PROBLEMS IN AIR TRAFFIC MANAGEMENT: II. PREDICTION OF SUCCESS IN AIR TRAFFIC CONTROLLER SCHOOL

BART B. COBB, M.S.*

*Selection Section

Psychology Branch

62-2

FEDERAL AVIATION AGENCY
CIVIL AEROMEDICAL RESEARCH INSTITUTE
AERONAUTICAL CENTER
OKLAHOMA CITY, OKLAHOMA

February 1962

PROBLEMS IN AIR TRAFFIC MANAGEMENT: II. PREDICTION OF SUCCESS IN AIR TRAFFIC CONTROLLER SCHOOL

BART B. COBB, M.S.

ABSTRACT

An analysis of scores for an extensive battery of psychological tests administered to a large number of air traffic controller (ATC) trainees indicated that such tests can make a useful contribution in the selection of personnel for ATC training. Five test areas which emerged as most predictive of ATC school performance were: abstract reasoning, numerical ability, spatial relations, non-verbal analogies, and a specially designed test of air traffic control problems.

The results of previous studies by Brokaw[1] and Trites[2] have demonstrated that an assessment of aptitudes, attitudes, and other personal attributes of air traffic controller trainees can be used to effectively predict (a) school performance measures attained in a basic training course and (b) ratings by supervisors representing on-the-job performance evaluation of work accomplished at operational facilities one to five years after completion of the basic training course. Both Brokaw's investigation and the five-year follow-up study by Trites were based on data obtained for a sample of subjects who began their training in 1956. In contrast, the present report concerns an experimental testing program involving the more current classes of the Air Traffic Controller School.

The Federal Aviation Agency (FAA) provides for two basic Air Traffic Controller (ATC) courses at the Aeronautical Center in Oklahoma City. The instruction and training offered in either of the two basic eight-week courses is oriented toward qualification for an Air Traffic Control Specialist (ATCS) Certificate. However, the objectives of the courses are different. One provides training relevant to work required of controllers at an Air Route Traffic Control Center and hence is referred to as the Enroute or Center course. The other, designated as the Terminal or Tower course, focuses on training of personnel for terminal or airport tower duties. The student input is programmed so that a new class begins each of the courses every two weeks. Trainees receive instruction in academic subjects directly related to air traffic control and are required to perform laboratory work simulating that which will be experienced after assignment to an operational facility.

Participants are selected by FAA Regional Headquarters personnel from eligibility rosters supplied by the United States Civil Service Commission (CSC). No formal assessment of aptitudes by testing procedures in involved in the CSC screening of applicants or in the FAA's final choice of personnel for training. Instead, selection is based upon previous aeronautical experience, particularly Air Traffic Control, Air Traffic Communications, and/or piloting experience. In most instances, this qualifying experience has been attained while on active duty with one of the military services.

The objective of the present study was the determination of the most effective psychological test measures for the prediction of training-course-performance criteria for students in relatively recent classes of the Enroute Controller School. The Brokaw study represented a major

point of departure in formulation of plans for the present investigation. Brokaw's findings indicated the efficacy of certain tests administered during 1956 in the prediction of both training school and on-the-job criteria. However, the need for additional research became apparent when considering issues relevant to:

(1) Differences in method and content between the 1956 and the 1960-61 training course.
(2) The possibility of differences between aptitude levels for the 1956 and 1960-61 training groups.
(3) The non-availability of United States Air Force aptitude tests for extended use with a civilian population.
(4) The inclusion of psychological tests measuring factor areas not previously covered by the 1956 experimental battery.
(5) The desirability of additional evidence to substantiate the general findings reported by Brokaw and Trites.

PROCEDURE

The project was begun in August, 1960, with the administration of an extensive battery of commercially developed psychological tests to a class of incoming ATC students. After testing of the first three classes, revisions were completed to yield a stabilized battery. Beginning on September 23, 1960, and continuing through April 21, 1961, trainees of all incoming ATC classes were administered this uniform group of tests. In addition to psychological-test-performance data, information was also collected regarding the age, experience in job-related fields, and educational background of each student. Upon completion of training, four criteria of training performance were developed for every student in each class. A listing of all criterion, background, and test variables is presented in Table 1 with descriptions provided for those specific tests which the study later indicated as being most significant.

TABLE 1

LISTING OF VARIABLES INCLUDED IN THE ANALYSES

(Descriptions are provided for only those phychological test variables which regression analyses indicated as being significant in the prediction of training course criteria.)

Criterion Variables:

A. *The Combined Academic-Laboratory Grade Average.*
B. *The Pass-Fail Criterion.*
C. *The Scaled Objective "Personality" Rating.*
D. *The Scaled Subjective "Personality" Rating.*

Background Variables:

1. *Age When Tested.*
2. *Sum of Coded Relevant Experience.*
3. *Coded Educational Background.*

Psychological Test Variables:

Variables 4-8 represent subtests of the Bennett-Seashore-Wesman Differential Aptitude Test (DAT) Battery, Form A.

4. *DAT-Space Relations.* A 45-item test of ability to visualize objects and forms in two or three dimensions. The task, for each item, is to indicate how many of five depicted solid figures can be made from an unfolded pattern.
5. *DAT-Numerical Ability.* A 40-item-test presenting a series of relatively simple numerical problems. Provides a measure of "number" ability.
6. *DAT-Abstract Reasoning.* A 50-item test wherein the task is to indicate, for each item, which of a series of choices (figures) properly carries out a principle of logical development exhibited by a sequence of figures. The test provides a nonverbal measure of reasoning.
7. *DAT-Language Usage, Part II.*

8. *DAT-Mechanical Reasoning.*

   Variables 9 and 10 represent Part I and Part II of the Air Traffic Problems Test which was developed under contractural arrangement in 1952 by the American Institute for Research for the Civil Aeronautics Administration.

9. *Air Traffic Problems, Part I.* A 30-item test presenting highly simplified versions of Air Traffic Control situations. Good performance is not necessarily dependent on past ATC experience. Flight data displays are presented for several inbound aircraft, all flying the same speed and course, but at different altitudes and with different ETA's. Given a basic 5-minute time separation rule, the examinee must decide, for each item, whether or not sufficient time separation exists between certain aircraft to permit changes to certain specified altitudes.

10. *Air Traffic Problems, Part II.*

    Variables 11-28 represent 18 scales of the 480-item California Psychological Inventory (CPI) booklet. The scales provide a comprehensive survey of the individual from a social interaction viewpoint, and are referred to below in terms of the factors measured.

11. *CPI-Ac* (*Achievement via Conformance*).
12. *CPI-Ai* (*Achievement via Independence*).
13. *CPI-Cm* (*Communality*).
14. *CPI-Cs* (*Capacity for Status*). An index of an individual's capacity for status (not his actual or achieved status). The scale attempts to measure the personal qualities and attributes which underlie and lead to status.
15. *CPI-Do* (*Dominance*).
16. *CPI-Fe* (*Femininity*).
17. *CPI-Fx* (*Flexibility*).
18. *CPI-Gi* (*Good Impression*).
19. *CPI-Ie* (*Intellectual Efficiency*). The degree of personal and intellectual efficiency which a person has attained.
20. *CPI-Py* (*Psychological Mindedness*). The degree to which the individual is interested in, and responsible to, the inner needs, motives, and experiences of others.
21. *CPI-Re* (*Responsibility*).
22. *CPI-Sa* (*Self Acceptance*).
23. *CPI-Sc* (*Self Control*).
24. *CPI-So* (*Socialization*).
25. *CPI-Sp* (*Social Presence*).
26. *CPI-Sy* (*Sociability*). Outgoing, sociable, participative temperament.
27. *CPI-To* (*Tolerance*).
28. *CPI-Wb* (*Sense of Well Being*). A scale identifying persons who minimize their worries and complaints, and who are relatively free from self-doubt and disillusionment.

    Variables 29-40 pertain to twelve subtests of the California Test of Mental Maturity (CTMM, Advanced Form A, 1957 edition).

29. *CTMM-Immediate Recall.*
30. *CTMM-Delayed Recall.*
31. *CTMM-Sensing Right and Left.*
32. *CTMM-Manipulation of Areas.*
33. *CTMM-Opposites.*
34. *CTMM-Similarities.*
35. *CTMM-Analogies.* A 15-item test, wherein seven drawings of different objects are presented for each item. The first object has a definite relationship to the second which the student must recognize in order to identify, by analogy, the drawing among the last four which is similarly related to the third drawing.
36. *CTMM-Inference.* A 15-item test, wherein printed statements for each item present two premises. The student must select the logical conclusion, based on those premises, from the four possible alternatives offered.
37. *CTMM-Number Series.*
38. *CTMM-Numerical Quantity, Coins.*
39. *CTMM-Numerical Quantity, Arithmetic.*
40. *CTMM-Verbal Concepts.*

    Variables 41-47 are representative of seven tests of the Moran Repetitive Measurements (RPM) battery. The battery is composed of highly-speeded perceptual, coordination, and memory tests. All RPM scores used in the present study were measures of performance representing initial administration.

41. *RPM-A, Aiming.* This test measures the ability to carry out quickly and precisely a series of movements requiring eye-hand coordination. Specifically, the student's task is to place a stylus point through the center of randomly positioned printed circles of .08-inch diameter.
42. *RPM-FC, Flexibility of Closure.*
43. *RPM-NF, Numerical Facility.*
44. *RPM-PS, Perceptual Speed.*
45. *RPM-SC, Speed of Closure.*
46. *RPM-V, Visualization.*
47. *RPM-SM, Social Memory.* This test measures the ability of a student to remember faces or photographs. After studying a group of 16 photographs (faces) for one minute, the student must turn to a second sheet and indicate recognition of the 16 faces from among a group of 32 pictures. Only 16 of these faces are the same. The three parts that compose the test are similar but different photographs are involved in each part.

Criterion Variables

The four criterion measures were derived through analysis and treatment of data entered by training school personnel on each student's final "Evaluation of Performance" record. The evaluation form used for an Enroute student differed only slightly from that of a Terminal student. This record was prepared by senior and supervising instructors from information submitted by each of the student's instructors. The criteria developed and used in the study were:

A. The Combined Academic-Laboratory Grade Average,
B. The Pass-or-Fail-the-Course Criterion,
C. The Scaled Objective "Personality" Rating, and
D. The Scaled Subjective "Personality" Rating.

A greater number of relatively independent criteria could have been developed and utilized. However, due to the large number of psychological test measures scheduled for analysis and limitations of available computer programs for handling more than a specified number of variables, only the four criteria were included.

The Combined Academic-Laboratory Grade Average. This criterion measure represented a mean (or average) of two separate averages; one based on a summation of all examination grades achieved by the student at various training levels for seven different academic subjects; the other based on final performance grades for laboratory-simulated air traffic control work.

Examinations covering the academic areas are administered at both an intermediate stage of the ATC course and at the conclusion of training. In the latter instance, the tests constitute the Air Traffic Control/Specialist (ATCS) Certification Examination. A failing grade of 70 or less at either stage of training for any academic area normally entails a "retake" of the examination for the specific areas involved. Retention in the training program prescribes that no more than two failing "retake" grades be obtained by the student at either stage of training and a trainee is automatically eliminated if he obtains failing grades for three or more academic subjects on the first administration of the ATCS Examination.

The final "Evaluation of Performance" form submitted for each Enroute Air Traffic Control trainee also provides three final laboratory grades (or averages) reflecting performance in (1) Strip Writing, (2) Control Procedures for ATC Position as Assistant, and (3) Control Procedures for Position as Journeyman. An unsatisfactory grade of 70 or less for any one of these areas will normally result in elimination of the student from training. Laboratory work in the Terminal or Tower course consists of training in simulated Airport/Approach Control operations and laboratory performance is reflected by entry of a single and overall grade (or average) which must be 70 or more in order for the student to pass the course.

The Combined Academic-Laboratory (A-L) criterion score developed for the present study was based on the mean (or average) of the Academic Grade Average and the Laboratory Grade Average attained by each subject. This criterion score was rounded to overcome difficulties in dealing with fractional values. Data of incomplete records for students who were eliminated or withdrawn from training were

treated similarly, but the averages were based on fewer numbers of grades.

The Pass-Fail Criterion. All students successfully completing the training course were considered as pass cases; those unsuccessful as fails. Students whose withdrawal was necessitated by illness, death in the family, and so forth, were deleted from the study.

The Scaled Objective "Personality" Rating. This criterion rating was prepared only for those students successfully completing the training course. It was a normalized 9-point score, or stanine, based upon a classification, by psychologists, of statements recorded by the senior and supervising instructors for the first four areas of the "personality" profile section of the School's Final Evaluation from. The statements concerned (1) Performance Under Stress, (2) Attitude Toward Instruction, (3) Ability to Work With Others, and (4) Job Interest. It was Training School policy to prepare personality profiles only for those students who successfully completed the course. For derivation of the criterion measure, the statements were classified as being either positive (favorable) or negative (unfavorable) and the algebraic sum of such frequencies was obtained as a raw score for each subject. Two sets of such scores, derived independently by different psychologists for all cases used in the study, correlated .92. In each instance, an average of the two values was assigned the passing student and stanining techniques were applied to the data of each class to determine the final ratings.

The Scaled Subjective "Personality" Rating. This was a normalized 9-point score, or stanine, based upon a different and subjective assessment of the same informative statements previously dichotomized in the derivation of the Objective Rating. Each of two psychologists, working independently, examined and compared the contextual meaningfulness of comments submitted for each student with those made for each classmate of the student, derived a ranking of the student for each of the four areas, and summed the assigned rankings to obtain a raw score for each individual. Independently derived scores by the two psychologists correlated .90. Consequently, each student was assigned an average of the two scores and stanining procedures (by class) were employed to establish the final rating.

Background Variables

The three background variables included in the analysis were: Age, Sum of Coded Relevant Experience, and Coded Educational Background.

Age. Chronological age (rounded to nearest birthday) at the time of administration of the psychological test battery was recorded for each student.

Sum of Coded Relevant Experience. This variable represented the sum of codes assigned for amounts of previous experience in various related work areas. Source of the information was the "ATCS Registration Sheet" (Form AC-508) which each student completed immediately after arrival at the FAA Center. A code, ranging from 1 through 9, was assigned for each of ten different types of experience. Three of the types pertained to Communications, six to Air Traffic Control, and one to Ground Control Intercept (GCI). The ten codes were summed to constitute the "experience" variable used in the present analysis.

Coded Educational Background. Educational information obtained from Registration Form AC-508 was coded on a 9-point scale basis (but did not represent a stanining procedure). For example, a non-graduate of high school was coded as 1; a high-school graduate as 2, and so forth. A code of 9 would represent six years or more of college.

Psychological Test Variables

The test battery was heterogeneous, consisting mainly of commercially developed tests providing measures of various aptitudes, attitudes, and perceptual abilities. Some were highly speeded; others were "power" tests. The selection of tests comprising the battery was based on numerous considerations. Some of the tests were those which Brokaw had reported as being highly predictive of training-course and on-the-job criteria; several represented substitutes for the USAF tests which had been available for his 1956 study, and others were selected on the assumption that they either provided more comprehensive and reliable measures of certain areas or because they provided measures relevant to the investigation of potential for measuring new or different areas.

Exclusive of lunch and rest periods, approximately seven hours were required to administer the battery. Though participation in the research program was voluntary, approximately 98 per cent of all incoming students cooperated.

Complete Data Records

A complete record for a pass case consisted of data for 51 variables, as represented by the 4 training-course-performance criterion measures, scaled ratings for the 3 background variables, and 44 test performance scores. A complete record for a fail case was similar but consisted of data for only 49 variables due to the non-existence of data for the two personality ratings. Cases with incomplete records were deleted from most phases of the study.

Samples

The uniform battery of psychological tests was administered to incoming students of all Enroute and Terminal classes for the period September 23, 1960, through April 21, 1961. Terminal classes were usually less than half the size of the Enroute. Due to this disproportionate accumulation of Terminal cases, it was decided that investigative efforts of the present study would be focused on the development and validation of performance-prediction equations for students of the Enroute course and that a similar and separate study would be undertaken (pending the availability of cases for both an experimental and a validation sample) to determine the most valid predictors of performance for the Terminal course. However, proceeding on the assumption that similarities, as well as differences, characterized both the students and the training of the two courses, it was also decided that regression equations developed on data of the Enroute classes would be applied to determine their efficacy in the prediction of performance criteria for a single and sizable sample of Terminal cases. Consequently, the three samples described below were established for the present study. The number of cases listed for each sample represent only those for which complete testing and training records were available and constitute the minimum *N's* involved in any phase of the analyses.

Sample 1, or Experimental Sample. One hundred twenty-four cases (95 pass and 29 fail subjects) of five 1960 Enroute classes were designated as an experimental sample and scheduled for analyses aimed at the development of criterion prediction equations.

Sample 2, or Validation Sample. A second group, composed of 172 cases (136 pass and 36 fail subjects) of eight 1961 Enroute classes was established as a validation sample on which to test the prediction equations derived from an analysis of data for the experimental sample.

Sample 3, or Terminal Tryout Sample. One hundred forty-eight cases (137 pass and 11 fail cases) representing thirteen 1960-61 Terminal classes, constituted a sample on which the prediction equations developed for the Enroute classes could be tested for appropriateness in forecasting performance criteria for the Terminal course.

Regression Analyses

Two matrices of intercorrelations were computed for the experimental sample. The first, based on data for the entire group of 124 cases, was a 49 by 49 matrix which included no "personality" ratings. The second was a 51 by 51 matrix representing the addition of the two "personality" ratings and based on data of only the 95 pass cases. Using these data, and following an iterative multiple regression technique developed by Greenberger and Ward[3] and Bottenberg[4], the best combinations of predictors for the training-course-performance criteria of the Enroute classes were selected. The inclusion of variables in a prediction equation was terminated when no further significant increase in the magnitude of the multiple correlation could be obtained.

For the initial phases of the study, seven regression analyses were accomplished on data of the experimental sample. The first four were based on data of the combined pass plus fail cases and resulted in the development of two prediction equations for the Academic-Laboratory criterion, and two for the prediction of the Pass-Fail criteron. For each criterion, one of the two equations was based on data of only the psychological tests, while the other included a consideration of Age, Experience, and Education. The remaining three regression analyses accomplished during the initial phases of the study were restricted to data of only the

95 pass cases, included a consideration of both test and non-test variables, and resulted in the development of prediction equations for the two "personality" ratings and for the Academic-Laboratory Grade Average. At later stages of the study, and eighth and ninth regression equation were computed for the experimental sample using the data of only five selected tests. One equation was for the prediction of Pass-Fail status; the other for the A-L Grade. The rationale for selection of these five tests will be explained in subsequent sections of this report.

Application and Validation of Equations

The seven regression equations based on data of the experimental sample were applied, for validation purposes, to the data representing the second sample of Enroute cases. These equations, for prediction of Enroute course performance, were also applied to the data of Sample 3 to determine their efficacy in the prediction of performance for the Terminal course. Finally, the eighth and ninth equations, based on data of only five selected tests, were applied to Samples 2 and 3.

All of the prediction equations were applied using various weighting techniques for the tests of each composite. These included b-weights derived from true betas, integer weights proportionate to established b-weights, and weights proportionate to the inverse of the standard deviation of each test. Inasmuch as test performance measures were in raw-score form, this latter method was equivalent to a simple unit-weighting of standard scores and yielded correlates rather than values approximating each criterion.

RESULTS

The interrelationships of the four performance criteria developed for the present study are shown in Table 2. The correlation coefficients (of .68, .71, and .62) for Academic-Laboratory Grade (A-L) and the Pass-Fail (P-F) criterion were not sufficiently high to justify the two as being interchangeable. The correlations based on data of the pass cases only, indicate that the A-L measure is relatively independent of the two "personality" ratings which were, as expected, highly related.

TABLE 2

Intercorrelations of Criteria

Sample and Variable		*Acad. Lab.*	*Pass-Fail*	*Obj. Pers.*	*Subj. Pers.*
Experimental Sample					
Pass plus Fails:	Acad-Lab.	1.00	.68		
Passes Only:	Acad-Lab.	1.00		.30	.28
	Obj. Pers.	.30		1.00	.80
	Subj. Pers.	.28		.80	1.00
Validation Sample					
Pass plus Fails:	Acad-Lab.	1.00	.71		
Passes Only:	Acad-Lab.	1.00		.36	.41
	Obj. Pers.	.36		1.00	.81
	Subj. Pers.	.41		.81	1.00
Terminal Sample					
Pass Plus Fails:	Acad-Lab.	1.00	.62		
Passes Only:	Acad-Lab.	1.00		.36	.46
	Obj. Pers.	.36		1.00	.81
	Subj. Pers.	.46		.81	1.00

TABLE 3

Development and Application of Regression Equations for Prediction of Academic-Laboratory Grade Average and Pass-Fail Status of ATC En Route Students

	Multiple Correlations and Beta Weights Derived Via Regression Analyses of Data for Exp. Sample[a]							
	When Only Psych. Tests Are Considered				*When Psych. Tests Are Supplemented by Background Variables*			
Variables	*Regr. No. 1*		*Regr. No. 2*		*Regr. No. 3*		*Regr. No. 4*	
Criterion:								
A. Acad.-Lab. Grade Avg.	R = .66				R = .69			
B. Pass-Fail Status			R = .52				R = .50	
	Beta Wt.	*Validity*[c]	*Beta Wt.*	*Validity*[c]	*Beta Wt.*	*Validity*[c]	*Beta Wt.*	*Validity*[c]
Tests:								
5. DAT-Numerical Ability					.17	.41		
6. DAT-Abstract Reasoning	.33	.54	.33	.40	.31	.54	.32	.40
9. Air Traffic, Part I	.22	.40			.16	.40		
14. CPI-Cs, Capac. For Status			−.21	−.03				
19. CPI-Ie, Intell. Eff.	.25	.29			.21	.29		
26. CPI-Sy, Sociability	−.24	−.04			−.21	−.04		
28. CPI-Wb, Sense Well Being			.19	.20				
35. CTMM-Analogies			.21	.35			.21	.35
36. CTMM-Inference	.19	.42						
41. RPM-A, Aiming	−.14	−.10	−.17	−.15	−.15	−.10		
Background:								
1. Age When Tested					−.26	−.38		
2. Coded Sum of Experience							.21	.21
	Validation Sample[b]*: Correlation of Predicted Criterion Values With Actual*							
Criterion:								
A. Acad.-Lab. Grade Avg.	.49				.56			
B. Pass-Fail Status			.35				.38	

[a] EXP. Sample: N = 124; 95 pass plus 29 fail subjects of five 1960 En Route Classes.
[b] Val. Sample: N = 172; 136 pass plus 36 fail subjects of eight 1961 En Route Classes.
[c] Validity coefficients greater than .16 are significant at the .05 level and those greater than .21 are significant at the .01 level.

The data of Table 3, based on the combined pass and fail cases and representing the results of four regression analyses, shows the best combinations of variables for the Academic-Laboratory Grade and the Pass-Fail criterion. The data are arranged to facilitate comparison between prediction equations based only on data of the psychological tests and those based on inclusion of data of the psychological tests and those based on inclusion of data for the background variables. All of the regression equations validated when applied to the second sample of Enroute cases. As would be expected, the correlations between the predicted and actual criterion values for Sample 2 were somewhat lower than the multiple correlations obtained for Sample 1.

The first regression equation indicates that if composite scores were derived through conventional b-weighting of scores for six of the psychological tests for students of the experimental sample, these would correlate .66 with the Academic-Laboratory criterion. Data for regression number 3 show that when background variables are included, this multiple correlation for the experimental sample is only increased to .69. Validation coefficients of .49 and .56 were obtained when these prediction equations were applied to the data of Sample 2.

Results for regression analyses 2 and 4 are also shown in Table 3. For Number 2, a composite of five psychological tests was found to correlate .52 with the Pass-Fail criterion. The data for analysis 4 illustrate that when background variables are included the experience variable emerges to reduce the number of test predictors required but does not contribute toward an increase in the multiple. (Actually, the multiple is smaller by .02 due to the fact that variables were deleted unless they contributed toward a statistically significant increase in the multiple correlation.) When applied to the data of sample 2, validation coefficients of .35 and .38 were obtained for equations 2 and 4, respectively.

The predictor composites shown in Table 4 for regressions 5, 6, and 7 were developed from data of both test and non-test variables for the 95 pass cases of the experimental sample. The multiple of .62 between the Academic-Laboratory Grade and the four predictors shown for regression 5 compares favorably with the .69 previously obtained for the combined pass and fail cases of the sample. Values predicted with this equation for cases of the validation sample correlated .37 with the actual criterion values.

Multiples obtained by regressions 6 and 7 for the two "personality" ratings were of comparable magnitude. This might be expected inasmuch as a correlation of .80 was obtained between these two criteria. When applied to the validation sample, equation 6 yielded scores which correlated .28 with the objective "personality" ratings actually received. A validity coefficient of .20 was found for equation number 7. In the context of this study such validities may be considered quite low.

Regression Analyses Eight and Nine

Two additional regression equations (Number 8 for A-L, and Number 9 for the P-F criterion) were computed for the experimental sample using the data of only five selected tests. The previous findings of Brokaw and Trites, supplemented by the results of regression analyses 1 through 7 of the present study, indicated the significance of five major test areas for the prediction of various criteria. For example, Brokaw showed that space relations, abstract reasoning, and air traffic problems were significantly related to one or more of his criterion measures. In the report by Trites, tests of the abstract reasoning type and of the space relations or orientation type were found to be predictive of ratings by supervisors representing on-the-job performance evaluation of controllers five years after completion of the training course. In the present study, tests of numerical ability, space relations, abstract reasoning, air traffic problems and analogies were significantly related to one or more of the criterion measures. Due to this convergence of evidence, regression analyses 8 and 9 were accomplished to determine the predictive efficiency of a battery composed of only five such tests. Data pertaining to the two prediction equations appear in Table 5.

Regression analyses based on data of the five tests for subjects of the experimental sample yielded a multiple of .60 for the A-L criterion and .47 for Pass-Fail status. The values predicted by regression equation 8 for the Enroute validation sample correlated .54 with the

Academic-Laboratory Grade Averages actually attained. Similarly, values predicted by equation 9 correlated .40 with the actual Pass-Fail status.

Comparability of Different Equations and Weighting Techniques

In contemplating relevance of findings in the present study to establishment of an operational testing program, it seemed appropriate to determine the efficacy of each equation in predicting criteria other than that for which it was specifically designed. For example, to what degree can a prediction equation designed for the A-L criterion be used in forecasting pass-fail status, or vice versa? On the basis of assumptions pointed out earlier in this report,

TABLE 4

Development and Application of Regression Equations for Prediction of Criteria for Only the Pass Subjects of the ATC En Route Course

	Multiple Correlations and Beta Weights Derived Via Regression Analyses of Data for Exp. Sample[a]					
Variables	*Regr. No 5*		*Regr. No. 6*		*Regr. No. 7*	
Criterion:						
A. Academic-Laboratory Grade Avg.	R = .62					
C. Objective Personality Rating			R = .47			
D. Subjective Personality Rating					R = .43	
	Beta Wt.	*Validity*[c]	*Beta Wt.*	*Validity*[c]	*Beta Wt.*	*Validity*[c]
Test Variables:						
4. DAT-Space Relations	.20	.42				
6. DAT-Abstract Reasoning			.16.	.21		
19. CPI-Ie, Intellectual Eff.	.33	.39	.17	.15		
20. CPI-Py, Psych. Mindedness					.14	.09
35. CTMM-Analogies			−.16	−.10		
36. CTMM-Inference	.23	.33			−.16	−.07
47. RPM-SM, Social Memory			.24	.19	.29	.22
Background Variables:						
1. Age When Tested	−.25	−.35	.15	−.00		
3. Educational Background			−.35	−.29	−.35	−.29
	Validation Sample[b]*: Correlation of Predicted Criterion Values With Actual*					
Criterion Variables:						
A. Academic-Laboratory Grade Avg.	.37					
B. Objective Personality Rating			.28			
C. Subjective Personality Rating					.20	

[a] Exp. Sample: 95 pass students of five 1960 En Route classes.
[b] Val. Sample: 136 pass students of eight 1961 En Route classes.
[c] Validity coefficents exceeding .19 are significant at the .05 level and those exceeding .26 are significant at the .01 level.

TABLE 5

Development and Application of Regression Equations Based on Consideration of Data For Only Five Selected Tests

Variables	*Multiple Correlations and Beta Weights Derived Via Regression Analyses of Data for Exp. Sample*[a]			
	Regr. No. 8		*Regr. No. 9*	
Criterion:				
A. Academic-Laboratory Grade Avg.	R = .60			
B. Pass-Fail Status			R = .47	
	Beta Wt.	*Validity*[c]	*Beta Wt.*	*Validity*[c]
Tests:				
4. DAT-Space Relations	.03	.35	−.08	.22
5. DAT-Numerical Ability	.13	.43	.12	.32
6. DAT-Abstract Reasoning	.34	.54	.27	.40
9. Air Traffic, Part I	.19	.40	.05	.23
35. CTMM-Analogies	.11	.36	.22	.35
	Validation Sample[b]*: Correlation of Predicted Criterion Values With Actual*			
Criterion:				
A. Academic-Laboratory Grade Avg.	.54			
B. Pass-Fail Status			.40	

[a] Exp. Sample: N = 124; 95 pass plus 29 fail subjects of five 1960 En Route Classes.
[b] Val. Sample: N = 172; 136 pass plus 36 fail subjects of eight 1961 En Route Classes.
[c] Validity coefficients greater than .16 are significant at the .05 level and those greater than .21 are significant at the .01 level.

it also seemed advisable to apply these Enroute equations to data of the Terminal classes. In order to investigate the degree to which an equation could be simplified without appreciable sacrifice in prediction accuracy, all equations of the present study were applied to the data of samples 2 and 3 using four different weighting techniques. Justification for all these lines of investigation becomes obvious when considering implications relevant to possible implementation of a test screening program for the selection of ATC trainees. For example, it would be highly desirable to have a battery composed of *relatively few tests* for which a *minimum number* of *simplified equations* could be used to effectively predict *multiple criteria,* and if sufficiently appropriate, to use the *same equations* in making predictions for applicants of either Enroute *or* Terminal ATC training.

The four different weighting techniques involved in computing four expected values per student for each of the eight basic equations were:

(a) Conventional b-weights, each representing a conversion of the true beta according to the proportionate relationship of the standard deviation of the criterion variable to the standard deviation of each test predictor variable. These b-weights were computed on experi-

mental sample data and applied to other samples using values carried to four places beyond the decimal.

(b) Integer, or two-place, weights which were directly proportionate to the established b-weights described above.

(c) Two-place weights proportionate to the inverse of the standard deviation of each test predictor variable (irrespective of betas and b-weights). This method is equivalent to a simple summation, or unit-weighting, of test performance scores after conversion to standard-score form.

(d) Two-place weights identical to those described for method "c" but including no negative weights. In other words, any predictor variable negatively related to the criterion was deleted from the prediction equation.

Correlations were obtained between all sets of predicted values and for all sets of predicted values versus actual criterion scores for cases of the Enroute validation sample and for the Terminal trainees of sample 3. A portion of the resulting matrix is presented in Table 6. It shows comparative data derived through application of two of the different weighting techniques for each of the basic regression equations 1, 2, 3, 4, 8 and 9. For example, the designation "1a" is for data resulting from application of basic regression equation number 1 using the conventional b-weights described above under method "a". Similarly, "1c" denotes utilization of method "c" which is equivalent to a simple summation of predictor test performance measures in standard-score form. The lowest correlation between values predicted by the two methods for any of the basic equations is .91, indicating that the latter simplified method is sufficiently appropriate.

In assessing the efficiency with which each equation can be used to predict both the criterion for which it was designed and the other

TABLE 6

COMPARATIVE RESULTS DERIVED VIA APPLICATION OF SEVERAL DIFFERENT PREDICTION EQUATIONS

Prediction Equation No.	*Crit. of Prediction*	*Intercorrelations of Values Predicted By Different Equations For Cases of the Enroute Validation Sample*											*Predicted Vs. Actual Criterion Values*			
													En R. Validation Samp.		*Terminal Tr. Samp.*	
		1a	*1c*	*2a*	*2c*	*3a*	*3c*	*4a*	*4c*	*8a*	*8c*	*9a*	*9c*	*A-L*	*P-F*	*A-L*	*P-F*
1a	Acad-Lab		.98	.65	.55	.90	.91	.56	.49	.84	.78	.71	.72	.49	.44	.50	.44
1c	Acad-Lab			.60	.54	.87	.91	.47	.40	.77	.71	.62	.67	.47	.42	.47	.41
2a	Pass-Fail				.97	.70	.65	.71	.66	.73	.70	.80	.64	.42	.35	.39	.38
2c	Pass-Fail					.61	.59	.59	.56	.58	.57	.68	.53	.36	.29	.33	.33
3a	Acad-Lab						.98	.56	.48	.81	.77	.70	.65	.56	.49	.51	.44
3c	Acad-Lab							.48	.41	.79	.76	.66	.66	.54	.48	.51	.41
4a	Pass-Fail								.99	.75	.72	.81	.67	.40	.38	.37	.36
4c	Pass-Fail									.68	.66	.76	.62	.35	.34	.35	.33
8a	Acad-Lab										.97	.93	.89	.54	.46	.49	.42
8c	Acad-Lab											.91	.83	.53	.46	.44	.37
9a	Pass-Fail												.91	.51	.40	.42	.34
9c	Pass-Fail													.46	.35	.45	.31

criterion, it should first be noted that values predicted with the A-L equations not only correlated highly with values predicted, by other A-L equations, but also with values stemming from application of equations designed for the prediction of Pass-Fail status. Though all of the regression analyses were based on data of the Enroute experimental sample, the validity coefficients further indicated that each of the equations may also be used to effectively predict both the A-L and the P-F criterion scores of Terminal-course students. It should be emphasized that the validity coefficients shown for the P-F criterion in all tables are actually point-biserial correlations which would be considerably larger if converted to biserial correlations. The latter type more generally approximate the Pearson product-moment type of correlation used for all other validity coefficients.

The validities obtained for both versions of equations 8 and 9 should be regarded with particular interest. As pointed out earlier in this report, these equations were based on consideration of data for only five selected tests (all of which were shown by previous studies to be significantly related to training school or job-performance criteria). The data of Table 6 suggest that equation 8c should be given preferential treatment. For example, values predicted by the simplified weighting technique of 8c correlate .97 with those predicted by 8a. Further, the validities indicate that 8c may be used for the prediction of either the A-L or the P-F criterion for Enroute *or* Terminal trainees. Last, but most importantly, equation 8c is based on data of only five selected tests, yet its validity coefficients compare very favorably with those obtained for equations derived from a consideration of the entire test battery.

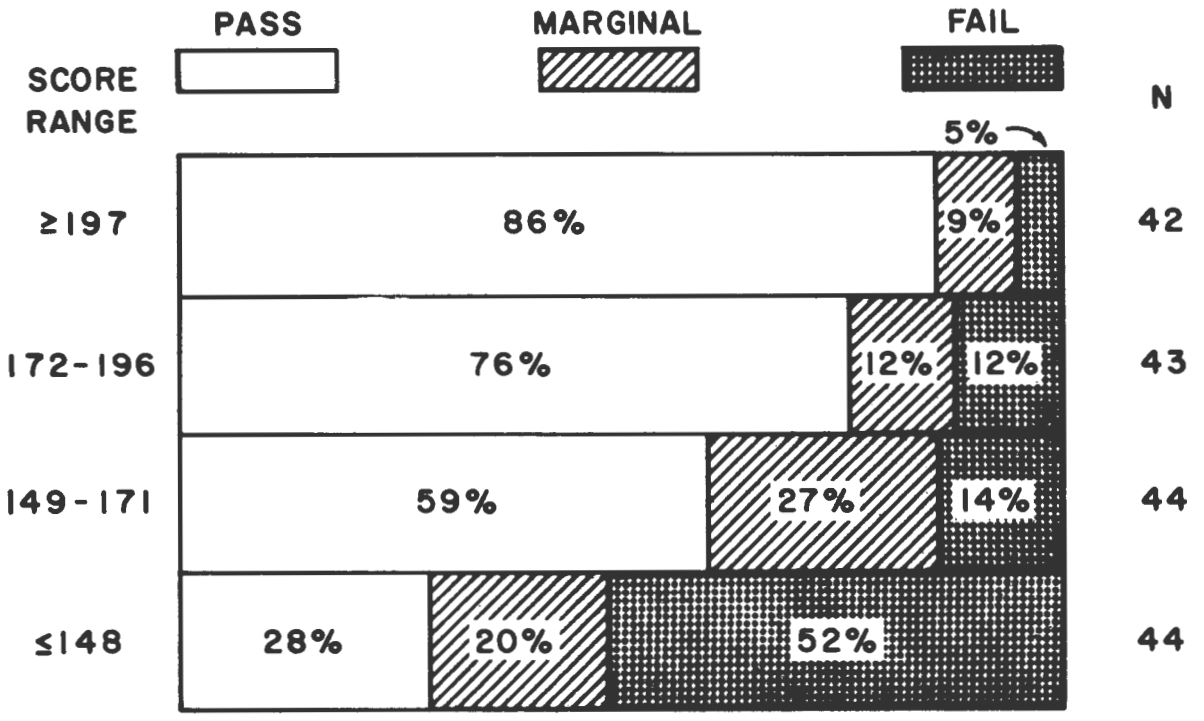

Figure 1 — Per Cent of Passing, Marginal, and Failing Students in the ATC Enroute Training Course By Approximate Fourths of Scores Predicted From Regression Equation 8c.

Predictive Efficiency of Equation 8c

It will be recalled that the application of prediction equation 8c was equivalent to a simple summation or unit-weighting of standard scores of performance attained by each student for DAT Numerical Ability, DAT Space Relations, DAT Abstract Reasoning, Air Traffic

Problems Part I, and CTMM Analogies. In each instance, the resulting score comprised a Correlate of, rather than a direct approximation or prediction of an actual criterion value. Figures 1 and 2 present a graphic illustration of the effectiveness with which such an equation was used to forecast training course performance criterion measures for students of the Enroute and Terminal Classes.

grades. In contrast, it should be noted that only 5 per cent of the 42 cases comprising the upper quarter of the distribution actually failed the course, whereas 86 per cent were non-marginal cases. A comparison of the data for the upper and lower intermediate quarters of the distribution further illustrates the potentialities of a test screening program for applicants of Enroute ATC training.

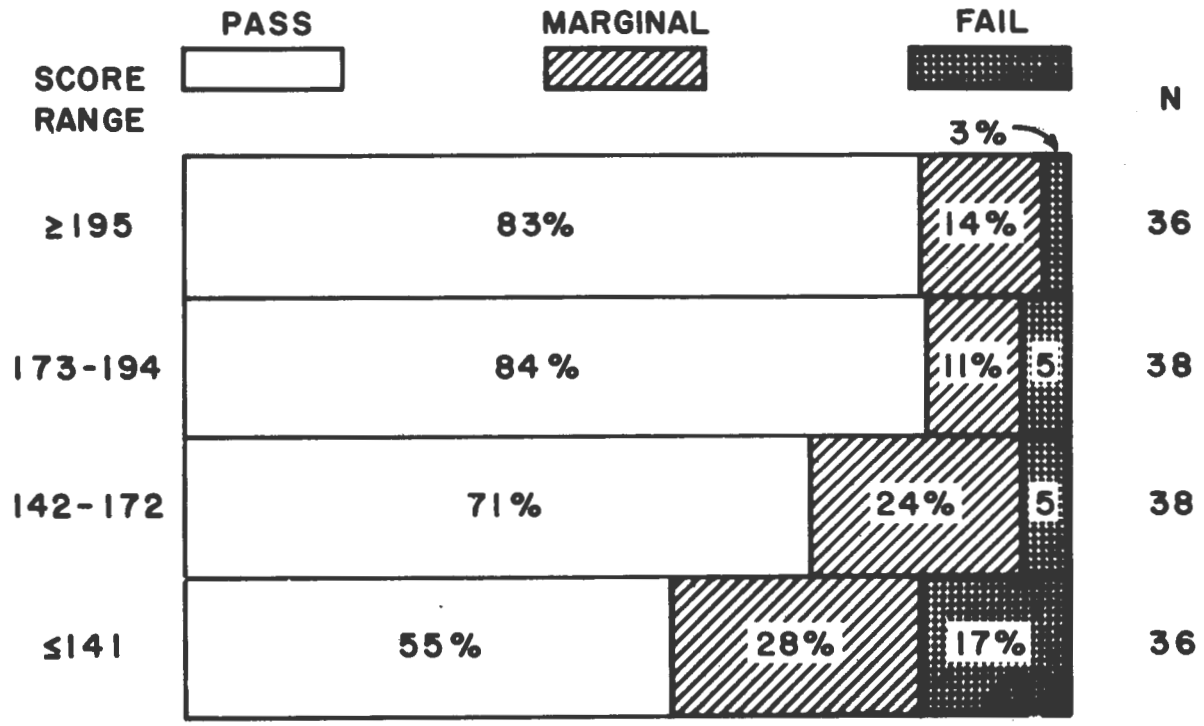

Figure 2 – Per Cent of Passing, Marginal, and Failing Students in the ATC Terminal Training Course By Approximate Fourths of Scores Predicted From Regression Equation 8c.

Enroute Validation Sample. Figure 1 is based on data of 173 students comprising eight 1961 ATC Enroute training classes. For each quarter of the distribution of scores predicted by equation 8c, percentages are shown for the passing (or "non-marginal"), "marginal", and failing students. "Marginal" trainees are those who passed the course but who are in the bottom 22 per cent of the Academic-Laboratory Grade distribution for the pass cases only. For the 44 cases representing the lower quarter of the distribution of predicted scores, 52 per cent failed the training course, 20 per cent passed with only marginal grades, and the remaining 28 per cent passed with relatively satisfactory

Terminal Trainee Sample. Even though equation 8c was developed specifically for prediction of performance criterion measures attained by students of the Enroute ATC training course, Figure 2 demonstrates the effectiveness with which it was applied to data of the Terminal classes of sample 2. Despite the fact that the attrition rate for these Terminal classes was only 7.4%, compared to 20.9% for the Enroute, Figure 2 illustrates that the majority of failing students have performed poorly on the five aptitude tests. Most of the marginal pass cases also appear in the lower two quarters of the predictor score distribution. Pending the availability of additional Terminal cases,

further studies will be undertaken to develop equations aimed specifically at prediction of performance criteria for the Terminal course. It is considered doubtful that equations can be developed which will yield a significant increase in predictive accuracy over that obtained with equation 8c.

DISCUSSION AND CONCLUSIONS

The present investigation indicated that a psychological battery can make a useful contribution in the selection of personnel for air traffic control work. Numerous equations derived through regression analyses of data for an experimental sample of Enroute trainees have proved highly valid in predicting the Pass-Fail status and the Academic-Laboratory Grade Average attained by students of either the Enroute or Terminal training course. In contrast, the variances of the two remaining and highly interrelated criteria (i.e., the Objective and Subjective Personality Ratings) appear to be relatively independent of the A-L Grade and P-F status, are less associative with the variances of the psychological tests included in the battery, and are therefore less predictable. Even though it may be assumed that both the A-L and P-F criteria are more appropriate for purposes of the present study, future research may demonstrate the usefulness of the "personality" ratings — not as criterion measures, but — as predictors of specific criteria pertaining to journeyman performance several years hence.

Results of this investigation supplement previous evidence submitted by Brokaw and Trites illustrating the potential effectiveness with which measures of five specific and major aptitudes, or factor areas, may be used in the screening of ATC applicants. Inasmuch as the predictions have been made for trainees who have already met qualification requirements involving an assessment of previous and job-related experience, implementation of such an operational test-screening program seems quite justified. However, it should be emphasized that the prediction equations which have been developed utilized scores obtained from commercially developed psychological tests. Because of susceptibility to compromise, these commercial instruments cannot be recommended for operational use. To overcome this difficulty, research is presently underway to determine the comparative efficiency of a corresponding battery of tests published and controlled by the United States Civil Service Commission.

Further studies are also contemplated regarding the assessment of additional factor areas. For example, a considerable portion of future research will be focused on development and evaluation of performance measures reflecting a student's ability to make rapid and accurate decisions while confronted with a continuous input of both diverse and relevant information.

REFERENCES

1. Brokaw, L. D.: School and job validation of selection measures for air traffic control training. WADC-TN-59-39, Personnel Laboratory, WADD, USAF, Lackland AFB, Texas, 1959.
2. Trites, D. K.: Problems in air traffic management: I. Longitudinal prediction of effectiveness of air traffic controllers. *Aerospace Med.*, 32:1112, 1961. Report No. 61-1, Civil Aeromedical Research Institute, Federal Aviation Agency Aeronautical Center, Oklahoma City, Okla., 1961.
3. Greenberger, M. H. and Ward, J. H. Jr.: An iterative technique for multiple correlation analysis. IBM Tech. Newsletter No. 12, International Business Machines Company, New York, N.Y., 1956.
4. Bottenberg, R. A.: An iterative multiple regression procedure using the IBM 650. WCLL-TM-60-3, Personnel Laboratory, WADD, USAF, Lackland AFB, Texas, 1959.

*U.S.GOVERNMENT PRINTING OFFICE: 1990-761-020/01176