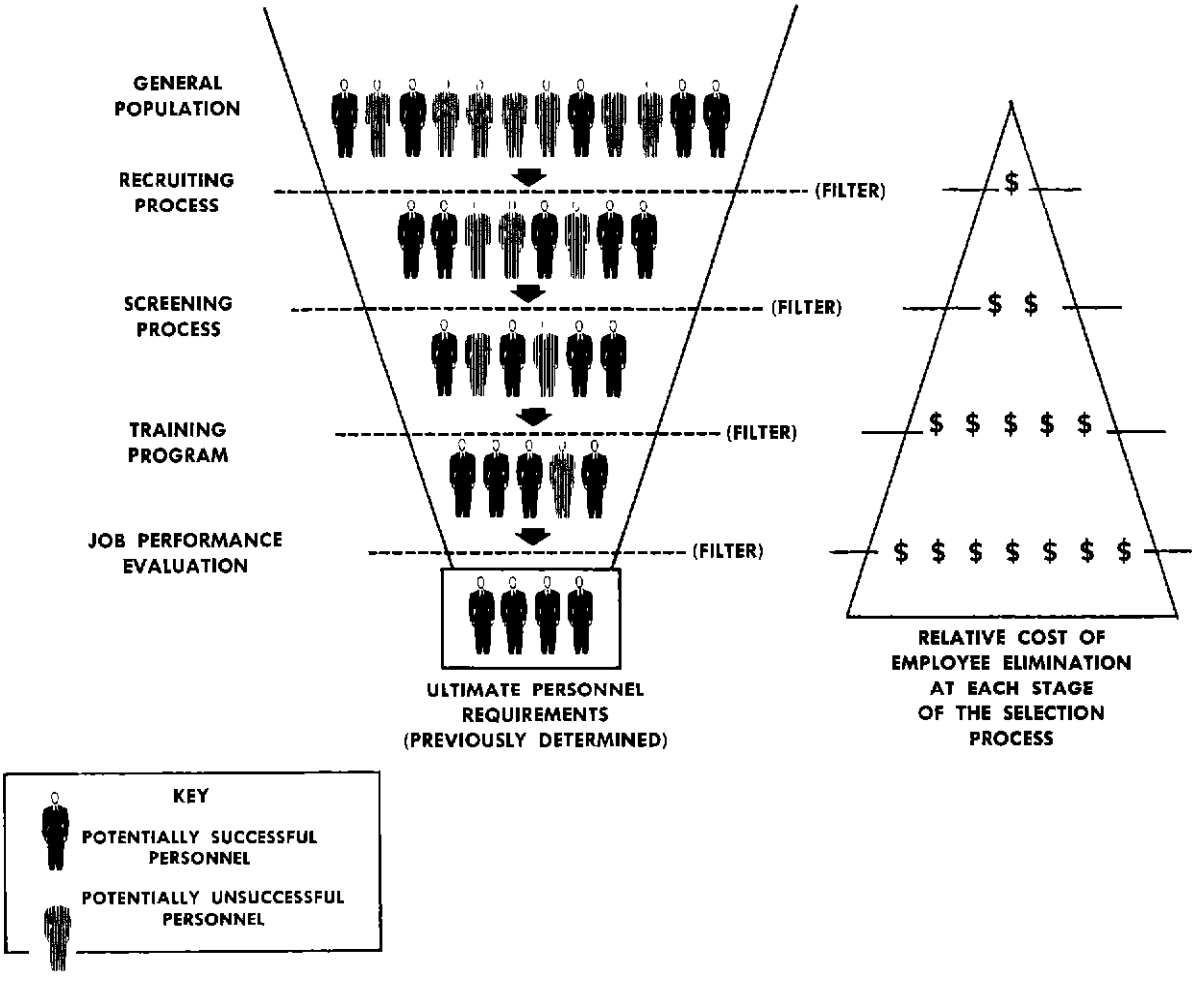


Figure 1 Selection as a Successive Filtering Process

Each of the four selective phases will be treated separately in sections that follow

A Recruiting

A selection program actually begins with recruiting The objective of recruiting is to attract applicants with specific interest in and qualifications for a job Ideally, such a program would provide only candidates who were destined to be successful on the job immediately, or more generally, after a suitable period of training Similarly, it would tend to increase the effectiveness of the training program by providing trainees who were well suited to learning the job skills and duties

Practically, a recruiting program should be directed toward that segment of the population most likely to have the necessary job qualifications An example of this appeal in a recruiting campaign might hinge around the age of potential employees Thus, if it were established that age was negatively related to job performance (i e , if younger persons tended to be more successful than older ones), then initial recruiting efforts could be oriented toward that age group with the greatest probability of success This could be done by means of advertisements placed where they were most likely to be seen by the most desirable candidates, such as schools, colleges, etc Included in these announcements would be such directly determinable requirements as age, education level and physical characteristics

Having appealed to that segment of the population having the greatest probability of job success, the employer would then begin the process of screening the interested applicants on the basis of qualifications not readily determinable by the recruiting process

B Screening

An effective screening process, which would be compatible with present Civil Service practices, would be subdivided into three parts They are review of the application blank, aptitude testing, and orientation interviews

1 The Application Blank The application blank is considered a preliminary screening device If a large number of applicants are being considered at one time, the examination of detailed application blanks becomes an involved task It is essential that those who show little or no promise be eliminated from the group, so that time and expense involved in subsequent selection steps be reduced However, it is difficult to classify applicants when the details of the application form are so numerous that they confuse the picture In such a situation, it may be well to determine in advance five or six definitive characteristics for use as a basis of preliminary screening Thus, if certain physical characteristics are necessary to the job, or applicants of certain ages are desirable, these may be used as criteria A checklist of relevant items can be prepared and each application checked as satisfactory, unsatisfactory, or marginally acceptable on each item The results of this technique can then be used for ordering applicants on a preferential basis Such a technique would appreciably increase the accuracy of selection and reduce the time necessary to process the blanks

2 Aptitude Testing The second phase of screening--aptitude testing--constitutes the most significant portion of the screening process The basic premise here is that tests which are found by research to be related to job performance can improve the prediction of job performance Most jobs require a number of specific aptitudes for successful performance A lack of such aptitudes can be predicted with a reasonably high degree of success using the appropriate tests However, although persons having superior aptitudes can be identified through the use of such tests, success on the job is affected by a multitude of factors which may be totally unrelated to the aptitude(s) measured by tests For example, such factors as motivation, domestic difficulties, personality conflicts with superiors or associates, personal illness, and other emotional stresses may have an important effect on an employee's adjustment to the job Some of these non-aptitude factors may be uncovered in an interview, but others may not Thus, testing will sometimes indicate selection of an individual who does not turn out well on the job, similarly, some applicants who might become good employees will be screened out by tests (But this same statement applies to the interview or any other device for selecting employees) Applicants who failed to make a passing score on the tests would be rejected at this point in the selection process, and those passing would be notified to appear at a designated time for an orientation interview

3 Orientation Interviews The orientation interview is an important phase of the screening process Although it can be used only in a limited

way as a screening technique in the lower job classifications of the federal employment structure, it can prove invaluable in effecting self-selection as a result of the exchange of information between employer and applicant More specifically, the interviews would serve (1) to provide applicants with better insight as to the nature of the job, and (2) to provide the employer with some insight as to the interest, general ability level, and personality characteristics of the applicant

In many cases the applicant himself, if given sufficient information regarding the nature of the job, is in a better position to judge the suitability of his own interests, aptitudes and aspirations as they pertain to the job For example, if the candidate's personal data indicated heavy family responsibilities, the permanence or steadiness of the work might be an important consideration It is the interviewer's responsibility in such a situation to inform the applicant that the need for simulator operators fluctuates and that it is not possible to guarantee continuous employment Given such information about the job and its requirements, the applicant may prove an effective self-selector

Thus the interview can become an important step in the selection process if the candidate receives, as well as gives information It is one of the few pre-hiring situations in which the applicant's motivation to succeed can be put to effective use By encouraging the candidate's active participation in the screening interview, the interviewer can evaluate as well as provide an opportunity for self-evaluation of those motivational aspects of

the individual which are so important, but which are so difficult to measure objectively Unless he voluntarily withdrew from consideration, or some disqualifying factor were uncovered, the applicant would be hired following this three-phase screening process

C The Training Program as a Selection Technique

If we assume a fairly high correlation between employee performance in the training situation and employee performance later on the job, the training program represents a situation which will provide a reasonably accurate indication of an employee's future job success In addition, instructors conducting a training program, of necessity, have a thorough and accurate knowledge of the job requirements, as well as an adequate opportunity for observing trainee performance In view of this, instructor evaluations should provide a refinement in selection over at least some of the earlier methods Thus, in this technique, the trainee is selected or rejected on the basis of achievement test grades and/or instructor evaluations with regard to job performance potential

D Job Performance Evaluation

Selection at this stage in the employment process is the most accurate, but also the most expensive That is, the ultimate suitability of the employee is more nearly determinable, but the elimination of personnel at this stage, after they have been through extensive screening and training, is done at a great cost in terms of money and effort A major deter-

minant of the effectiveness of this phase in employee selection lies in the availability of reliable and valid performance evaluations Frequently the lack of objective measures of performance, the shortcomings inherent in subjective rating scales, and the lack of opportunity for adequate observation by supervisory personnel makes the obtaining of effective employee evaluations a major problem

E General Comments

The above discussion was presented to establish a frame of reference and a rationale for the work done in the present investigation and for the recommendations which follow from it

The following statements summarize the treatment of selection as it was applied in the present study.

1 Personnel selection is considered a continuous process both because of fluctuating employer needs and because of a certain amount of unavoidable attrition

2 The selection procedure is divided into four major phases. recruiting, screening, training, and job performance evaluation These phases also represent an ordering as to. (a) occurrence in the employment process, (b) increasing accuracy, and (c) increasing expense In view of these considerations, the objective of this selection study is to make possible the elimination of potentially unsuitable employees at the earliest possible time

3 The selection process represents a trade-off situation, in which there may be incompatible objectives, i e , overselectivity with respect to one characteristic may lead to undesirable results in others Thus, effective selection is dependent upon a determination of an optimum combination of job qualifications by the employer

III. THE BASIC PERSONNEL DECISION

Research can provide objective answers to many questions, but often research poses questions as well Rarely does research yield a single alternative, rather it provides the human decision-maker with a variety of alternatives with a probable set of consequences for each Ultimately the administrator must give subjective values to these consequences and select a course of action This study is no exception It can indicate what is likely to happen if applicants with certain characteristics are selected to be simulator operators, but the ultimate decision as to what constitutes a satisfactory operator can only be made by entering subjective weights into the prediction equation, the entering of these subjective weights is an administrative action which can only be made by the professional staff of the FAA's simulation facility

Nevertheless, the writers of this report feel an obligation to at least call attention to some of the factors which might influence decisions based on this report There are two classes of factors which need to be considered.

those extrinsic to the study which depend upon the importance of simulation in the research program of the Federal Aviation Agency, and those intrinsic characteristics of the study which effect the degree of confidence which should be placed in the results

A Extrinsic Factors

The problem of overselection has already been described As pointed out, the consequences of selecting personnel with qualifications greater than those required to do the job usually are manifested in boredom, high turnover, and an increasing apathy which results in deteriorating job performance and morale Sometimes, however, the very nature of the job to be performed demands overselection Where the consequences of poor performance are serious enough to warrant it, overselection is a means of obtaining at least short-term benefits by decreasing training time and temporarily increasing the accuracy of performance It is believed that the simulator operator's job is such a task

The importance of the simulator "pilot's" role in ATC simulation was emphasized previously in this section Since the Federal Aviation Agency is using this tool as a means of gathering evidence which may effect national aviation policy, the confidence placed in the findings of simulation must be fully justified A multi-million dollar program of airways modernization is under way Many of the equipments of this evolving system will be tested in the simulated environment at the National Aviation Facilities Experimental Center It would be false economy to jeopardize the validity of this evaluation through the inadequate performance of a few simulator operators

The basic personnel decision faced by FAA in staffing the air traffic control simulator systems is whether or not to risk the potential long-term disadvantages of overselection in order to obtain maximum short-term performance In view of the critical role of simulation in FAA policy determination, it is our belief that only one decision is possible--select the best possible personnel within the already stringent limits of a GS-2 job classification

There is another extrinsic factor which also provides supporting evidence for this recommendation This study was done in its entirety on the relatively simple operator's consoles of the TDC optical simulator The simulator consoles being manufactured by Aircraft Armaments, Inc are not only more complex, but place a premium on the operator's participation in the recording of data during a simulation run Thus, the operator's role will become even more important in the evaluation of future ATC systems using these more sophisticated simulation devices

B Intrinsic Factors

Two intrinsic limitations of this research should also be considered in making the basic personnel decision

1 The study suggests long-term deterioration of performance as a result of overselection, but this cannot be considered conclusive evidence that such an effect will be duplicated in the NAFEC environment

2 The correlations obtained in this study are probably not true indicators of the validities of the various selection instruments studied, rather these instruments are probably considerably better than have been demonstrated here

A number of differences in the environment in which the study was carried out may have produced the apparent long-term loss of prediction effectiveness, which was found at Indianapolis Because of the nature of the labor market in the Atlantic City area, such a long-term effect may not occur at NAFEC Only a follow-up study at some future time would determine the applicability of the findings at TDC to this new environment Obviously, then, limited emphasis should be given to this possibility If the deterioration in performance or high turnover does not occur, then FAA has gained effective performance at no additional cost If the effect does appear, it can be combated by motivational devices such as upgrading, providing a titular promotional structure, or even by using transfer to other positions within the research activities at NAFEC Even at the worst it can be coped with by maintaining a large pool of candidates and by supporting a continuous training program of replacement operators

At first glance, the validity coefficients reported in this study for the Civil Service Examination and the supplementary test battery do not appear high However, this study was done under conditions which made it impossible to obtain an indication of the true effectiveness of these instruments as predictors of job success A glance at Figure 1 on page 4 will show that

selection assumes that all of the population of applicants is available to be studied However, in this study, only successful candidates on the Civil Service Examination were allowed to continue through the other stages of the selection process The task of identifying the potentially successful operator has thus been made much more difficult since the population has been shrunken or "truncated" by the elimination of part of the candidates Our interpretation of the results of this study is, therefore, always on the conservative side and the actual effectiveness of the screening instruments is probably considerably higher than the values reported

In the recommendations and implications which follow, we have tried to place ourselves in the position of the administrator of FAA's simulation programs Each of the recommendations is predicated on the basic personnel decision to obtain and train the best possible simulator operators--a decision which we believe is dictated by the importance of simulation as a tool in air traffic control research

IV. IMPLICATIONS OF THE RESULTS

A Implications for Recruiting

The results of this study indicated a tendency for younger operators to perform better than older ones

In addition, during the course of this contract task, the investigators had frequent and extensive contact with an operating simulation system, as well as numerous interviews and informal conversations with simulation

instructors, supervisors and pilot operators As a result of this experience, it became quite apparent that a general interest in aviation and a desire to do the job well are essential for success as an operator

In view of these observations, it is recommended that the appeal of the Civil Service announcement be slanted, and the advertising of it be emphasized, such that the greatest likelihood of attracting the best-suited applicants is realized

Typical nearby sites and activities which may provide such a source of suitable operators follows

1 Local area schools and colleges

2 Orientation of high school groups by participation in Vocational Guidance dissemination programs

3 Flying clubs and Civil Air Patrol centers

4 Youth centers

5 Libraries

6 A periodic "open house" program at NAFEC for local residents, including tours of the simulation facilities

Thus, by intensifying the recruiting program, a larger and more suitable population of applicants would be made available to choose from

B Implications for Screening

There are three general implications pertaining to screening

First, more extensive and systematic use could be made of the application blank (Standard Form 57) as a preliminary screening device A

systematic approach to evaluating application blank information such as that discussed earlier is recommended Some of the items which should be considered are

1 Date of birth (age)

2 Height

3 Previous Work Experience--A review should be made of previous work experience for relevance to simulator operator tasks Considerable interpretation of the work history may be necessary, in order to determine the possibility of transfer of skills to the task of operating a target generator Such jobs as switchboard operators, taxi cab dispatchers, or laboratory technicians might be sufficiently related to provide an experience advantage The skills required to learn and perform such previous jobs might coincide with those skills outlined in the Simulator Operator's Job Specifications (Appendix A) A history of successful performance of previous jobs involving manipulative skills, verbal fluency, and response accuracy would be relevant experience which could be of value in this phase of the screening process

4 Education Level--Because of a trend toward more widespread education in our society, non-completion of high school may be a more significant indicator of an unsuitable applicant now than in the past, notwithstanding the possibility of extenuating circumstances in certain cases In the case of students with some college training, those with a science-oriented background may be particularly desirable

Second, continued use of the Civil Service Examination as the primary selection instrument is recommended Raising of the cutting score on this battery is a possibility which should be investigated further The effect of such an increase in the minimum qualifying score cannot be determined from the data of this study In the absence of such support, it may be desirable to make possible finer screening at later stages in the selection process by providing a large pool of candidates from which to choose This can be accomplished by maintaining an open register and by increased emphasis on recruiting If a surplus of applicants has been provided as a result of intensive recruiting, more selectivity can be exercised by choosing the highest scorers from the group who achieved at least the minimum score in order to fulfill the job requirements

Third, the orientation interview should be adopted as a part of the selection process Because this step represents a learning process for the potential operator, it should be conducted by simulation instructors or supervisors who have a thorough and accurate knowledge of the job requirements to impart to the applicant These interviewers are not likely to be experienced in interviewing techniques, therefore, the situation should be structured for them, by providing an interview guide such as that presented in Appendix D of this report Such a method should lead to greater uniformity in the information gained as well as that disseminated

C Implications for the Training Program

In the NAFEC simulation situation, no operator trainees were eliminated as a result of their performance during the training program However,

a statistical comparison of the training grades with later job performance evaluations indicates an accuracy of prediction of later job success equivalent to that provided by the Civil Service Examination It is therefore recommended that trainee performance in the training program (as measured by the final grade at the completion of the training program) be used as a basis for retaining or eliminating operator candidates A decision regarding the cut-off score should be reached after more details surrounding the future training program are made available The instructors for the operator training program conducted at NAFEC during the summer months of 1959 should, because of their experience with that program, be consulted in the determination of this cut-off score

D Implications for Job Performance Evaluation

Since, in this phase of selection, the ultimate suitability of the operator is most nearly determinable, performance evaluations can be used as a basis for the final elimination of unsuitable operators Although, as pointed out earlier, this is the least desirable point in the selection process to accomplish the elimination of unsatisfactory operators, job performance evaluation serves other important functions It offers a fair and objective approach to the inevitable selection process required for decisions regarding employee upgrading, designation for supervisory training, or determining the operator's capability of performing satisfactorily on other more complex equipments or in more responsible positions The fluid nature of the demand for simulator operators at NAFEC also makes job performance evaluation a very important

stage in the selection process In periods of reduced demand, such evaluations can be used to reward effective operator performance and simultaneously provide the best possible team of operators through further differentiation between those individuals who are merely satisfactory and those who are superior in their performance of the simulator "pilot's" task. The instruments used for measuring job performance in the present study proved to be reliable and valid indices, and therefore should be considered for implementation in the future Since the stanine ranking scale proved as effective as the eight-trait rating scale, and was more efficient, it is recommended for future use in obtaining job performance evaluations

Because of the fact that rater opportunity for observing the operators proved to be an important factor contributing to the reliability of ratings, several implications emerge regarding the use of the rating scales First, the supervision and on-the-job training of new operators should be more definitely structured That is, supervisors should be advised in advance of their forthcoming evaluation duties, to orient them toward an "evaluation-consciousness" Secondly, supervisors and instructors reported evidence of considerable learning on the job after completion of the training program Therefore, in order to permit the adequate cultivation of job skills, as well as to insure sufficient time for on-the-job observation, the ratings should not be obtained until two or three months after completion of the training program

In order to facilitate the identification of operators by raters, a labeled group photograph should be used in conjunction with the rating scale In addition, it is recommended that use be made of peer ratings in situations where there is an opportunity among the group to be rated, to observe the characteristic (s) to be rated

Again, determination of the cut-off point for retention or elimination should be deferred until more details are available concerning the number of personnel needed and those available

SECTION TWO

INTRODUCTION

A key to the effectiveness of dynamic air traffic control simulation lies in the performance of the simulator operators. They comprise the largest portion of the human element in the simulation process, hence, the development of effective operator selection and training programs is a vitally important goal. The critical importance of adequate personnel staffing was recognized early by the Federal Aviation Agency, and provision was made for the development of a simulator operator selection program as a part of Contract FAA/BRD-27 with Courtney and Company

I. HISTORICAL DEVELOPMENT OF THE PROBLEM

The original intent of the contract task under which the present research was conducted was to assist in the development of a selection and training program for staffing a proposed dynamic air traffic control simulation system This system, which was being built for FAA largely by Aircraft Armaments, Inc , Cockeysville, Maryland, consists of 50 radar target simulators (and related support and data recording equipments). When properly staffed with "pilot" operators this system provides the simulated outside-world portion of the air traffic controllers' environment

This simulator system was scheduled to begin operation early in 1960, therefore, the selection program was planned for implementation several

months prior to that date However, because of delays in the development and production of this and related equipment, and a change in the immediate needs of the FAA, the following factors effected a redirection of the work aimed at developing a selection system

First, the discontinuance of the Civil Aeronautics Administration and the Airways Modernization Board, and formation of the Federal Aviation Agency, resulted in the moving of an operational dynamic air traffic control simulation system from the Civil Aeronautics Administration's Technical Development Center (TDC), at Indianapolis, to the National Aviation Facilities Experimental Center (NAFEC), at Atlantic City. Secondly, this move created an immediate need for staffing the newly-located TDC simulator, and at the same time, provided an opportunity for an investigation of the problems involved in simulator operator selection.

This staffing need was foreseen sufficiently in advance of the actual transfer of the system to permit a preliminary investigation of the problem while it was still operating at TDC

This investigation served a number of useful purposes. It provided

1. A better insight into the nature of the simulator system and the problems involved in staffing it,
2. The opportunity to carry out an activity analysis of the simulator "pilot's" job in an operational setting, the activity analysis in turn, provided a matrix of job-related skills from which selection instruments might be chosen,

3 A group of experienced simulation supervisors who were helpful in the development and evaluation of the operator performance evaluation measures, and

4 A situation for testing the validity of the currently used Civil Service Examination and other variables potentially useful as predictors of operator success

With this background, Courtney and Company personnel were able to make more significant contributions to the actual selection and training of operators for the TDC system in its new setting at NAFEC.

Thus, the investigation reported herein is concerned with an analysis of data obtained from the TDC simulator equipment, its simulator operators, and the simulation supervisors in two operational settings (1) a preliminary investigation of the operating system while it was at TDC, Indianapolis, and (2) a more extensive investigation of the same system after it was transferred to NAFEC, Atlantic City.

II. DYNAMIC SIMULATION IN AIR TRAFFIC CONTROL RESEARCH

In recent years, the use of dynamic simulation techniques for the study of air traffic control problems has become quite extensive Evidence indicates (1, 2, 3, 4, 5, 10, 15) that application of these techniques already has led to improvements in the ATC system. These efforts have also generated a number of new concepts which are expected to have application in future systems. Both the magnitude of funds appropriated, and the extent

of research effort--past, present, and future-- by the Federal Aviation Agency, are indicative of the importance attached to dynamic simulation as a research and development tool.

An extensive, long-range program for improving the nation's air traffic control system was initiated by the Airways Modernization Board, and is currently being carried on by the Federal Aviation Agency The Bureau of Research and Development of FAA (at NAFEC) currently has one simulation system in operation, and has three others in various stages of completion Each of these systems requires approximately 50 trained simulator "pilots" for normal operations. Simulation techniques will be extensively used in order to bring out the strong points of the air traffic control system and delineate its weaknesses for modification purposes or further development. In view of this, and of the continuing FAA needs for simulation studies, the development of effective operator selection and training programs is considered a vitally important goal.

III. GENERAL DESCRIPTION OF AIR TRAFFIC CONTROL SIMULATION SYSTEMS

It seems appropriate at this point to include a brief description of the simulators for those readers not acquainted with these systems.

First, the human element is retained in the controllers' and "pilot's" positions and the total flow of information in any segment of the simulated system duplicates the flow in the actual system. In one large room, the movement of individual aircraft is simulated by radar target generators (See Figure 2.) Each of these target generators is manned by an operator ("pilot"). (See Figure 3 and 4)

Courtesy of FAA

Figure 2. The Radar Target Generators (Aircraft Armament Simulator)

Courtesy of FAA

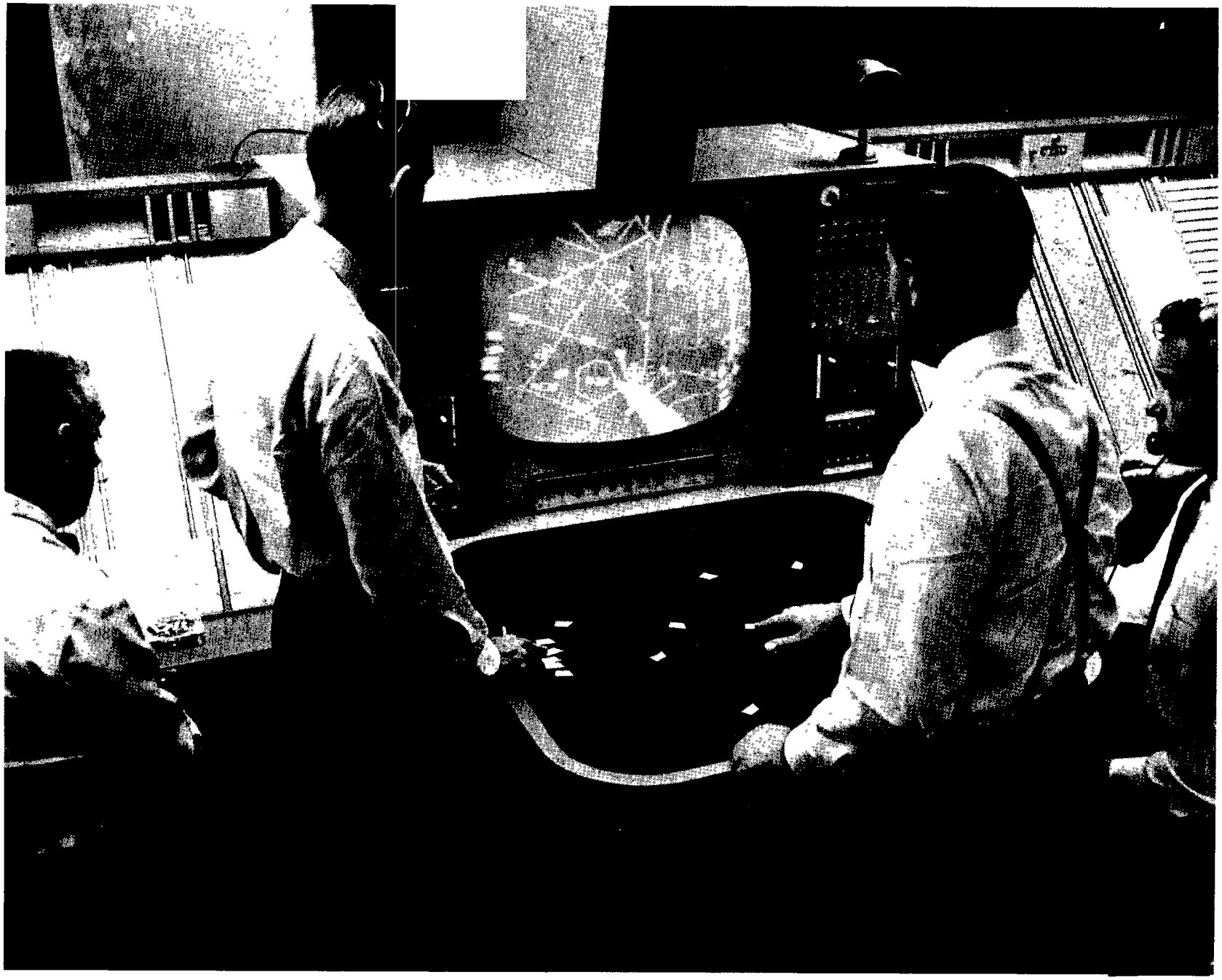
Figure 3. Simulator "Pilots" at the Target Generator Consoles (TDC Simulator)

Courtesy of FAA

Figure 4 Simulator "Pilots" at the Target Generator Consoles (Aircraft Armaments Simulator)

By manipulating switches, knobs and other controls on his (her) console, each operator "flies" a pre-arranged flight plan, as modified by controllers' instructions given by "radio " As with real aircraft, the "pilot" can vary speed, altitude and direction of flight An additional and equally important operator duty centers around the recording of certain aspects of each flight for subsequent problem analysis This is done by each operator, both through the actuation of appropriate data recording switches on his console, and by the keeping of a tabular-form log

In the simulated ATC facility, skilled air traffic controllers are provided consoles and associated displays closely approximating the real-world working environment Outputs from the "piloted" target generators appear to controllers in a separate room as aircraft moving on a radar display (See Figure 5)



Courtesy of FAA

Figure 5 Controllers at their Consoles in the Simulated ATC Facility

Simulation of the air/ground and ground/ground communications network is accomplished by a multi channel interphone system Data collection and reduction equipment is also an integral part of the over-all system

Thus, by using traffic samples of predetermined characteristics, innumerable traffic situations can be created A simulation problem may use one a few, or all of the target generators to produce targets on the controllers' displays, each appearing at predetermined places and times of entry into the control area under study Each "pilot" operator then manipulates the controls on one target generator console and periodically reports the progress of his "flight" to the controllers in accordance with his clearance and subsequent amendments

The validity of the results and conclusions derived from this kind of laboratory experimentation is obviously greatly affected by the simulator operator The degree of realism provided by him in simulating both aircraft performance and pilot behavior--as they apply to the interaction with controllers--is crucial Therefore, the policies and procedures governing the recruiting, screening and hiring of trainees to assume the role of pilots in these simulation studies is equally crucial The aptitude of those trainees selected then becomes an important influence upon the content and pace of the subsequent training required to bring them to an acceptable level of competence

IV. ACTIVITY ANALYSES

The first step in any program for selecting personnel for a specific job specialty should be a comprehensive analysis of that job It is essential to determine both the specific activities which are carried out and the conditions under which they are carried out Thus, before we can hope to develop effective tests for the selection of simulator operators, we must have precise and specific knowledge about the operators' duties and the circumstances surrounding their accomplishment Such detailed information was obtained by means of systematic activity analyses performed on two of the dynamic simulator systems. These analyses necessarily constituted a prelude to the main part of the present study, since they provided a basis for answering the fundamental question regarding the degree of similarity between the two systems under study. Stated another way, in so far as the two jobs are essentially similar, the results and conclusions derived from the TDC system selection studies would logically be applicable to the Aircraft Armaments system.

A The Aircraft Armaments Simulator System

In the case of the Aircraft Armaments simulator, the activity analysis was more difficult than usual because the system was not yet in an operational status. The analysis was therefore conducted primarily on the engineering model of the target generator during acceptance tests of the equipment Under these conditions, typical simulator problems were developed by the

investigators, then "flown" by them as simulator pilots to determine the responses and activities involved. An account of this investigation and the recommendations resulting therefrom, were presented in a previous report of this series (6).

This equipment is more complex and considerably more flexible as a research tool than its predecessors, but was still in its development stages when this study was undertaken Consequently, it could not be studied under operational conditions. However, every effort was made to reproduce the total working situation in the "flights" to determine specific job activities.

B The TDC Simulator System

This system, which offers the capability of simulating up to 42 aircraft at a time, has many characteristics in common with the system under development As mentioned earlier,the TDC simulator was in operation at Indianapolis at the time this study was begun. This situation offered an opportunity for a preliminary investigation in an operational setting However, such an investigation would be of value only if the two systems were sufficiently similar to permit application of results found on the operating system to the proposed system To resolve this question of similarity, an activity analysis of the operator's task on the TDC simulator was carried out under operational conditions while this system was located at Indianapolis.

C Job Specifications

Activity analysis data from the two simulator systems, together with interview data from simulator operators and operator supervisors at TDC, defined a sizeable list of abilities and traits required to perform the tasks involved in the two systems. A critical examination of these results seemed to indicate that the similarity between tasks was sufficient to justify the extrapolation of experimental results to the developmental Aircrafts Armaments system. Appendix A presents a composite list of the traits and abilities hypothesized as pertinent to the job of simulator pilot operator for both simulators. This list constitutes the Job Specifications, and provides a logical basis for developing or choosing selection instruments.

V SELECTION INSTRUMENTS AND CRITERION MEASURES

A Selection Instruments

The test used for selecting operators of the target generators at TDC, and at NAFEC, was a Civil Service Examination battery consisting of tests #105A, #105B, and #96. These are paper and pencil tests designed to measure aptitude for learning, and ability to follow oral instructions. However, the validity of this battery as a predictor of success as a simulator operator had not been determined, and its value as a selection device had been questioned by simulation supervisors at the Technical Development Center. As mentioned earlier, one of the purposes of this study was to carry out a validation of this Civil Service battery.

In addition, on the premise that this selection test was less than adequate, a multi-aptitude battery of tests, the Employee Aptitude Survey, (9) was selected for experimental administration to, first, the experienced operators at TDC, and second, to the relatively inexperienced operators at NAFEC This was done for the purpose of isolating testable traits which were identified with effective performance as simulator pilots. These traits would provide a basis for future development of a valid selection device to supplant or augment the Civil Service Examination

B. Criterion Measures

Having determined, for purposes of this study, which predictors were to be validated, the next step was to obtain a criterion of success as a simulator operator No direct, objective measures of job performance were available

The task of the simulator operator is such that the investigators concluded that intermediate measures of job performance could best be derived from ratings of the observed performance of the simulator operators The question of whether the method of evaluation should be based on subjective rating of elements of this performance (molecular evaluation), or on subjective ratings of the performance as a whole (molar evaluation), was settled by a compromise.

This compromise consisted of (1) an eight-trait rating scale, which measures various aspects of job performance and attitude (molecular evaluation), and (2) an over-all ranking scale (molar evaluation). Hence,

operator evaluations were obtained from supervisors on both scales Further explanation of the development and use of these instruments is presented in Section Three

VI. STATEMENT OF THE PROBLEM

The initial concern of this study was the development and evaluation of selection measures for staffing the newly-located Technical Development Center target generators The secondary concern of the study was to develop recommendations for staffing similar simulator systems

The specific elements or problems which developed out of consideration for the staffing of the simulator systems were

1. A determination of the similarities in operator job requirements between two currently-complete or nearly-complete dynamic simulators, namely, the Technical Development Center target generator, and the Aircraft Armaments target generator,
2. A testing of the validity of the current instrument (the Civil Service Examination) for selecting simulator operators,
3 Development or selection of other predictor variables based upon job requirements derived from the results of the activity analyses,
4 Development of reliable criterion measures to make possible an evaluation of the predictor variables,

5. A testing of the validity of the predictor variables in two operational settings of the TDC target generators,

   (a) At TDC, with experienced operators, and

   (b) At NAFEC, with newly-hired operators,

6 A comparison of results obtained on the predictor variables, criterion measures, and validity coefficients of the two different operational settings,

7. A comparison of the use of training grades and job performance measure as the criterion measure at NAFEC, and,

8 Development of optimum operator performance prediction techniques, based upon the NAFEC results.

SECTION THREE

METHODS AND PROCEDURES

The methods and procedures used in this study will be presented in this section in their order of development The first sub-section deals with a detailed description of the selection measures The second deals with the development of the criterion measures The third represents a description of the samples of operators being considered, and the procedures used in getting the measurements on the samples The last sub-section deals with the statistical methods for treating the data

I SELECTION INSTRUMENTS

A Civil Service Examination

The primary instrument used for selecting ATC simulator operators is a Civil Service Examination battery, consisting of tests #105A, #105B and #96 These are paper and pencil tests designed to measure aptitude for learning and ability to follow oral instructions Competitors are rated on a basis of 100 points, and must attain a composite rating of 70 to qualify for the register This composite is made up of weighted subscores from the three tests, and, in the case of certain military veterans, includes the test scores plus a veteran's preference bonus of 5 or 10 points (In this latter case, a maximum score of 105 or 110 is possible)

As indicated in the previous section, one of the factors leading to the present study was supervisor dissatisfaction with previous operator selection resulting from use of this examination at TDC, Indianapolis, and a desire to secure quantitative data regarding its suitability as a predictor of job performance at NAFEC

B Employee Aptitude Survey

This study also attempted to determine which measurable aptitudes should be included in a successful prediction instrument A Job Specifications list (Appendix A), which was derived as a result of the simulator operator activity analyses on two simulator systems, and interviews with TDC personnel, served as the basis for selecting factors which seemed amenable to measurement by paper and pencil testing procedures

A critical review of currently available multi-aptitude batteries led the investigators to select the Employee Aptitude Survey as most nearly satisfying the requirements of the situation. The battery is comprised of nine 5-minute tests and one 10-minute test The tests are entitled Verbal Comprehension, Numerical Ability, Visual Pursuit, Visual Speed and Accuracy, Space Visualization, Numerical Reasoning, Verbal Reasoning, Word Fluency, Manual Speed and Accuracy, and Symbolic Reasoning.

Aside from its content value, the Employee Aptitude Survey has the following convenient features which make it particularly suitable to the present experimental situation (1) each test is confined to a single 8-1/2 by 11 inch sheet, with complete and concise instructions and examples on one

side, and the test items themselves on the reverse side, (2) the time required is very short, and (3) scoring, in most cases possible by IBM Scoring Machine, is facilitated by stencils for right and wrong responses

The battery was reportedly "developed to provide measures of a diversity of special abilities required by a wide variety of specific jobs at all levels on the occupational ladder", thus, all the abilities measured by the various tests were not expected to be applicable to the present job situation, but the pertinent ones could be empirically selected and weighted to maximize prediction accuracy

II. DEVELOPMENT OF CRITERION MEASURES

In order to determine the effectiveness of the selection instruments, a measure of the proficiency of the simulator operators was necessary To meet this need, the investigators developed rating instruments to measure different facets of simulator operator performance

A Eight-Trait Rating Scale

The experienced supervisors at TDC could reasonably be expected to have developed considerable insight into those traits desired of successful operators Consequently, their opinions were carefully considered in the selection and definition of traits to be measured. An eight-trait rating scale was formulated, using a method similar to that of Wilson, Mackie, and Buckner (16). The traits selected were (1) Knowledge of the Job,

(2) Learning Potential, (3) Communicating Abilities, (4) Attitude toward Work, (5) Accuracy, (6) Adaptability, (7) Dependability, and (8) Group Compatibility In essence, the rating instruments provided a definitive description of each of these traits, with descriptive phrases characterizing intermediate values and extremes for each trait Thus, each trait scale was presented as a continuum rather than as a limited number of discrete categories A copy of the rating form is presented in Appendix B

B Stanine Ranking Scale

A stanine ranking scale was developed as a second type of criterion measure This instrument incorporated a number of concepts which were intended both to avoid some of the common shortcomings of rating scales, and to facilitate its effective use by untrained raters

First of all, the size of the operator samples being evaluated made a complete ranking impractical Therefore, a group-order, or forced distribution ranking method was chosen In this method the rater is told the number of cases that fall into each category of the ranking scale This not only facilitates the ordering of personnel, but, also eliminates the major types of errors made by raters. Since the rater, in conforming to the forced distribution, must put the prescribed number of persons into each of the categories, he cannot commit the leniency error of assigning most operators high ratings, or the central tendency error of clustering operator ratings around the mean of the total group

Secondly, the forced ranking scale was divided into nine categories, and frequencies for each category were assigned in accordance with the normal frequency distribution This constitutes a stanine or "standard nine" scale, as presented in Garrett (8). It is simply a subdivision of the normal curve into nine intervals, in which each interval (category) has a "width" of one half a standard deviation, (except for the extreme classes which are open-ended). Figure 6 shows this graphically, with the percentages of people who would be expected to fall in each interval.

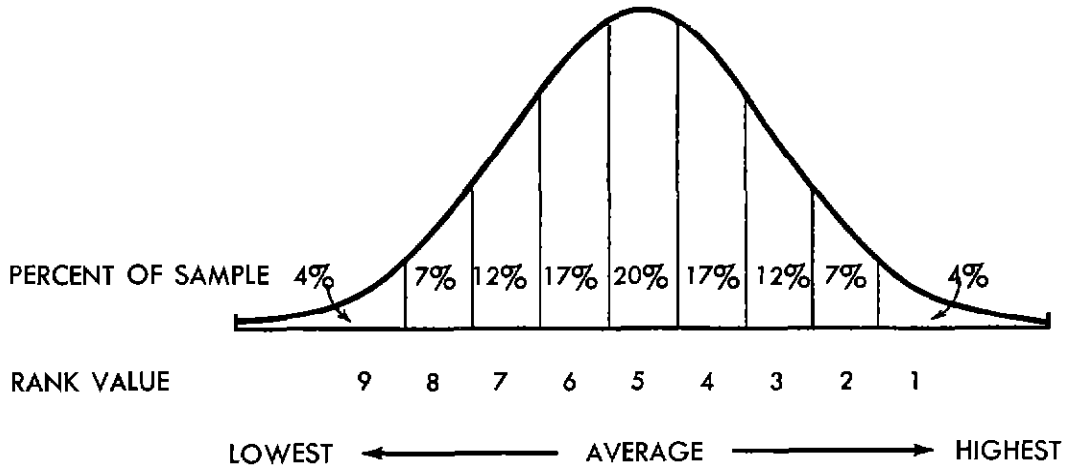


Figure 6. Per Cent of Sample and Corresponding Rank Value Produced by Superimposing Stanine Intervals on a Normal Distribution

The use of the stanine ranking distribution assumes that, on the average, the operators are equal in ability and that there is equal dispersion on either side of the mean It also assumes that the operator ability for each group is normally distributed The sample of job applicants in this case seemed of sufficient size to justify this assumption, and although the operators as a group seemed of above average ability, the investigators felt that the normal hypothesis could nevertheless be assumed.

Use of this ranking scale required first, that each rater check off (on a roster of operators) those whom he knew well enough to rank, and second, that the investigators use that number of operators checked, together with the percentages shown in Figure 6, as a basis for computing the number of operators to be assigned each of the rank values from 1 to 9 by the rater. For example, in a case where 50 operators were known, two operators would be assigned a rank of one, and two operators would receive a rank of nine The number of operators to be assigned these ranks was determined by multiplying the per cent of the sample corresponding to the rank value by the number of operators known and rounding the figure to the nearest unit By similarly using the percentages in Figure 6, the number of operators in each rank was determined and marked on the ranking form before returning it to the rater for the actual evaluation task. A copy of the ranking scale appears in Appendix C

III. DESCRIPTION OF OPERATORS, PROCEDURES, AND RATERS

A. The TDC Situation

Civil Service Examination scores on each of the 49 simulator operators in the TDC sample were obtained from the Civil Service Commission, but, since these operators had been hired at different times in the past, no information was available regarding the conditions surrounding their appointments.

Of the 49 operators in the TDC sample, 44 were women and 5 were men. Their mean age was 35 9 years, with a standard deviation of 10.6 years Their education level was 11 7 years, with a standard deviation of 1.5 years Their job experience ranged from 2 to 84 months with a mean age of 20 7 months and a standard deviation of 20 9 months.

The Employee Aptitude Survey was administered by Courtney and Company personnel to these operators during their last day on the job Their cooperation was requested for the purpose of "assisting in research to improve operator selection procedures for staffing another simulator," and obviously, the results could have no effect on their job status

These operators, unlike the NAFEC group, had had no formal training program to prepare them for the job. The orientation of new operators was done on an individual or small group basis by one of the simulation supervisors First, an operator would be given a familiarization lecture and tour, this was followed by a period of sitting at a console observing and listening to experienced operators during simulation runs Then, gradually the

beginning operator would be integrated into the job, with coaching and assistance by supervisors and other operators

Operator proficiency ratings were obtained from nine simulation supervisors using both of the rating scales These raters constituted a heterogeneous group from the standpoint of tenure in the position, extent to which they knew the operators, and opportunity for direct observation of them The evaluations were made after the simulation facility at TDC had been disbanded, and most of the raters had been transferred to other facilities, hence the independence of ratings can reasonably be assumed

B The NAFEC Situation

The Civil Service Examination was administered by the Civil Service Commission to 135 applicants for the simulator pilot operator positions at NAFEC Of this number tested, 77 were eligible for appointment to fill approximately 50 job openings (Appointments here, as at TDC, were at a GS-2 level, with an annual salary of $3255) A portion of the 77 eligible candidates were automatically eliminated from further consideration by virtue of their not appearing for operator orientation interviews These interviews were conducted by personnel who had been assigned the task of training those candidates finally hired Although the interviews were intended primarily (1) to familiarize candidates with the nature of the job, and (2) to provide the instructors with some insight as to the general ability level of their trainees, a certain amount of additional selection was effected

in that a number of the candidates declined appointment to the position following the interviews

A brief account of the interview technique used follows First, the instructors had had little or no prior interviewing experience Thus, in order to achieve a reasonable degree of uniformity in the information presented by the interviewers, and that obtained from applicants, an interview guide was developed Since this form served as an on-the-spot "frame of reference" for interviewers, the data obtained was not readily quantifiable and objective evaluation of its effectiveness was not possible However, interviewer opinion indicated that is was a moderately successful aid to conducting the interviews and evaluating the applicants A sample of this interview guide appears in Appendix D

Thus, as a result of these testing and interviewing techniques, 51 applicants (28 male and 23 female) were appointed as operator trainees Their mean age was 32 5 years, with a standard deviation of 9 5 years Their education level was 11 8 years, with a standard deviation of 1 4 years

The Employee Aptitude Survey was administered by Courtney and Company personnel to these newly hired operator trainees at NAFEC during their first day on the job, hence, the group tested with this battery consisted only of subjects who had been selected by the Civil Service Examination and the interviews The operator trainees were instructed that the testing was for research purposes and the results would have no influence on their job status Although it was also announced that the scores would not be revealed publicly, it seemed evident that the motivation of the group was high

The operators were given a five-week course of instruction consisting of lectures, classroom exercises and on-the-job orientation Weekly achievement tests were administered and the composite of the test grades formed a final training grade This grade was included in the analysis of data as a possible immediate criterion measure

In order to permit the operators time to become reasonably proficient on the job, and to enable the raters to observe adequately most of the operators, a two-month period elapsed after the completion of the training course, before operator evaluations were conducted using the eight-trait rating scale and the stanine ranking scale

The individuals who performed the ratings and rankings comprised three groups.

1 One group consisted of three instructors who had (a) taught the operators during the five-week training program, and (b) participated in on-the-job training

2 A second group consisted of eight operators who were generally regarded by their instructors as superior among the group and who were being considered for supervisory positions For this reason they will be referred to as "potential supervisors " Since these operators had the advantage of continuous and direct working contact with their peers, the investigators felt that they were in a position to provide adequate operator evaluations The ratings from this group were probably the most independent since their evaluations were conducted in a session supervised by the investigators

3 The final group of raters consisted of four air traffic control specialists who were conducting the simulation problems as well as doing some on-the-job training This group had had the least opportunity for observing the operators and were expected to provide less reliable ratings than the other raters

IV. STATISTICAL TECHNIQUES

In order to treat the results of the TDC and NAFEC operator samples most efficiently, a number of statistical techniques were necessary Most of the techniques used dealt with the study of relationships, i e , rater reliabilities, correlations among predictors and criterion measures, and the maximizing of prediction and comparison of predictors A description of the methods used in determining these relationships is presented below

A Reliability of Ratings

The reliability of the criterion measures, i e , the extent to which the different raters tended to agree in their ratings of the operators, was determined by means of an intraclass correlation coefficient This method, which was described by Ebel (7), is essentially an analysis of the variance of the individuals rated and of the within-persons error variance The

correlation is derived from the formula

$$r_{tt} = \frac{\tilde{\sigma}_p^2 - \tilde{\sigma}_e^2}{\tilde{\sigma}_p^2}$$

where

r_{tt} represents the average intercorrelation or the reliability for mean ratings from k raters,

$\tilde{\sigma}_p^2$ is the variance between persons, and

$\tilde{\sigma}_e^2$ is the within-person variance

Since all the raters did not rate every person, an estimate of the average number of raters per person was determined following a formula of Snedecor recommended by Ebel (7)

B. Trait Weighting

Since a statistical analysis would be greatly facilitated through the use of a single index or value for the criterion, some combination of the traits was desirable An attempt was made to weight the various criterion measures by the factor $\frac{1}{1-r_{tt}}$. However, the results indicated that the process did not produce an appreciable gain in the reliability of the composite criterion. (It only affected the reliabilities in the third decimal point or within the rounding error) Therefore, the criterion was formed simply by combining standard scores for the eight traits and the over-all rank

C Intercorrelation Matrices

Pearson product-correlation coefficients were used in determining the interrelationships among the various predictors and criterion measure This technique was also applied in obtaining the intercorrelations of the traits and ranking score.

D Multiple Correlation Coefficients

In order to determine the effectiveness of the combined instruments in the NAFEC study and to compare the Civil Service Examination with the Employee Aptitude Survey, the Wherry-Doolittle Test Selection Method was used (8) This method selects the variables analytically and adds them one at a time until a maximum shrunken multiple correlation is obtained

SECTION FOUR

RESULTS, DISCUSSION, AND CONCLUSIONS

In order to facilitate comparisons between the results of the two samples of simulator operators this section will completely treat the evaluation of criterion measures of both samples before going to the validation of prediction indices The reader is urged, however, to bear in mind the fact that the two samples were different in many respects, one of which was the operational setting, i e , the TDC sample measures were obtained at the close of operations, while the NAFEC sample measures were obtained in the early stages of operation of the newly located system

I EVALUATION OF CRITERION MEASURES

A Technical Development Center Data

The dynamic ATC simulator system at TDC provided an operational situation for testing the validity of the multiple factor aptitude battery and the criterion measures which had previously been selected It was felt that the advantages accruing from this preliminary investigation outweighed the disadvantages arising from the lack of similarity to the NAFEC setting

1 Reliability The intraclass correlation coefficients used in determining the reliability of the supervisors' ratings of operator effectiveness are presented in Table 1

Table 1

Reliability of Eight-Trait Ratings and Stanine Ranking Scores (TDC Sample)

Trait	r_{tt}	Trait	r_{tt}
Knowledge of the Job	81	Accuracy	79
Learning Potential	78	Adaptability	76
Communicating Abilities	76	Dependability	.81
Attitude toward Work	77	Group Compatibility	82
		Stanine Ranking	82

All coefficients are significant beyond 01 level

The reliabilities range from 76 for Communicating Abilities and Adaptability, to 82 for Group Compatibility and the stanine ranking, with a median coefficient of 79 Thus, there appears to be little difference in the relative effectiveness of the measures as far as reliability is concerned A point to be noted is that the stanine ranking scale is as reliable as any of the eight-trait ratings Thus, the stanine ranking scale would seem to be the measure to use if time and effort considerations tended to curtail criterion information collection and an over-all rating were desired

2 Intercorrelations of Criterion Measures In order to determine the relationship of the various indices of job performance, the trait scores were intercorrelated The results of the intercorrelations of the eight-trait ratings and the ranking score are presented in Table 2

Table 2

Intercorrelation of Eight-Trait Ratings and Stanine Ranking Scores (TDC Sample)

Indices	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Knowledge of the Job	84	81	24	82	86	63	22	86
(2) Learning Potential		88	10	90	.89	51	13	87
(3) Communicating Abilities			34	89	86	71	28	89
(4) Attitude toward Work				35	26	70	62	36
(5) Accuracy					85	68	29	92
(6) Adaptability						64	14	87
(7) Dependability							51	66
(8) Group Compatibility								33
(9) Stanine Ranking								

35 significant beyond 01 level

27 significant beyond 05 level

The results indicate that except for the correlations of Attitude toward Work and Group Compatibility with Knowledge of the Job, Learning Potential, and Adaptability, the intercorrelations were significant beyond the 05 level The intercorrelations between Knowledge of the Job, Learning Potential, Communicating Abilities, Accuracy, Adaptability, and the stanine ranking score were above 80 and, hence, were higher than the reliability coefficients for the same indices The high degree of intercorrelation between the various indices, except for Attitude toward Work and Group Compatibility, suggests that an overall "job proficiency" factor is being measured, possibly best represented in the stanine ranking, since the correlations of the traits on it are generally higher than on the other traits The fact that the traits intercorrelate so highly would tend to indicate that they are elements contributing to this over-all job performance If there is a second element being measured by the criteria it centers around Attitude toward Work and Group Compatibility traits and might be denoted as an "agreeableness" factor The fact that it generally does not correlate with the job proficiency aspects, except in the 30's on stanine ranking, indicates that it might not be important for successful operation of the console, and thus for worker efficiency

3 Choice of Criterion Measure In order to develop the most comprehensive index of job performance, the investigators tried to determine an optimum composite criterion measure To aid in the selection of the measure, use was made of the reliability of the indices as well as the intercorrelations among them

As previously mentioned, an attempt was made to weight the various traits according to their reliability, but since the change that occurred in the weighting procedure was only in the third decimal place, the approach was abandoned

A summed standard score approach, combining the ratings of all supervisors and all traits, was chosen for the criterion measure The reasons for the choice were that (1) the increased number of measurements going into the determination of the combined measure increased its reliability, and (2) the relative lack of independence among the traits indicated that a general proficiency factor was being measured Hence, the use of this over-all combined score seemed appropriate

The evaluation of the criterion measures at TDC indicates that

1 The measures of job performance are reliable, as shown in Table 1,

2 A high degree of interrelationship exists between the measures, indicating the presence of a general job proficiency measure,

3 The above findings supported the use of an over-all summed standard score as the criterion measure

B NAFEC Data

The need for staffing of the TDC simulator after its transfer to NAFEC provided an appropriate situation for testing the instruments for selection This transfer gave the investigators the opportunity to try out the criterion measures and predictor variables in a new operational situation where

experience on the job was constant, and lack of motivation on the part of the tested population was not a problem

1 Reliability of Indices The most essential step in the evaluation of a criterion measure is the determination of its reliability The intra-class correlation coefficient was again used to determine the reliabilities on each of the eight traits of the rating scale and the stanine ranking The reliability coefficients for the three groups of raters are presented in Table 3 The average number (k)[1] of ratings upon which the coefficients are based is also presented

As indicated in the table, all reliability coefficients except that for Group Compatibility, as rated by the team members, are significant beyond the 01 level Generally speaking, the training instructors and the potential supervisors were more consistent raters than the simulation team members

The "All Raters" column in Table 3 presents the reliability coefficients for the three rater groups considered collectively The reliability coefficients for the combined groups range from 81 for Group Compatibility to 94 for Knowledge of the Job It is noteworthy that the over-all stanine rank reliability coefficient was quite high (92), and, as the results of the TDC study also indicate, may have been a sufficiently reliable index of job performance if time considerations did not warrant the use of the more extensive eight-trait rating scale

[1] All raters did not know all operators equally well, thus they were given the option of rating or not rating each operator The average number of raters rating each group was computed as described in Section Three

Table 3

Reliability of Eight-Trait Ratings
And Stanine Ranking Scores (NAFEC Sample)

	Instructors (n=3)		Potential Supervisors (n=8)		Team Members (n=4)		All Raters	
Trait	r_{tt}	k*	r_{tt}	k*	r_{tt}	k*	r_{tt}	k*
Knowledge of the Job	88	2 88	92	6 22	69	3 22	94	12 20
Learning Potential	90	2 94	92	6 56	77	3 43	94	12 93
Communicating Abilities	86	2 92	90	6 40	37	3 45	92	12 71
Attitude toward Work	84	2 92	80	6 16	59	3 47	86	12 50
Accuracy	86	2 96	92	5 95	70	3 45	92	12 30
Adaptability	87	2 92	91	5 97	64	3 41	93	12 24
Dependability	90	2 92	82	6 14	45	3 39	.89	12 40
Group Compatibility	81	2 92	74	6 50	18**	3 47	81	12 89
Stanine Ranking	85	2 96	90	5 77	66	3 15	92	11 83

*k = Average number of raters entering into the determination of the trait ratings

** = Unreliable index, all other significant beyond the 01 level

Since, as would be expected, the reliability coefficients for all three groups combined were somewhat higher than for any of the three groups separately, a composite measure from all of the raters was again chosen for use in determining the relationship among the criterion measures

2 Intercorrelations Among the Criterion Measures An examination of the intercorrelations among traits and ranking, as shown in Table 4, indicates a high relationship among the variables This relationship suggests that only one aspect of performance was being measured and that it was producing a pervasive halo-effect throughout the criteria Actually, only one trait, Group Compatibility, gives even the slightest indication of measuring any unique elements of the operators' performance In most cases, the intercorrelations are as high as the reliabilities of the indices

3 Selection of Criterion Measure Although the intercorrelations among the trait ratings and rankings were higher than had been anticipated, this outcome did aid in the decision to use a collective criterion measure Since weightings of the criteria based on their reliability did not appreciably increase their value, it was decided to sum the standard scores for the trait ratings and ranking for use as the criterion measure Inclusion of all of the criterion measures was decided upon because of these high intercorrelations It was also decided to use the three groups of raters in order to achieve increased reliability Thus the final criterion measure consisted of the summed standard scores for the simulator operators on the eight traits and the stanine ranking, performed by the combined three groups of raters

Table 4

Intercorrelation of Eight-Trait Ratings, Stanine Ranking Scores, and Composite Criterion Scores (NAFEC Sample)

Indices	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Knowledge of the Job	94	97	82	96	97	87	68	92
(2) Learning Potential		92	82	94	93	82	58	90
(3) Communicating Abilities			81	96	96	87	68	89
(4) Attitude toward Work				83	84	90	72	86
(5) Accuracy					97	89	63	88
(6) Adaptability						90	63	93
(7) Dependability							69	86
(8) Group Compatibility								64
(9) Stanine Ranking								

All of the correlation coefficients are significant beyond the 01 level

C Comparison of TDC and NAFEC Criterion Measures

Since the activities involved in the determination of criterion measures at TDC and NAFEC were similar in most instances, it seems appropriate to compare the results obtained from the two investigations A point to keep in mind throughout this review is that the TDC study involved supervisor raters who knew the operators for varying lengths of time In the NAFEC study, on the other hand, ratings were derived from three different groups of raters Raters in each of these groups knew the operators for the same length of time but the groups differed in the extent to which they had an opportunity to observe the operators The team members had occasion to observe the operators only for approximately a five week period, whereas the other two rater groups had had approximately twelve weeks for observation

1 Reliability In order to facilitate direct comparison of the reliability coefficients, the coefficients from the various groups of raters in the NAFEC study presented in Table 3 were each adjusted to represent the same number of raters In the TDC study, the number chosen was the number of supervisory ratings The results of the adjustment are presented in Table 5

As is immediately evident from the table, the instructors provided the most reliable trait ratings Of the four groups of raters, this group had the attention of the operators in a very confined situation--the classroom--and hence, had a better opportunity to know the operators individually This group was also oriented toward operator evaluation because of the necessity

Table 5

Comparative Reliability Coefficients
Of Operator Performance Ratings (TDC and NAFEC Samples)

Trait	TDC		NAFEC		
	k*	Supervisors**	Instructors	Potential Supervisors	Team Members
Knowledge of the Job	4.98	.81	93	90	78
Learning Potential	4 50	78	93	88	82
Communicating Abilities	5 00	76	.91	88	46
Attitude toward Work	4 96	77	90	.76	.68
Accuracy	4 96	79	.91	91	.77
Adaptability	5 00	.76	92	90	.72
Dependability	4 98	81	94	79	.55
Group Compatibility	4 98	82	.88	68	24***
Stanine Ranking Scale	5 04	82	91	.89	76

* k= average number of ratings in the TDC sample used to adjust the reliability coefficients of the NAFEC ratings

** Duplication of the reliability coefficients presented in Table 1 for use of comparison

*** Insignificant index, all others significant beyond 01 level

for giving final grades for the training period This "set" carried over to the on-the-job training phase as well as the instructors' continued interest in the development of the operators Also, the instructors shared a common office and a large part of office talk centered around the evaluation of the operators Thus, although the instructors stated that they had independently completed the criteria evaluations there seems little doubt that the high degree of relationship among their results was due to judging from common frames of reference developed over an extended period of working together

The "potential supervisors" had a number of aspects in common with the instructors First, they knew their "peers" for the same length of time as the instructors Secondly, they were evaluation-conscious, since they were given periodic achievement tests during the training period which provided an indication of their absolute, as well as relative progress Also, because of their physical proximity to one another, as well as a common communications channel with the controllers, they were provided with frequent and immediate feedback as to the adequacy of their own and of their peers' performance

A note on the independent ratings of this group lies in the fact that the scales were administered to all raters in a controlled situation and the group did not know beforehand that they were going to rate their "peers" This analysis indicates that in this type of situation peer ratings may not only be adequate but may also be better indices of operator performance than the ratings of any other group

The fact that the TDC supervisors' reliabilities are generally not as high as the NAFEC instructors' or the "potential supervisors'" appears to be out of keeping with the fact that they had a longer time in which to know their operators than the NAFEC raters This discrepancy is probably attributable to the evaluation-consciousness prevalent at the new installation at NAFEC, and to the absence of a formal operator training and evaluation program at the TDC facility Another aspect in the NAFEC data which was an outgrowth of the TDC study was the use of an identifying photograph of the group of simulator operators Many of the TDC supervisors had felt that they knew more of the operators by sight or by "nickname" than they rated, but could not identify them by name alone on the rating scales Hence, a group photograph of the operators was provided to each rater in the NAFEC study to facilitate operator identification It was felt that the possibility of making errors in identifying operators (rating the wrong person) was practically eliminated

Results indicate that the team members in the NAFEC simulation problems were decidedly the poorer raters One explanation for this is that they had the least opportunity to observe the operators Also, they were not directly involved in the evaluation of the operators, and only did so as a gesture of good will for the investigators An ordering of the rater groups, according to their opportunity for observation of the operators, places the instructors at NAFEC first, as having had the most opportunity, next were the NAFEC "potential supervisors", then the TDC supervisors,

and finally the NAFEC simulation team members A comparison of the reliability coefficients (Table 5) across these groups indicates a strikingly similar ordering Thus, it is concluded that opportunity for observation by raters was the primary consideration in obtaining reliable operator peformance ratings in this study

2 Intercorrelations of Criterion Measures Comparisons were made between the results of the intercorrelation matrices of the eight-trait ratings, and ranking score at TDC (Table 2, page 55) and at NAFEC (Table 4, page 61) The results indicate that the degree of relationship among the ratings is much higher in the NAFEC study where the median intercorrelation is 875 with a range of 63 to 97 In the TDC study the median intercorrelation was 68 with a range of 10 to 92

One major difference among the traits in the two studies is the intercorrelation pattern evidenced by Attitude toward Work In the TDC study the median correlation for this trait is 355 while in the NAFEC study it is 825 This may have to do with the changing nature of the job with time or with the initial confounding of measurement of this trait with job performance in the NAFEC study The former aspect possibly is a better explanation since with time the task may be less challenging and it is more reasonable that the less efficient operator would retain a greater motivation for improvement than one who is already confident of his effectiveness

Because of the very high intercorrelations among the measured traits (Table 4) the results further indicate that in the NAFEC study there is only

one factor measured In the TDC data there is some evidence of a second factor grouping around the Attitude toward Work and Group Compatibility traits

One interpretation of these results is that "good performance," at least during the early months on the job, seems largely dependent upon the possession of a general ability or proficiency factor The influence of this factor seems sufficiently pervasive and significant to mask out other would-be measurable factors, such as those incorporated into the rating scale That is, only after extended experience on the job is the initial advantage of having this general ability factor offset by such other traits as Attitude toward Work, Dependability, Group Compatibility, etc Perhaps by then the job offers less of a challenge to the more capable operators and tends to reduce or negate their early all-pervasive advantage on operator performance traits

As mentioned earlier, the design of the eight-trait rating scale was based on the premise that a number of fairly independent abilities and characteristics contribute to effective performance as a simulator operator, and that rater evaluations on the basis of these traits would discriminate among operators more effectively This premise was not substantiated by the NAFEC results and only partially so by the TDC results

3 General Evaluation The analysis of the criteria used in this study indicate

a) The reliability of the ratings is a function of adequate observation of operators and interest and "set" toward evaluation

b) Selected operators from the group to be rated can reliably rate their "peers" with a high degree of accuracy when there is an adequate opportunity to observe the group's performance

c) Job proficiency on this task, particularly in the early stages of the operation, tends to be based on a global factor which masks any of the sub-traits from which it is derived Hence, use of a stanine ranking scale seems as adequate as the more time consuming eight-trait rating scale

d) Use of an identifying photograph of the group of operators as in the NAFEC study facilitated the raters' task and possibly contributed to the increased reliability of the NAFEC ratings.

II VALIDATION OF PREDICTIVE INDICES

The results of the validation will be presented first for the TDC situation with experienced operators and secondly, for the NAFEC situation with newly-hired operators Finally, a comparison of the results of the two settings will be made

A Technical Development Center

The results of the TDC data are considered to be preliminary for a number of reasons (1) they represented an experimental development and tryout of the rating scales, (2) the operators had received no formal training for their jobs, (3) the experimental test battery was administered to the operators on their last day on the job, and (4) the operators had known for some time that the operation at TDC was to be terminated, therefore, job motivation was probably less Finally, before their cooperation with this study, the supervisor-raters did not have the necessity to evaluate the operators, therefore, they were not evaluation-conscious

Despite these shortcomings, it was still thought desirable to collect the various measures for an empirical determination of the validity of the various predictive indices The results of the Pearson product-moment correlation coefficients for the predictor variables, namely, test scores, Age, Years of Education, Months of Experience, and the composite rating for the 49 simulator operators at TDC, are presented in Table 6

Table 6

Intercorrelation of Predictors and Criterion Measures (TDC Sample)

Indices	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
(0) Months of Experience	38	20	- 02	- 05	00	- 13	- 19	- 29	- 30	- 15	- 06	- 16	- 36	31
(1) Age	- 22		15	- 32	- 12	- 17	- 16	- 34	- 43	05	- 30	- 29	- 49	- 23
(2) Years of Education			25	32	01	24	13	34	33	17	28	01	24	31
(3) Verbal Comprehension				10	11	12	32	22	08	49	- 06	31	07	03
(4) Numerical Ability					22	12	44	57	15	42	45	32	39	08
(5) Visual Pursuit						04	29	24	- 03	02	42	35	19	07
(6) Visual Speed and Accuracy							01	29	21	39	23	- 15	42	- 11
(7) Space Visualization								56	24	34	18	50	33	04
(8) Numerical Reasoning									34	33	27	36	60	01
(9) Verbal Reasoning										09	19	12	54	38
(10) Word Fluency											13	11	17	- 11
(11) Manual Speed and Accuracy												24	23	10
(12) Symbolic Reasoning													22	18
(13) Civil Service Examination plus Veterans Bonus														- 05
(14) Composite Criterion														

35 is significant at 01 level
27 is significant at 05 level

1 Civil Service Examination The validity study of the Civil Service Examination (- 05) indicates that the examination was no better than chance as a determinant of operator performance It is of interest to note that this conclusion bears out the supervisors' feelings that the examination was not effective

The Civil Service Examination was related to the Numerical Ability (39), Visual Speed and Accuracy (42), Numerical Reasoning (60), and Verbal Reasoning (59) tests of the Employee Aptitude Survey at the 01 level of significance, and to the Space Visualization (33) at the 05 level of significance It was negatively related to Months of Experience (- 36) and Age (- 49) at the 01 level of significance That is to say older people tended to score lower on the examination Another interpretation would be that people who score high on the test do not remain on the job However, this latter interpretation may be an artifact due to the placement of college students on the job who have fewer months of experience because of their transitory nature, nevertheless, the fact that Education (24) was not significantly related to the Civil Service Examination was contrary to this explanation

2 Age and Education Age was not significantly related to success as a simulator operator It was negatively related to six of the predictor variables Numerical Ability (- 32), Numerical Reasoning (- 34), Verbal Reasoning (- 43), Manual Speed and Accuracy (- 30), Symbolic Reasoning (- 29), and the Civil Service Examination (- 49) This follows the general findings

of research workers in the field of differential psychology that younger people perform better on paper and pencil tests Another possible explanation which may apply in this situation is that the young people are closer to the classroom situation and more accustomed to taking tests, the complete significance of which is unknown to them, while the older people are more detached and more annoyed by the apparent waste of time and other factors brought on by such a situation

Age was also related to Months of Experience (38) Thus, older persons tended to have a longer tenure, which is a typical finding

Education (31) was one of the two predictors which was significantly related to the index of successful operation of the target generator at the 05 level of significance It was related to the same Employee Aptitude Survey tests as Age, except in a positive direction, however, it was not related to the Civil Service Examination, Age, or Months of Experience

3 Employee Aptitude Survey Only one of the ten tests of the Employee Aptitude Survey was a significant predictor of the operator proficiency measure That test, Verbal Reasoning, had a validity coefficient of 38 (significant at the 01 level) The Verbal Reasoning factor correlated highly with the Civil Service Examination (54) and was also significantly related at the 05 level to the Numerical Reasoning factor (34)

The results of the intercorrelations of the Employee Aptitude Survey (Table 6) indicate that the tests are fairly independent since there are 25 non-

significant intercorrelations However, ten of the intercorrelations are significant at the 01 level and nine are significant at the 05 level The Numerical Reasoning and Word Fluency factors enter in significantly on a number of the factors

4 Months of Experience The length of time an operator spent on the job was significantly related to the criterion measure This is the usual, more experience/better performance element However, this factor, which cannot be truly called a predictor, has some interesting relationships with the predictor variables It is negatively related to the Civil Service Examination (- 36) at the 01 level of significance, and is also negatively related to one of the significant predictors. Verbal Reasoning (- 30) and to the Numerical Reasoning (- 29) at the 05 level This seems to indicate that either (a) people with less ability (as measured by the predictors) were initially hired, or (b) through self-selection, the more capable people tended to leave the position after a period of time

5 Overview It is felt that the general results of the selection study at TDC are not favorable for a number of reasons First, the instrument that had been used for selection--the Civil Service Examination--was not related to the criterion of successful simulator operation Second, only one of the ten tests of the Employee Aptitude Survey (Verbal Reasoning) selected on the basis of supposed correspondence to the traits thought necessary from the activity analyses, was significantly related to the criterion measure

However, the conditions of testing may have influenced these results Third and finally, is the relationship of the Months of Experience variable Although this variable is related to successful operation of the console it was negatively related to a number of the test variables including the Civil Service Examination

Thus, of 13 possible predictors, only two were related to the criterion measure One of these was the Verbal Reasoning Test and the other was Years of Education Neither of these factors were being used as a selection device

B NAFEC Results

The results which follow represent a collection of selection data under more nearly optimum conditions Following the transfer of the TDC simulator system to NAFEC, data was gathered on the newly-trained operators This not only eliminated the difficulty of isolating the effects of varying Months of Experience variable, but also permitted the study of an immediate criterion (training grades) as well as an intermediate criterion (job performance rating)

Pearson product-moment correlation coefficients were computed among the predictor variables, namely, test scores, Age, Years of Education, and the composite ratings for the 51 simulator operators Correlation coefficients of these variables on the training grades are also included The results of the intercorrelations are presented in Table 7

Table 7

Intercorrelation of Predictors and Criterion Measures (NAFEC Sample)

Indices	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1) Age	24	64	04	08	- 08	- 07	24	04	17	- 15	- 27	08	12	- 32	- 18
(2) Years of Education		29	20	35	34	21	17	14	30	40	07	33	32	23	13
(3) Verbal Comprehension			51	29	37	31	53	42	56	16	12	38	38	08	31
(4) Numerical Ability				45	68	62	71	61	43	40	43	46	50	36	68
(5) Visual Pursuit					39	58	44	42	27	40	33	31	31	26	31
(6) Visual Speed and Accuracy						52	58	53	51	53	32	66	59	47	63
(7) Space Visualization							53	67	32	36	68	46	44	35	54
(8) Numerical Reasoning								72	45	24	44	41	35	15	54
(9) Verbal Reasoning									38	30	53	42	34	16	37
(10) Word Fluency										24	08	44	38	31	23
(11) Manual Speed and Accuracy											24	42	40	46	30
(12) Symbolic Reasoning												11	09	24	45
(13) Civil Service Examination													94	51	47
(14) Civil Service Examination plus Veterans' Bonus														52	50
(15) Composite Criterion															52
(16) Classroom Training Grade															

27 is significant at the 05 level
35 is significant at the 01 level

1 Civil Service Examination The Civil Service Examination plus the Veterans' Bonus proved to be the most effective prediction instrument, since it correlated 52 with the criterion Since the Civil Service Examination plus bonus served as the basis for selection, the index is affected to a greater extent by truncation than the other indices and the true validity coefficient, when corrected for truncation, would be higher It is interesting to note that the Civil Service Examination score alone was 51, hence, although not significant, the veterans' bonus points raised the validity coefficient

The Civil Service Examination was related to nine of the ten tests of the Employee Aptitude Survey The relationships which are significant at the 01 level are Verbal Comprehension (38), Numerical Ability (46), Visual Speed and Accuracy (66), Space Visualization (46), Numerical Ability (41), Verbal Reasoning (42), Word Fluency (49), and Manual Speed and Accuracy (42) The test measuring Visual Pursuit (31) was significantly related to the Civil Service Examination at the 05 level Of the ten tests, only the Symbolic Reasoning test failed to significantly relate to the Civil Service Examination These significant correlations suggest that the Civil Service Examination is measuring general aspects of intelligence Actually, the high degree of correlation among tests of the Employee Aptitude Survey suggests that the factors are not independent and probably contain a second order, general ability factor. It is probably this general ability factor, which the EAS sub-tests and the Civil Service Examination have in common, that accounts for the high intercorrelation

2 Age It is interesting to note the relationship of the non-test variables on the criterion Age (- 32) was significantly related to the criterion at the 05 level Thus the older operators tended to be rated as less effective performers on the job This is true even though Age is related negatively at the 05 level of significance on only one test variable, Symbolic Reasoning (- 27), and positively (01 level) on Verbal Comprehension (64) However, neither of these traits were related to the criterion Thus, although younger individuals generally tend to be better performers on the job as well as on paper and pencil tests, this was not true for the test variables in this study Actually, the present sample of operators suggests evidence of a complete reversal of the usual findings of the relationship between age and years of education It is a general finding, because of the recent emphasis on education as well as the enforcement of compulsory education, that younger people have more education and thus, age is significantly related to education only in a negative way Actually, the obtained correlation of 24 between Age and Years of Education suggests (though lacking statistical significance) the reverse of this That is, the current older group tended to have more formal education that the younger group, but did not perform any better on the aptitude test, and apparently were not motivated by the job, since Age was negatively related to success on the job

3 Years of Education The number of Years of Education (23) was not significantly related to success as a simulator operator However, it was significantly related to five of the Employee Aptitude Survey scores,

and to the Civil Service Examination Two of the correlations were significant at the 01 level, Visual Pursuit (35) and Manual Speed and Accuracy (40) This is somewhat contrary to the general finding that these traits, in contrast to Verbal Comprehension and Reasoning tests, are less affected by Years of Education Verbal Comprehension (29), Visual Speed and Accuracy (39) and Word Fluency (30), were significantly related at the 05 level to Years of Education The Civil Service Examination (33) was significantly related to Years of Education at the 05 level

Thus, Years of Education was not related to the criterion, but was related to three of the significant predictors (Visual Speed and Accuracy, Word Fluency, and Civil Service Examination, plus Bonus) at the 05 level It was related to Manual Speed and Accuracy at the 01 level of significance

4 Employee Aptitude Survey Four of the tests of the Employee Aptitude Survey, namely, Numerical Ability (36), Visual Speed and Accuracy (47), Space Visualization (35), and Manual Speed and Accuracy (46) were significantly related to the criterion at the 01 level, Word Fluency (31) was significantly related at the 05 level Verbal Comprehension, Visual Pursuit, Numerical Reasoning, Verbal Reasoning and Symbolic Reasoning were not significantly related to the criterion measure

A further look at the intercorrelation matrix in Table 7 for the Employee Aptitude Survey indicates that 32 of the intercorrelations are significant beyond the 01 level and seven at the 05 level Only six of

the intercorrelations are not significantly related, and only one of the intercorrelations among the tests significantly related to the criterion, is not significantly related Thus, as stated previously, there is evidence of a second order, general ability factor from the results of the intercorrelations of this study

5 Multiple Regression Analysis of the Employee Aptitude Survey

In order to get a more precise description of what tests in the Employee Aptitude Survey were optimally predicting the criterion measure, the Wherry-Doolittle Test Selection Method, as described by Garrett (8), was utilized The results of the regression analysis for the selection of the simulator operators are presented in Table 8 Six of the ten tests entered into the determination of the multiple correlation In the multiple regression analysis, four of the five tests that were significantly related to the criterion were positively weighted on the criterion and two of the tests not related to the criterion were negatively weighted and acted as suppressor variables The order in which the tests were selected was

Numerical Ability, Manual Speed and Accuracy, Numerical Reasoning, Space Visualization, Word Fluency, and Verbal Reasoning

Table 8

Tests Selected from EAS Battery with Weights for Maximum Validity Using Wherry-Doolittle Method (NAFEC Sample)

Order of Selection	Predictor Variable	Multiple Correlation	Shrunken Multiple Correlation	Beta Weights
1	Numerical Ability	469	469	3128
2	Manual Speed and Accuracy	532	518	2634
3	Numerical Reasoning	547	519	- 1551
4	Space Visualization	566	526	2613
5	Word Fluency	580	527	1540
6	Verbal Reasoning	593	529	- 2042

The multiple correlation was 59, however, after Wherry's shrinkage correction is applied to the multiple correlation, it is reduced to 53 which is no better than the results obtained with the Civil Service Examination The shrunken multiple correlation represents the maximum correlation with the criterion after correction has been made for chance fluctuations in the correlation coefficients

6 Training Grades It is interesting to note the relationships of the variables with the training scores of the operators All of the test factors except Word Fluency are related to the training grade Actually, the correlations of the tests scores with the training grade were much higher than with the composite criterion On the Employee Aptitude Survey, Numerical Ability (68), Visual Speed and Accuracy (63), Space Visualization (54), Numerical Reasoning (54), Verbal Reasoning (37) and Symbolic Reasoning (45) were significant at the 01 level, as was the Civil Service Examination (47) Verbal Comprehension (31), Visual Pursuit (31), and Manual Speed and Accuracy (30) were significantly related at the 05 level Age and Years of Education were not related to training score

As is to be expected, there is a significant relationship between the training grade and the criterion measure (52) However, it is far from being a perfect relationship This is further supported by the change of the correlation coefficients of the tests as predictors of training success and on-the-job performance Word Fluency, which was not significant in predicting training grades, was a significant predictor of the criterion measure, while Verbal Comprehension, Visual Pursuit, Numerical Reasoning, Verbal Reasoning, and Symbolic Reasoning, which significantly predicted training grades, did not predict the job performance criterion The high degree of intercorrelation among the traits suggests that a general learning ability (or test-taking ability) factor was being measured which is a better estimate of paper and pencil training examination grades

7 Optimum Prediction from NAFEC Study To get an index of the optimum prediction of operator success, an additional regression analysis using all of the quantifiable predictors was performed, and is presented in Table 9

Table 9

Best Combination of Predictor Variables with Weights for Maximum Validity Using Wherry-Doolittle Method (NAFEC Sample)

Order of Selection	Predictor Variable	Multiple Correlation	Shrunken Multiple Correlation	Beta Weights
1	Civil Service Examination Plus Bonus	516	516	4135
2	Age	642	632	- 3652
3	Manual Speed and Accuracy	673	655	1994
4	Word Fluency	690	665	1725

The results of this analysis are more favorable since the multiple correlation was 69 and after Wherry's correction for shrinkage, it is 67 Four of thirteen predictors entered significantly into the determination of

the multiple correlation Their order was Civil Service Examination plus Veterans' Preference, Age, Manual Speed and Accuracy, and Word Fluency The weights were in the same direction as the relationship of the variables on the criterion measure Hence, only Age entered into the multiple regression equation with a negative weight

One of the weaknesses of the present study lies in the fact that scores were not available on the total sample of applicants, but only on those who were selected This made the possibility of correcting the validity coefficient of the Civil Service Examination for truncation impossible, and hence, eliminated the possibility of comparison with the validity coefficients of the test scores on the Employee Aptitude Survey derived from a multiple correlation However, the fact that the multiple correlation derived from the Employee Aptitude Survey was 53 in comparison to the 52 on the Civil Service Examination plus bonus indicates that there is not a significant difference between the coefficients

Because of the number of tests involved in the two batteries and the time involved to test the individuals, the Civil Service Examination plus bonus is a more efficient instrument than the Employee Aptitude Survey Practically, the Civil Service Examination should probably remain as the selection instrument, particularly when the difficulties involved in changing a selection system are considered

III COMPARISON OF RESULTS OF THE TWO SAMPLES

The results show that there are no significant validity coefficients common to the TDC and NAFEC samples In other words, the use of the significant predictor variables found in the TDC results to select the operators at NAFEC would have been no better than chance, and vice versa (if the latter situation were possible)

The Civil Service Examination, which was no better than chance in predicting the criterion of successful operation at TDC, emerged as the most significant predictor of success in the NAFEC data The results at TDC verified the supervisors'feelings that the Civil Service Examination was ineffective in predicting operator performance, hence, indicated a need for more effective selector instruments However, the results from the Employee Aptitude Survey showed that only one of its tests, Verbal Reasoning, was effective, i e , positively correlated with the criterion Also, whereas Years of Education and Verbal Reasoning were significant predictors of successful operation of the consoles at TDC, neither of these variables entered in significantly in the NAFEC results

In the NAFEC setting, the results were completely reversed The Civil Service Examination was the best predictor, in fact, it was as good as any combination of tests on the Employee Aptitude Survey, as shown by the results of the Wherry-Doolittle Test Selection method

Five of the ten tests of the Employee Aptitude Survey were related to the criterion, namely, Numerical Ability, Visual Speed and Accuracy, Space Visualization, Word Fluency, and Manual Speed and Accuracy Verbal Reasoning, significant in the TDC results, was not significant in the NAFEC results Still another point of discrepency is that Age was negatively related to the criterion in the NAFEC study, while not being related in the TDC data, however, the TDC coefficient was in the same direction and tending toward significance

How can these results be reconciled? Here, we have two samples of operators working on the same target generator This selection study, using the same predictors, and same criterion, came up with the results that are directly opposed to each other!

One of the possible explanations is that the training program at NAFEC was responsible for the difference That is, as a result of the thorough indoctrination of the trainees regarding the nature and purpose of the air traffic control system, the function of simulation studies, and the significance of their contribution as "pilot operators" to these developmental efforts, their attitudes, interests, and motivation were favorably affected Without such an influence, or perhaps after an extended period on the job, this "esprit de corps" might not exist

Another possible explanation is that the ability and learning factors are important in the early stages of this position, when the simulation operation has intrinsic interests After a certain period of time, interest

in the system declines, and the operation of the target generators becomes routine Thus, the brighter operators would tend to get bored and, consequently, their performance would degenerate while those with lesser ability would maintain a higher level of motivation and better performance

The conditions of collecting the data may also have affected the results The NAFEC group had much more interest in the job They were better motivated in the Employee Aptitude Survey tests session, since it was given on the first day on the job, and despite being informed otherwise the operators felt that their performance would affect their status in the position This can be contrasted to the TDC group, which was called in for testing for half a day after the operation of the simulation system had been closed down Also, the significance of improving the selection system at TDC could have no personal meaning to the operators, since it did not involve a direct benefit to them, or to their simulation facility Another factor is that the supervisors who performed the operator evaluations at TDC did not know the operators as well as the NAFEC raters did, as can be inferred from the fact that the average number of ratings from the total possible raters was less at TDC

Whatever the reason the contrasting results of the two studies indicate that until further research is performed on these selection instruments at NAFEC, their usefulness as predictors of operator success over

a long period of time would be very tentative It may well be that the TDC results were due to an artifact of the unusual conditions present at the TDC, Indianapolis installation However, a longitudinal follow-up of the NAFEC operators to see how time affects their performance is necessary to clarify the results

IV CONCLUSIONS

The results of this study support the following conclusions

A Activity Analyses

1 The activity analyses on the Technical Development Center simulator system and the Aircraft Armaments simulator systems indicate that the operator job requirements are approximately the same

2 The job requirements found by the activity analyses aided in the selection of the test factors measured by the Employee Aptitude Survey

3 The activity analyses, coupled with TDC supervisors' assistance, aided in the development of criterion measures of simulator operator performance

B Criterion Measures

1 The eight-trait rating scales of job performance were reliable criteria in both the TDC study and in the NAFEC study

2 There was little indication of independence in the job performance traits as measured by the eight-trait rating scales

3 The stanine ranking scale was a reliable criterion measure in both studies

4 The eight traits of the rating scale correlated highly with the stanine ranking scale, giving further indication that one factor was being measured--over-all job performance--creating a "halo-effect" in the eight-trait ratings

5 The reliability of the ratings of the rater groups appeared to be a function of their opportunity to observe the operators The order from best to poorest was NAFEC instructors, NAFEC 'potential' supervisors, TDC supervisors and NAFEC team members

6 The use of an identifying photograph with a roster of the names of the operators aided the reliability of the NAFEC ratings

7 A composite sum of the standard scores of the raters on all the rating scales gave the most stable index of job performance in both the TDC results and the NAFEC results

C Validation Results

1 The Civil Service Examination was not effective in long-term prediction of operator performance as indicated by the TDC results

2 The Civil Service Examination was effective in short-term prediction as indicated by the significant predictors of operator performance on training grade, and on-the-job performance measures in the NAFEC results

3 Only one of the ten tests of the Employee Aptitude Survey, Verbal Reasoning, was a significant predictor of operator performance in the TDC results

4 Five of the ten tests of the Employee Aptitude Survey were related to the job performance criterion in the NAFEC results The tests were Numerical Ability, Verbal Speed and Accuracy, Space Visualization, Word Fluency, and Accuracy

5 The results of the Wherry-Doolittle technique applied to the tests of the Employee Aptitude Survey in the NAFEC results, indicated that the multiple correlation of the best combination of its tests on the criterion was no better than prediction from Civil Service Examination plus Veterans' Preference Bonus

6 Years of Education was a significant predictor of success in the TDC results

7 Age was negatively related to performance in the NAFEC results

8 The results of the Wherry-Doolittle technique applied to all the predictors of the NAFEC results raised the shrunken multiple correlation to .665 by combining the results of the Civil Service Examination plus Bonus, Age (negatively weighted), Manual Speed and Accuracy and Word Fluency

9 A comparison of the TDC and NAFEC results shows there are no significant validity coefficients common to both studies

REFERENCES

(1) Anderson, C. M. , and Dowling, C. E. EVALUATION BY SIMULATION TECHNIQUES OF A PROPOSED TRAFFIC CONTROL PROCEDURE FOR THE NEW YORK METROPOLITAN AREA Civil Aeronautics Administration, Technical Development Report No 245, August 1954

(2) Anderson, C. M., McKnight, F. S., Vickers, T. K., and Yost, M.H. EVALUATION BY SIMULATION TECHNIQUES OF PROPOSED TRAFFIC-CONTROL PROCEDURES FOR THE NORFOLK TERMINAL AREA. Civil Aeronautics Administration, Technical Development Report No 185, January 1955

(3) Anderson, C. M., and Vickers, T. K. APPLICATION OF SIMULATION TECHNIQUES IN THE STUDY OF TERMINAL-AREA AIR TRAFFIC CONTROL PROBLEMS. Civil Aeronautics Administration, Technical Development Report No 192, November 1953

(4) Astholz, P T , and Vickers, T. K. A PRELIMINARY REPORT ON THE SIMULATION OF PROPOSED ATC PROCEDURES FOR CIVIL JET AIRCRAFT. Civil Aeronautics Administration, Technical Development Report No 352, May 1958

(5) Baker, R. E., Grant, A. L., and Vickers, T. K. DEVELOPMENT OF A DYNAMIC AIR TRAFFIC CONTROL SIMULATOR. Civil Aeronautics Administration, Technical Development Report No 191, October 1953 AD 38596

(6) Danaher, J. W., Colman, K. W., and Courtney, Douglas THE SIMULATOR PILOT IN A DYNAMIC AIR TRAFFIC CONTROL SIMULATOR AN ACTIVITY ANALYSIS (Interim Report) Courtney and Company Report #25, Project N, 15 February 1959 AD 214081

(7) Ebel, R. L. ESTIMATION OF THE RELIABILITY OF RATINGS. Psychometrika, 16, 1951, 407-424

(8) Garrett, H. E. STATISTICS IN PSYCHOLOGY AND EDUCATION (5th Ed) New York Longmans, Green and Company

(9) Grimsley, Glen, Ruch, F. L., Warren, N. D , and Ford, J. S. THE EMPLOYEE APTITUDE SURVEY Los Angeles Psychological Services, Inc , 1952-1958

(10) Gulliksen, Harold THEORY OF MENTAL TESTS New York John Wiley & Sons, Inc , 1950

(11) Hixson, W. C., Harter, G. A., and Warren, C. E. A RADAR SIMULATOR FOR USE IN AIR TRAFFIC CONTROL STUDIES. Wright Air Development Center, Technical Report 53-418, January 1954. AD 33463

(12) Johnson, Beatrice E , Williams, A. C., Jr , and Roscoe, S. N. A SIMULATOR FOR STUDYING HUMAN FACTORS IN AIR TRAFFIC CONTROL SYSTEMS National Research Council, Committee on Aviation Psychology, Report No. 11, June 1951.

(13) Snedecor, G. W. STATISTICAL METHODS. Ames, Iowa The Iowa State College Press, 1956

(14) Thorndike, R. L. PERSONNEL SELECTION TEST AND MEASUREMENT TECHNIQUES. New York John Wiley & Sons, Inc , 1949

(15) Vickers, T. K. ATC LABORATORY J of Air Traffic Control, Vol 1 (3), January 1959, 4-7

(16) Wilson, C. L., Mackie, R. R., and Buckner, D. N. THE USE OF A PERFORMANCE RATING SCALE IN THE MEASUREMENT OF SHIPBOARD PERFORMANCE OF ENLISTED NAVAL PERSONNEL PART II of Final Report in Five Parts on RESEARCH ON THE DEVELOPMENT OF SHIPBOARD PERFORMANCE MEASURES. Human Factors Research Corporation, February 1954

APPENDIX A

AIR TRAFFIC CONTROL SIMULATOR OPERATOR JOB SPECIFICATIONS

(Derived from the Activity Analyses, Personnel Interviews and Analyst Experience)

I Physical Requirements

- A Minimum height 5 feet
- B Minimum age 18 years
- C Sex male or female
- D Motor response capabilities
  - 1 Use of both hands simultaneously
    - a Mounting of display maps*
    - b Gross coordination (positioning of toggle switches, etc)
    - c Fine coordination (fine adjustment of controls)
  - 2 An amputation of leg or foot, in most cases should not disqualify an applicant

II Sensory Requirements

- A Vision--normal acuity (corrected, if necessary)
  - 1 Depth perception not critical
- B Hearing--sufficiently sensitive to understand messages received through a headset (Hearing aid not permissible)

III Cognitive requirements (Mental Alertness and Intelligence)

- A Ability to follow oral and written instructions
- B Ability to learn new procedures and routines with a minimum of practice and individual instruction
- C Verbal ability--understand English and speak it fluently and without obvious impediment or heavy foreign accent
  - 1 Ability to copy and read back extended instructions and to make verbal position reports
- D Ability to read and interpret information presented on display maps
- E Basic arithmetic ability (use of Aircraft Navigation computer, type D-4, for solving speed-time-distance problems)

---

* This requirement applies only to the Aircraft Armaments simulator

IV Personality and Emotional Requirements

A Conscientious attention to details of task to insure accuracy and consistency of accomplishment (Particularly in data recording aspects of the job)

B Ability to maintain a desired performance level under varying work loads

C Dependability or persistence

1 Ability to accomplish an assigned task ("flight") without continuous supervision

2 Ability to remain seated at a control console for extended periods (2 hours or more)

D Adaptability and deftness in getting along with others and working in close proximity to them

V Interest Requirements

A A keen interest in learning about the aviation field in general, and air traffic control in particular

B Interest in contributing to the improvement of ATC as a member of a group research project

INSTRUCTION TO RATERS

You will find the names of a group of simulator operators listed in this booklet They are to be rated by you on several characteristics which you will recognize as important for success as a simulator operator

A characteristic is described in a section at the upper left corner of each page Below this section are four categories containing the various degrees of the characteristic in question Some persons in your group will be pretty well described by one particular group of statements, while others may seem to fall between two groups of statements Therefore a scale index provides a guide for finer judgment within each broad category Your task is to indicate with one check mark the position of each operator in your group with respect to the trait in question

A sample rating sheet is shown below to indicate how a completed rating might look Here each person has been compared with the others and rated on the trait of JOB KNOWLEDGE On the sample page, L L Furgeson is among the best of his group in this trait, therefore a check mark appears near the top of the column under his name Notice also that the check marks have been well spread out and that no two people received exactly the same rating These are two characteristics of well-performed ratings

Observing the following procedures will help make your ratings most valuable

1) Consider each operator in comparison to the others in the group as you rate each trait There will be a few outstanding persons and a few who are not so good, while most will fall somewhere in between

2) Rate the best and the poorest operator in each trait first With these limits established, rate the remainder of the group in relation to them

3) If possible, avoid giving tie ratings Before you rate two persons exactly equal, consider carefully whether there is not some small difference between them on this particular trait

4) Remember that most persons' ratings will vary from trait to trait Because a man is high (or low) in one trait does not necessarily mean he will be high (or low) on the others Each trait should be considered separately For example, it is improbable that the same operator will be highest (or lowest) in <u>every</u> trait in the scale Experience and ability generally run together, but they are not always perfectly related

5) THIS IS IMPORTANT These ratings are to be used for research purposes <u>only</u> and will be kept confidential and anonymous We sincerely request your frank and honest impressions

KNOWLEDGE OF THE JOB Consider the following factors in your rating of each person How well does each operator know the technical duties of the job? Does he (she) know all the fine points or just enough to get by? How often does someone have to help get the job done? Does he (she) know more or less than the other operators in the group?	SCALE INDEX	Brownell, E J	Carson, J A	Furgeson, L L	Harrison, T J	Moore, F G	Parker, F L	Thompson W N	Turner, S E
Has as much job knowledge as the very best operator Knows all the fine points of the job Never needs help on the job				✓					
Practically never needs help on the job Has very little to learn about the job except for a few of the most technical aspects			✓				✓	✓	
Has adequate knowledge to turn in a satisfactory job Sometimes needs help on the job, still has a few things to learn		✓				✓			
Has much to learn about the job Needs supervision very frequently, since he (she) lacks much of the fundamental knowledge necessary to do the job alone					✓				
Not Observed									

KNOWLEDGE OF THE JOB Consider the following factors in your rating of each person How well does each operator know the technical duties of the job? Does he (she) know all the fine points or just enough to get by? How often does someone have to help get the job done? Does he (she) know more or less than the other operators in the group?	SCALE INDEX																	
Has as much job knowledge as the very best operator Knows all the fine points of the job Never needs help on the job																		
Practically never needs help on the job Has very little to learn about the job except for a few of the most technical aspects																		
Has adequate knowledge to turn in a satisfactory job Sometimes needs help on the job, still has a few things to learn																		
Has much to learn about the job Needs supervision very frequently, since he (she) lacks much of the fundamental knowledge necessary to do the job alone																		
Not Observed																		

LEARNING POTENTIAL Consider How rapidly and thoroughly did this operator absorb initial training? Was extra training required to achieve satisfactory performance? Did he (she) indicate unusual insight during training? In brief, compared with other individuals in this group, how much effort was required to train this operator?	SCALE INDEX
Learns very rapidly and thoroughly Grasped the essentials quickly and was able to absorb the training without difficulty and with a minimum of explanation and repetition	
One of the better operators with regard to learning potential Performed effectively after minimum training and in most cases did not require additional explanation of concepts	
A reasonably receptive trainee May have required some extra training or explanation on occasions, however, retention of the material learned was adequate with periodic review	
Experienced difficulty on some phases of the training Had difficulty grasping procedures or operating instructions Required extra training to achieve satisfactory performance	
Not Observed	

COMMUNICATING ABILITIES Consider How well does he (she) use voice procedures and maintain radio discipline? How well does he (she) copy and follow flight clearances? Does he (she) reply p·ompt1y and sufficiently loud and clear when called on the radio?	SCALE INDEX
Uses exceptionally good voice procedures and maintains excellent radio discipline Rarely makes mistakes in copying, reading back and executing clearances Responds promptly and properly to calls	
Uses very good voice procedures and rarely breaks radio discipline Can be depended upon to copy, read back and execute properly all but the more difficult clearances	
Is not as conscientious about radio discipline and voice procedures as most operators in the group Sometimes has difficulty in copying and executing clearances	
Ranks among the poorer operators regarding use of voice procedures and radio discipline Is likely to require repetitions or corrections before properly copying and executing a flight clearance	
Not Observed	

ATTITUDE TOWARD WORK Consider If extra effort is called for does he (she) respond with enthusiasm? Does he (she) seek to learn more about the job or make suggestions to improve group performance? Is the operator only interested in the work as a "job" or is he interested in the implications of the research?	SCALE INDEX
Consistently does more than expected Likes to master new tasks Has asked for copies of reports or shown other indications of interest in research findings	
Alert and generally does more than required Tries harder to improve performance than most others in the group Is interested in producing worthwhile simulation results	
Does his (her) share of the work, but seldom does more than required May occasionally show flashes of interest in the work and tries to improve own performance	
Contributes as little as possible to the work Interest is limited to the job as a source of income Rarely demonstrates behavior indicating concern for simulation results	
Not Observed	

ACCURACY Consider How accurate is the operator in his (her) work? Does he (she) make errors in setting up for a simulation problem, e g , wrong speed set in? Does the operator occasionally turn the wrong direction or allow the target to wander off an assigned track?	SCALE INDEX
Is exceptionally accurate in doing assigned work One of the very best operators in this group Makes few, if any, errors in set up and closely follows assigned routes and clearances	
Better than most operators at avoiding errors Executes turns, climbs and other maneuvers with a reasonable degree of accuracy Demonstrates better than average accuracy in performance	
Makes occasional errors but usually detects and corrects them Is not as accurate in operating equipment as most operators in this group	
Makes relatively frequent errors Requires supervision and sometimes has to be told to correct an error Needs to be reminded of the importance of accurate performance	
Not Observed	

ADAPTABILITY Consider When assigned to a new simulation situation, how readily does this operator adapt to new procedures or operational requirements? How well does he (she) relate past experience to new requirements? Is his (her) performance in a new situation less effective than that of others in the group?	SCALE INDEX
Is exceptionally quick to grasp the requirements of a new situation Adapts readily with a minimum of explanation Performance is minimally affected by a change	
Can adapt to a new situation more readily than most operators and requires less training and explanation than most to achieve the desired performance level	
Can adjust adequately to new problem requirements after some additional explanation and training He (she) is less capable of applying past experience and knowledge to a new situation than most operators	
He (she) has difficulty in adjusting to a new problem situation Requires detailed explanation and training to perform effectively in new situations	
Not Observed	

DEPENDABILITY Consider Can this operator be counted on when needed? Does he (she) carry out assignments without prodding or supervision? Is he (she) prompt and present when scheduled?	SCALE INDEX																		
Needs very little supervision and works hard without being urged to do so Is rarely late or absent without a good reason																			
Requires less than average degree of supervision in executing assignments Attendance record is good																			
Can generally be depended upon to perform tasks adequately but requires more than average amount of supervision																			
Will "fluff off" assigned work if not closely supervised Rates among the least dependable operators																			
Not Observed																			

COURTNEY AND COMPANY

GROUP COMPATIBILITY Consider Is this operator well liked and respected by all associates? Is he (she) a good team worker--contributing to morale among fellow workers? Or have you noted instances where the individual has been a disruptive influence on the morale of the group?	SCALE INDEX
Exceptionally well-adjusted person Goes out of his way to help others or in other ways provides a favorable influence for the group	
Gets along well with both fellow operators and supervisors Is accepted in the group and is rarely, if ever, a disturbing influence or source of dissension	
Agreeable and usually cooperative May occasionally be temperamental or a source of bickering or griping, but is reasonably well accepted by fellow operators	
Either by own choice or group's attitude is not clearly accepted by others May be tolerated but has been a negative factor in the morale of the group	
Not Observed	

APPENDIX C

STANINE RANKING SCALE

RANKING OF AIR TRAFFIC CONTROL SIMULATOR OPERATORS
(Part II)

In Part I of this rating scale you were asked to rate each simulator operator on each of eight categories

For research purposes, we would like to get an independent measure of the relative merit of the people you have been rating. Therefore, without referring to your previous ratings on individual traits, we would like you to rank the operators on their over-all effectiveness on the job. While doing this, assume that you have been given the task of staffing a new simulator and that you are to select from the attached roster of your operators those most capable of manning it.

Since it would be difficult to assign ranks from ___ through ___ (as several operators may be practically equal in over-all competence), we have simplified your task by permitting you to assign the same rank to several operators Nine ranks have been arbitrarily chosen, and the number of persons to be assigned each rank is indicated below

	Most Competent				Average				Least Competent
Rank	1	2	3	4	5	6	7	8	9
No. of persons assigned that rank	___	___	___	___	___	___	___	___	___

Now refer to the attached roster of operators. Mark a "1" in the "Rank" column beside the ___ best operators in the group. Then mark a rank of "9" beside the ___ least competent operators . Then mark a rank of "2" beside the ___ second best operators. Continue assigning the prescribed number of operators to each rank until all operators have been ranked.

Note It is understood that all operators in your group are reasonably good or they would not retain their jobs. However, we are interested in the competence of each operator relative to all others in the group.

In the space headed "Remarks", make any comment that you feel should be considered by us in analyzing your rankings. Any additional comments may be written on the back side of the roster

Name of Rater _____

Organization _____

Courtney and Company, Inc.

Date _____

ROSTER OF OPERATORS

NAME	RANK	REMARKS

APPENDIX D

INTERVIEW GUIDE

Interviewer _____

Date _____

INTERVIEW GUIDE FOR USE WITH SIMULATOR-PILOT-OPERATOR CANDIDATES

1 Name _____ Address _____
Last First Middle Street City State

Phone Number _____

2 Age _____ 3 Sex _____ 4 Marital Status _____

5 Number of dependents _____ 6 Education (yrs completed) _____

7 Weight _____ 8 Height _____

9 Present place of employment _____

10 Type of Work _____

11 Previous job experience or skills (number of months at each)

Typing _____, Drafting _____, Switchboard Operator _____,

Flying _____, CAP _____ Photography _____,

Operation of a Machine? _____, What kind? _____, Use a compass? _____

12 Ever work as a supervisor? _____, Over how many people? _____

13 What do you do in spare time? _____ Hobbies? _____

14 General physical condition _____

Are you now or have you ever been troubled with

diabetes _____, stomach trouble _____, high blood pressure _____,

nervous breakdown _____, frequent headaches _____

15 Wear eye glasses? Never_____, Sometimes_____, Always_____

16 Color blindness_____

17 Missing fingers_____ which one(s) _____

18 Missing arm_____ which one _____

19 Missing leg_____ which one_____

20 Missing foot_____ which one_____.

21 Wear a hearing aid?_____ which ear_____

Every have trouble hearing conversational sound level?_____.

22 Speech impediments?_____

23 When would you be available to begin work?_____

24 Can you drive an auto?_____

Here show illustrations and photos of the target generator and over-all simulation system to the candidate Describe the duties of an operator and the importance of accuracy and attentiveness in performing them Advise candidate about the one-month training period, the periodic proficiency tests and the general competitive and probationary nature of the first year on the job Those who successfully qualify will be retained, others will be dropped Work will be part-time, with the possibility of becoming regular full-time Job will require extended periods (2 or 3 hours) sitting at console Additional clerical duties may be assigned as necessary

25 How will part-time nature of work affect you? _____

26 Have you any questions? _____

27 How would you get to and from work if hired? _____

28 Do you think you would like this kind of work? _____

Why? _____ (If not, make sure to fill out appropriate form)

INTERVIEWER'S IMPRESSIONS AND COMMENTS

GENERAL APPEARANCE (neatness, bearing, composure, etc):

MENTAL ALERTNESS

ABILITY TO VERBALIZE

DEGREE OF INTEREST DISPLAYED

GENERAL IMPRESSION OF CANDIDATES' SUITABILITY FOR THE JOB

Interviewer's Name _____

Date _____