PROBLEMS IN AIR TRAFFIC MANAGEMENT:

V. Identification and Potential of Aptitude Test Measures for Selection of Tower Air Traffic Controller Trainees

Bart B. Cobb, M.S.

Approved by

Stanley R. Mohler

STANLEY R. MOHLER, M.D.
DIRECTOR, CARI

Released by

M. S. White

M. S. WHITE, M.D.
FEDERAL AIR SURGEON

July 1965

FEDERAL AVIATION AGENCY

Office of Aviation Medicine
Civil Aeromedical Research Institute
Oklahoma City, Oklahoma

Problems in Air Traffic Management: V. Identification and Potential of Aptitude Test Measures for Selection of Tower Air Traffic Controller Trainees

BART B. COBB, M.S.

ABSTRACT

A study of over 200 Terminal Air Traffic Control Specialists indicated that their training performance could be well predicted by a composite of four aptitude tests measuring: numerical ability, non-verbal abstract reasoning, ability to solve simplified air traffic problems, and verbal abstract reasoning. Pre-employment experience directly related to air traffic control was also found to contribute to the prediction of training performance. Although the aptitude tests were related in the expected direction with a measure of job performance, the relationships were small. The aptitude of verbal abstract reasoning is both unique and important for the prediction of Terminal ATCS performance. Previous research has failed to identify such an aptitude as being of importance for the prediction of Enroute ATCS performance. With the exception of verbal abstract reasoning, the group of Civil Service Commission tests presently being used to select all ATCS trainees provides adequate measures of those aptitudes which this study identified as most significant for the prediction of Terminal performance and is still appropriate for selection of Terminal ATCS trainees.

THE RESULTS OF previous research conducted at the Federal Aviation Agency's (FAA) Civil Aeromedical Research Institute (CARI) have indicated the contribution which aptitude test measures can make in the selection of personnel for air traffic control training.

From the Selection Section, Psychology Branch, Federal Aviation Agency, Civil Aeromedical Research Institute, Aeronautical Center, Oklahoma City, Oklahoma.

Opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the views or the endorsement of the Federal Aviation Agency.

More specifically, it has been shown that aptitude test measures can be used to predict [2] school performance in centralized basic training courses, and [4] supervisory ratings of on-the-job performance at operational facilities one to five years after completion of basic training.

However, there are two different groups or types of air traffic controllers (ATC's) involved in the FAA's management of air traffic. The present study represents the first major effort toward identifying specific aptitude measures that may have potential for further improving the prediction of performance criteria for one of the two types of trainees, i.e., Terminal (or Tower) trainees. Much of the previous research has been focused upon analysis of data for trainees preparing for control work at any one of the 24 continental Air Route Traffic Control Centers. Such controllers are referred to as Enroute or Center controllers, while those working in Terminal or Airport Tower facilities are designated as Terminal controllers.

Separate basic training courses were formerly provided by the FAA Academy at the Aeronautical Center in Oklahoma City for personnel recruited for training in the different types of control. However, the Enroute course was discontinued in October of 1962 with further emphasis being directed toward on-the-job training at operational facilities. The Terminal course continued to be offered.

The Academy's two courses are characterized by both differences and similarities in regard to instruction and objectives. Common to both courses are seven academic or study areas. One of the seven pertains to Air Route Traffic Control (ARTC) which receives major emphasis in the Enroute Course. Another is that of Terminal Area Traffic Control (TATC) which receives greater emphasis

Reprinted from AEROSPACE MEDICINE
Vol. 35, No. 11, November, 1964. © Aerospace Medical Association

in the Terminal course. Instruction and training in either course is oriented toward qualification on an Air Traffic Control Specialist Certification Examination. The examination and qualification requirements are identical for trainees of either type. A considerable portion of training also consists of laboratory work simulating that which each type of trainee will experience after assignment to an operational facility.

Until mid-1962, all trainees were selected on the basis of medical and job-experience requirements—with attainment of the latter usually resulting from military service. No formal assessment of aptitudes by testing procedures was involved in the screening of applicants for training.

As part of a research program aimed at improvement of selection techniques, administration of an extensive experimental aptitude test battery to all Enroute and Terminal trainees arriving at the Academy was initiated in September of 1960. The objective was to determine the specific aptitude areas having the best potential for prediction of performance measures. Due to substantially larger class inputs, data in sufficient quantity to permit an analysis was first accumulated for the Enroute trainees. Prediction equations, which were developed through iterative multiple regression techniques, and validated on subsequent samples, revealed that the best composite for prediction of Enroute training-course performance criteria should include tests of numerical ability, space relations, abstract reasoning, air traffic problems and figure analogies. These Enroute prediction equations were also applied and found to be highly valid for Terminal trainees. Results for early phases of this research were reported at the 32nd annual meeting of the Aerospace Medical Association.

In subsequent studies, prediction equations based on combinations of aptitude tests found most valid for criteria of Enroute samples, have been applied to data of Terminal samples with repeated success. However, the present study—in contrast to those previously reported—constituted the first major detailed analysis actually based on Terminal-trainee data and aimed at determining the specific composite of test measures and respective weighting factors most useful for the prediction of performance criteria for Terminal controller trainees.

PROCEDURE

During the period September 1960 through July 1961, an extensive battery of commercially developed psychological tests, yielding 44 different scores, was administered to 212 incoming students of the Academy's Terminal ATC training course. Data regarding age, previous job-related experience, and educational background, were systematically obtained and coded for all 212 cases. Five performance criteria were developed for which the number of cases with complete data records varied in accordance with the attrition rates for various stages of training. Four of the criteria reflected performance for the basic eight-week course. The fifth represented post-training-course job performance at an operational facility. A listing of all aptitude, background, and criterion variables is presented in Table I with descriptions provided for those specific measures which the study later revealed as most useful for predictive purposes.

TABLE I. LISTING OF VARIABLES INCLUDED IN THE ANALYSES

(Descriptions are provided for only those variables which one or more of the regressions indicated as being significant in the prediction of training-course criteria.)

Criteria

A. Academic Grade Average for the Terminal ATC Training Course
B. Laboratory Grade Average for the Terminal ATC Training Course
C. Combined Academic-Laboratory Grade Average for the Terminal ATC Course
D. Pass-Fail Status for the Terminal ATC Training Course
E. Average of Supervisors' Ratings for Performance at an Operational Facility

Aptitude Tests

1. DAT Space Relations: A 45-item test of identifying solid figures figures that can be made from unfolding patterns.
2. DAT Numerical Ability: A 40-item test of arithmetic computational skill.
3. DAT Abstract Reasoning: A 50-item test wherein the task is to indicate which of a series of choices (figures) properly carries out a principle of logical development exhibited by a sequence of figures.
4. DAT Language II.
5. DAT Mechanical Reasoning: For each of 68 items, the student studies a drawing depicting the operation of a physical or mechanical principle to determine the correct statement describing a situational outcome resulting from application of the principle.
6. Air Traffic Problems, Part I: A 30-item test presenting highly simplified versions of air traffic control situations. Performance is not necessarily dependent upon past ATC experience. Given a basic 5-minute time separation rule, the student must determine whether aircraft may be permitted to change altitude without violating the rule.
7. Air Traffic Problems, Part II.

Variables 8-19 pertain to twelve subtests of the California Test of Mental Maturity (CTMM, Advanced Form A, 1957 edition).

8. CTMM Immediate Recall.
9. CTMM Delayed Recall.
10. CTMM Sensing Right and Left.
11. CTMM Manipulation of Areas.
12. CTMM Opposites.
13. CTMM Similarities.
14. CTMM Analogies: Seven drawings of different objects are presented for each of 15 items. The first object has a definite relationship to the second which must be recognized in order to identify, by analogy, the drawing among the last four which is similarly related to the third drawing.
15. CTMM Inference: For each of 15 items, the task is to comprehend printed statements presenting two premises; then select, from among four alternatives, the most logical conclusion.
16. CTMM Number Series.
17. CTMM Number Quantity, Coins: A 15-item test necessitating mental manipulation of interrelated amounts of money and number of coins.
18. CTMM Number Quantity, Arithmetic: Solving of fifteen word problems presenting simple arithmetical and mathematical situations.
19. CTMM Verbal Concepts.

Personality Scales

Variables 20-37 represent 18 personality scales derived from the 480-iten, California Psychological Inventory

20. CPI-Ac (Achievement via Conformance).
21. CPI-Ai (Achievement via Independence): This scale identifies those factors of interest and motivation which facilitate achievement in any setting where autonomy and independence are positive factors.
22. CPI-Cm (Communality).
23. CPI-Cs (Capacity for Status).
24. CPI-Do (Dominance).
25. CPI-Fe (Femininity).
26. CPI-Fx (Flexibility).
27. CPI-Gi (Good Impression).
28. CPI-Ie (Intellectual Efficiency).
29. CPI-Py (Psychological Mindedness).
30. CPI-Re (Responsibility).
31. CPI-Sa (Self Assurance).

32. CPI-Sc (Self Control).
33. CPI-So (Socialization): An indication of the degree of social maturity, integrity, and rectitude which the individual has attained.
34. CPI-Sp (Social Presence).
35. CPI-Sy (Sociability).
36. CPI-To (Tolerance).
37. CPI-Wb (Sense of Well Being).

Speeded Tests

Variables 38-44 represent the highly speeded tests of the Moran Repetitive Measurements (RPM) battery.

38. RPM Aiming.
39. RPM Flexibility of Closure.
40. RPM Number Facility.
41. RPM Perceptual Speed: The task is to detect and cross out all numbers in a row that are like a circled number at the beginning of each row.
42. RPM Speed of Closure.
43. RPM Visualization.
44. RPM Social Memory: Measures the ability to remember faces or photographs.

Background Variables

45. Age: Chronological age when tested.

Variables 46-49 represent data coded on a 9-point scale and variables 50-51 each represent a summation of similarly derived codes for several elements.

46. Educational Background: Coded years of formal education.
47. GCI (Ground Control Intercept) Experience.
48. Pilot Experience.
49. Air-to-Ground Radio Operator Experience.
50. Sum of Air Traffic Experience: Summation of codes for several types of previous experience directly related to air traffic control. Types included were VFR Tower, Approach Control Tower, Radar Approach Control, Center, GCA, and RAPCON.
51. Sum of Communications Experience: Sum of coded amounts of Station Experience, Ground-to-Air, and Point-to-Point communications experience.

Enroute Equation Score

52. Composite Score Predicted by Previously Developed Enroute ATC Equation.

Psychological Test Variables:—All of the 44 scores derived from the test battery were included in the analysis. The 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. Numerous considerations were involved in the selection of tests comprising the battery. Some had been demonstrated by previous research to have predictive potential for ATC trainees; a few were selected to supplement, or substitute for, previously used tests on the assumption that they would provide more comprehensive and reliable measures of specific areas, and several were included because of their apparent relevance to investigation of potential for additional aptitude areas.

Background Variables:—The background variables included: Age of Trainees; Coded Educational Background, and five variables representing coded amounts of varying types of previous job experience. Age was simply the chronological age (rounded to nearest birthday) of an individual at the time of test battery administration. Coded Educational background represented a 9-point scaling of educational attainment level. For example, a non-graduate of high school was coded as 1; a high school graduate as 2, and so forth. Six or more years of college was coded as 9.

Amounts of experience in each of twelve different types of related work areas were also coded on a 9-point scale. For analysis purposes, however, the number of experience variables was reduced from twelve to five by development of two summary variables which included all but three of the individual types of experience. The first summary variable, designated as Sum of Air Traffic Experience, represented a total of coded amounts of experience directly related to control of air traffic. The second, named Sum of Communications, represented a summation of codes for communications experience not directly related to air traffic control. Previous research had indicated appropriateness of such grouping procedures. Codes for Ground Control Intercept (GCI), Pilot Experience, and Air-to-Ground Radio Operator Experience constituted the three remaining and nonsummary experience variables included for analysis.

Criterion Variables:—The four training-course performance variables were: Academic Grade Average; Laboratory Grade Average; the Combined Academic-Laboratory Grade Average, and Pass-Fail status. All were derived from data entered by training school personnel on each student's final "Evaluation of Performance" record.

The Academic Grade Average was based on a summation of all achievement and examination grades attained at various training stages for seven different academic study areas; the Laboratory Grade Average represented the final performance grade assigned each individual for laboratory-simulated air traffic control work, and the Combined Academic-Laboratory Grade Average was simply an arithmetic mean of the two averages. All such criterion scores were rounded to overcome difficulties of dealing with fractional values.

For the Pass-Fail criterion, all students successfully completing the training course were considered as pass cases; those unsuccessful, as fails. Satisfactory completion of the course and ATCS certification demanded ultimate attainment of a satisfactory grade at every training stage for every study area. At any stage of training, failing grades of less than 70 in three or more of the seven subject areas or in laboratory work, generally resulted in automatic elimination from training. Students having failing grades in no more than two of the subjects were normally permitted a retake of pertinent examinations in order to avoid elimination. It should be pointed out that a few of the fail cases included in the present study actually have higher grade averages than do some of the passes. This is due to the fact that they might have done quite well in some of the subjects other than those which they failed, whereas a few of the pass students were able to attain only minimal grades in all subjects. Previous studies have shown that the failure rate for the Terminal course is generally less than ten per cent. From a research standpoint, the low failure rate contributes toward additional significance of the grade averages as performance criteria.

The Average Supervisory Rating represented a mean of all performance evaluations submitted by work supervisors of an individual at an operational facility. The use of a special rating form permitted each supervisor to rate a controller, on a five-point scale, for each of a series of work-related and proficiency items.[6] For the present study, these element ratings were coded on a five-point scale, and a weighted average was computed to represent the overall rating by each supervisor. In most instances, an individual was rated by more than one supervisor and the assigned criterion value represents an average of all such ratings rendered. Unfortunately, the usefulness of this criterion in the present study may be questioned because of probable biasing factors relevant to data collection. More specifically, several individuals who had been administered the aptitude tests and whose records were included in other phases of the study, had left the FAA before the facilities received a request for performance evaluations. Ratings, made in retrospect, were submitted for only a portion of such cases. The data obtained for these individuals were excluded from the study on the presumption that mere acknowledgement of their separated status might have tended to evoke artificially lower and more critically-reflective ratings from supervisors. On the other hand, the non-availability or exclusion of data for the employment-terminated cases forces a consideration of probable attenuation effects and severely limits the definitiveness with which results of the present study may be interpreted.

Samples:—A total of 212 incoming students for the Terminal course were administered the battery of aptitude tests during the period of September 1960 through July 1961. Two samples of comparable size were established. The first, based on 99 pass and 7 fail cases of eight classes tested during the earlier portion of the time period, was designated as the Experimental Sample. These data were scheduled for analyses aimed at development of prediction equations for the various criteria. Data for the remaining 96 pass and 10 fail cases constituted the Validation Sample on which to test the predictive efficiency of the equations.

Regression Analyses:—After deletion of 5 cases which had incomplete data records, two matrices of intercorrelations were computed for the Experimental Sample. The first matrix was based on combined data for 95 pass and 6 fail cases and included intercorrelations of all psychological, background, experience, and criterion variables. An additional variable was included for purposes of comparison and reference only. It represented the composite scores derived through application of an equation which previous research had shown to be most predictive of criteria for trainees of Enroute ATC classes.

The second matrix was based on data of only the 95 pass cases of the Experimental Sample and excluded the Pass-Fail criterion variable. The correlations within each matrix, except those involving the Supervisory Rating criterion, were based on a constant or common N of cases.

Using these data and following an iterative multiple regression technique developed by Greenberger and Ward[3] and Bottenberg,[1] the best combinations of predictors for the performance criteria were selected. Inclusion of variables in a predictor equation was terminated when no further significant increase in the magnitude of the multiple correlation could be obtained.

For each criterion, efforts were first focused on development of a series of equations wherein selection of variables for the predictor composite was successively restricted to a consideration of different groupings of only the psychological measures. It should be pointed out that: 19 of the 44 test measures reflected performance on rather general and typical aptitude tests, 18

TABLE II. TRAINING-COURSE-CRITERIA VALIDITY COEFFICIENTS FOR EACH VARIABLE INCLUDED AS A PREDICTOR IN ONE OR MORE REGRESSION EQUATION

	Experimental Sample †				Validation Sample †			
Variables Significant in One	For Combined 95 Pass and 6 Fail Cases			90 Passes Only	For Combined 96 Pass and 10 Fail Cases			74 Passes Only
Or More Prediction Equation	Acad.	Lab.	A-L.	Supv.	Acad.	Lab.	A-L.	Supv.
Aptitude:								
DAT Spatial Ability	.19	.10	.16	.03	.33	.07	.17	.02
DAT Number Ability	.37	.12	.31	.06	.51	.17	.37	.19
DAT Abstract Reasoning	.34	.30	.38	—.14	.53	.34	.45	.15
DAT Mechanical Reasoning	.18	.00	.13	—.19	.19	—.23	.01	.06
Air Traffic Problems I	.40	.29	.42	.00	.38	.28	.35	.13
CTMM-Analogies	.30	.03	.22	—.04	.18	.04	.10	.06
CTMM-Inference	.43	.32	.44	.11	.21	.30	.20	.06
CTMM-Number Quantity, Coins	.34	.39	.42	.06	.38	.17	.28	—.02
CTMM-Number Quantity, Arith.	.34	.20	.35	.19	.52	.26	.43	.25
Personality:								
CPI-Ai, Achievement Ind.	.26	.09	.21	.08	.18	—.07	.09	.12
CPI-So., Socialization	.46	.25	.43	.31	—.02	.17	.06	—.07
Speed Tests:								
RPM-PS, Perceptual Spd.	.18	.26	.26	—.01	.10	.10	.05	.26
RPM-SM, Social Memory	—.03	.10	.02	—.20	.17	.07	.14	—.22
Background:								
Sum of Air Traffic Exper.	.01	.28	.14	—.12	.19	.38	.26	.16
Intercorrelations of Criteria								
Acad.: Academic Grade Avg.		.45	.87	.44		.49	.76	.23
Lab.: Laboratory Grade Avg.	.45		.82	.32	.49		.89	.12
A-L: Acad.-Lab. Grade Avg.	.87	.82		.46	.76	.89		.22
Supv.: Supervisor's Rating Avg.	.44	.32	.46		.23	.12	.22	

† For the experimental sample, all supervisory-rating correlations (r's) are based on a common N of 90 cases and those involving other criteria on a common N of 101. In contrast to the common-N procedure, every r for the validation sample is based on the maximum N of cases with data available for each pair of variables. For r's involving supervisory ratings, the N was or approached 74; for other criteria, the N usually was 106 or approached 106. The minimum r's for the respective significance levels of .05 and .01 are: $\geqq .20$ and $\geqq .27$ for $N = 90$; $\geqq .23$ and $\geqq .30$ for $N = 74$; and $\geqq .19$ and $\geqq .25$ for $N = 101$ or $N = 106$.

were actually personality scales, and 7 represented scores on very highly-speeded perceptual-type tests.

Data for the respective background and experience variables were also quite distinctive. Therefore, additional series of regressions, representing successive considerations of age, education, and previous-job-experience measures, were accomplished to determine the supplemental predictive contribution of these variables when used in combination with various groupings of psychological test measures.

A total of 38 different regressions were accomplished in only the initial phase of the study and 18 additional, but simplified, equations were developed in subsequent phases. Validity coefficients for the Experimental Sample are presented in the first portion of Table II for each variable that emerged as a significant predictor in one or more of the 56 equations. Intercorrelations for four of the five criteria are also shown. (The latter portion of the table presents similar data obtained for the Validation Sample during later phases of the study.)

The ultimate objective of these extensive analyses was determination of the most practical and valid equation(s) for prediction of the various criteria of Terminal-Trainee performance. Within the limitations of this report, presentation and comparison of data for all 56 equations is not feasible. However, several aspects concerning their general nature will be described in conjunction with the subsequent summary of developmental and predictor-composite selection procedures.

As a first step toward selection of the most useful prediction equation, 7 of the initial 38 regressions were discarded at the outset. The 7 were for the pass-fail criterion and represented analyses based on intercorrelations of the first matrix computed for the Experimental Sample. Inasmuch as the data of only 6 fail cases were included in the total of 101 cases upon which all correlations of the matrix were based, the 7 pass-fail equations were considered as unreliable and therefore deleted.

The remaining 31 regressions were carefully reviewed to ascertain similarities and differences in data for those exhibiting the best predictive potential. After a comparison of the multiple correlations, the variables of each predictor composite, and the beta weights of the respective variables, a total of 14 regression equations were selected for retention and further study. The conventional b-weights for these initially retained regression equations were next computed in order that they might be applied to the data of the Validation Sample. The b-weight for each variable of a predictor composite represented a conversion of the beta in accordance with the proportional relationship of the standard deviation (S.D.) of the predictor variable to that of the criterion variable.

Development of Simplified Prediction Equations:—Before initiating validation procedures, two additional groups of equations were developed. The first such group consisted of 14 equations designed to test the assumption that no appreciable loss in predictive accuracy would result through use of more simplified weights substituted for the conventional and more complex system of b-weights. In order to establish integer weights for these 14 equations, two sets of independently-derived single-digit weights were first computed. Based on a consideration of data for variables of each predictor composite, one of the digital weights was grossly proportionate to the inverse of a predictor variable's standard deviation and the other was grossly proportionate to its conventionally established b-weight. Except for rounding procedures to avoid fractions, the final assigned integer weight for each variable included in any of the 14 equations actually represented an average of the two independently-derived digital weights. These equations, were numbered as 1A through 14A, with 1A corresponding to previous regression 1; 2A with 2, and so forth.

A final group of 4 equations, representing even further simplification, was developed. These were designated as "Summarily-Derived" equations because of the method underlying their development. No variables were included in the predictor composites except those which preceding equations had indicated to be most potentially useful. The weights for the variables of each predictor composite were derived under an arbitrarily imposed restriction that no weight would exceed the magnitude of 4. The b-weights and integer weights for the most significant of the previous equations were first converted, on a grossly proportionate basis, to weights within the restricted range of 1 through 4. The procedure resulted in establishment of two new sets of weights for each composite. After comparing all sets, weights were adjusted to establish four new "Summarily-Derived" equations, numbered as 15, 16, 17 and 18, which will be presented and discussed in subsequent portions of this report.

Application and Validation:—For purposes of validation, the 14 initially retained regression equations, the 14 corresponding and simplified equations, and the 4 summarily-derived equations were all applied to data representing the second sample of Terminal cases. For purposes of permitting a more reliable and comparative assessment of the contrasting weighting techniques, the equations were also applied to data of the Experimental Sample. To facilitate further comparison, and to determine the efficacy of each equation in predicting criteria other than that for which it was specifically designed, all predicted values were correlated with the actual data of every criterion variable.

RESULTS

The results indicated that no equation was valid for the prediction of Average Supervisory Rating. As previously explained, this was the criterion variable for which the data was suspect of being attenuated. For the remaining criteria, all equations validated when applied to the data of the second sample. In addition to the criterion for which designed, each equation also exhibited relatively good potential for the prediction of multiple criteria. Excluding Supervisory Rating, all validities and correlations were statistically significant at the .01 level.

Prediction results for the 14 simplified equations were highly comparable to those obtained for the corresponding and more complex b-weighted regression equations. In many instances, the corresponding correlations were identical and there were no differences exceeding .04, thus supporting the contention that no appreciable sacrifice in predictive accuracy would result through use of the more practical and simplified equations.

Of the 14 simplified equations, the best potential for the prediction of multiple criteria was exhibited by three of the four equations originally formulated for the prediction of the Academic Grade Average. These equations, numbered as 2A, 3A and 4A, constituted the first of two groups retained for further comparative study in determination of the most practical valid equation. As further evidence regarding the efficacy of simplified weighting techniques, the results stemming from application of the four summarily-derived equations also warranted retention and consideration of this entire group for the selection of the best equation(s).

Comparison of Seven Equations:—The data of Table III permit a comparison of equations 2A, 3A and 4A with the four summarily-derived equations in terms of predictor-variable weights, validities, and other correlations. Data for the Enroute equation are also presented.

In reviewing the contents of the seven Terminal equations, the predominance of four predictor variables should first be noted. Six of the predictor composites include the CTMM-Inference test and all seven equations include the tests of Air Traffic Problems, CTMM-Numerical Quantity-Coins and CTMM-Numerical-Quantity-Arithmetic. These four tests apparently constitute a common "core," providing measures of those aptitudes most significantly related to Terminal-trainee performance.

Equation 2A has the most distinctive and complex composite. Even though it includes all of the "core" tests except Inference, it is the only retained equation embodying two or more other variables. It includes: the two CPI personality scales of Achievement and Socialization; the highly speeded RPM Perceptual Test, and the Sum of ATC Experience.

TABLE III: COMPARISON OF THE THREE BEST PREVIOUSLY DEVELOPED EQUATIONS WITH THE FOUR EQUATIONS BASED ON SUMMARILY-DERIVED COMPROMISE WEIGHTS

	Digital Factor Weights Of Predictor Variables							1/SD
	Three Most Practical Of Previously Developed Equations			Four Additional & Newly Developed Equations With Summarily-Derived Weights †				Wts. of Original Enroute
Predictor Vars.	2A	3A	4A	15	16	17	18	Equation
Aptitude:								
DAT Spatial								.06
DAT Number								.13
DAT Abst. Reas.								.14
Air Traf. Prbs. I	2	1	2	2	2	2	1	.18
CTMM-Analogies								.49
CTMM-Inference		4	5	4	4	4	1	
CTMM-NQ-Coins	1	1	1	2	2	2	1	
CTMM-NQ-Arith.	4	3	3	3	3	3	1	
Personality:								
CPI-Ai, Achiev.	3							
CPI-So, Social.	3					2	1	
Speed Tests:								
RPM-PS, Perc. Spd.	1							
Background:								
Sum of ATC Exp.	2		1		3			
Criterion Vars	Experi. Sample: Correlation Of Predicted Vs. Actual Values ‡							
Academic Grd.	.63	.54	.53	.53	.48	.63	.63	.46
Laboratory Grd.	.48	.36	.43	.39	.48	.43	.44	.23
Acad.-Lab. Grd.	.65	.53	.57	.55	.57	.63	.64	.42
Pass-Fail	.48	.39	.42	.42	.40	.50	.52	.42
Avg. Supv. Ratg.	.21	.12	.09	.10	.08	.22	.22	—.01
	Valida. Sample: Correlation Of Predicted Vs. Actual Values ‡							
Academic Grd.	.42	.51	.51	.52	.50	.46	.44	.53
Laboratory Grd.	.34	.36	.42	.36	.45	.39	.37	.25
Acad.-Lab. Grd.	.37	.44	.45	.44	.46	.42	.41	.39
Pass-Fail	.31	.28	.31	.29	.32	.35	.35	.27
Avg. Supv. Ratg.	.29	.19	.21	.17	.22	.14	.12	.16
	Combined Samples: Correlation Of Predicted Vs. Actual Values ‡							
Academic Grd.	.51	.52	.52	.52	.49	.53	.52	.49
Laboratory Grd.	.37	.35	.41	.36	.46	.39	.38	.23
Acad.-Lab. Grd.	.44	.45	.47	.45	.48	.46	.46	.38
Pass-Fail	.37	.32	.36	.34	.36	.41	.42	.33
Avg. Supv. Ratg.	.27	.17	.16	.15	.15	.19	.18	.09

† Weights were assigned variables of the four additional equations, 15-18,, under an arbitrarily imposed restriction that none would exceed 4, yet be somewhat proportionately related to the most effective of previously used sets of weights.

‡ All r's are based on maximum N's. Most r's for Supv. Rating are based on N's of 90 for Exp., 74 for Val., & 164 for combined, with most r's for other criteria based on 106, 106, and 212. The minimum r's for the respective significance levels of .05 and .01 are: For N = 90, $\geq$ .20 & $\geq$ .27; N = 74, $\geq$ .23 & $\geq$ .30; N = 164, $\geq$ .16 & $\geq$ .21; N = 106, $\geq$ .19 & $\geq$ .25, and N = 212, $\geq$ .13 & $\geq$ .18.

The four "core" tests are included, but weighted differently, in the six remaining equations. These four tests alone comprise the predictor composites for equations 3A and 15; are supplemented only by the CPI-Socialization scale to yield composites for 17 and 18, and are supplemented only by Sum of ATC Experience to form the predictor groups of 4A and 16.

The correlations of actual criterion values with predicted composite scores (are shown in the lower portion of Table III and) reflect the highly comparable degree with which each of the equations may be used to forecast multiple criteria. However, a comparative study reveals certain differences relative to selection of the most practical equation.

When correlations involving Supervisory Rating Average are excluded, the data indicate slightly superior predictive efficiency for equations 3A, 4A, 15 and 16. All of these equations are remarkably similar in several respects. Based on a comparison of all arrays of correlations for the Validation Sample with those of the Experimental Sample, the differences are smaller for these equations than for the remaining four. This exhibition of less shrinkage in correlation from sample to sample is indicative of greater consistency-of-measurement for these four equations. It should also be noted that, even though different patterns of weights are utilized, the predictor composite of 3A corresponds to that of 15 and the composite of 4A matches that of 16. However, the summarily-derived weights for equations 15 and 16 yield slightly better validation results.

Final Selection of Two Terminal Equations:—Equations 15 and 16 may therefore be designated as the most useful of all Terminal equations for the prediction of multiple criteria. Equation 15 is based on scores of only the four "core" tests. Except for the inclusion and weighting of ATC Experience, equation 16 is identical to number 15 in every respect and the two equations yield highly comparable results for the Validation Sample. (see Table III.) Composite scores based on the four test variables of equation 15 correlate slightly higher with Academic Grade than do the scores predicted by number 16. For the remaining criteria, the correlations favor the latter. However, the only appreciable contribution of ATC Experience as a supplementary variable appears to be in the prediction of the Laboratory Grade Average.

Differences Between Terminal and Enroute Equations:—Even though a comparison of all corresponding data indicates the superior effectiveness with which the four "core" tests of the Terminal composite may be used for the prediction of multiple criteria, the correlations based on application of the Enroute equation are very substantial and, in some instances, approximate those obtained with the specially developed Terminal equations. In view of previous research, wherein the predictive efficiency of the Enroute equation has been repeatedly demonstrated for Terminal classes, its validation on the present samples was anticipated. The emergence and validation of a predictor composite so contrastingly different from that of the Enroute equation had not been anticipated. With exception of the Air Traffic Problems Test, the predictor variables for the two types of equations are indeed different. Despite these apparent differences, however, a recent unpublished analytic study has illustrated the high degree of similarity between the two groups of tests in terms of factorial content.

A non-verbal abstract reasoning, or induction factor, and number facility are heavily represented in the measures provided by either composite. Factorially, the Abstract Reasoning and Number Ability tests of the Enroute composite are characterized by less complexity and, as their names might imply, exemplify these two factors. In contrast, all tests selected for the Terminal composite are factorially more complex and, in addition to the two factors most pertinent to Enroute prediction, embody a third factor which is presumably responsible for its superior predictive efficiency for Terminal trainees. This additional factor seems to be a syllogistic reasoning, or secondary induction, factor in-

TABLE IV. COMPARATIVE PREDICTIVE EFFICIENCY OF EQUATIONS 15 AND 16 AND THE PREVIOUSLY DEVELOPED ENROUTE EQUATION

Terminal Equation 15 Based On Scores Of Only Four Aptitude Tests					Terminal Equation 16 Based On Four Aptitude Tests And Sum Of ATC Experience					Enroute Equation Based On Five Aptitude Tests Selected By Previous Research				
Experimental Sample														
Per Cent of Students In Each Approximate Fourth Of The Predicted-Score Distribution Who Failed, Marginally Passed, or Non-Marginally Passed The Terminal ATC Training Course														
Score Range	% Fail	% Marg.	% Pass	N	Score Range	% Fail	% Marg.	% Pass	N	Score Range	% Fail	% Marg.	% Pass	N
≧ 182		11	89	27	≧ 175		4	96	26	≧ 1895		11	89	27
115-131	4	20	79	25	156-174	4	24	72	25	1695-1894	4	16	80	25
98-114	4	24	72	25	139-155		26	74	27	1415-1694	8	32	60	25
≦ 97	18	30	52	27	≧ 138	23	31	46	26	≦ 1414	15	26	59	27
Validation Sample														
Per Cent Of Students In Each Score Range Grouping (Based On The Experimental Sample) Who Failed, Marginally Passed, or Non-marginally Passed The Terminal ATC Training Course														
Score Range	% Fail	% Marg.	% Pass	N	Score Range	% Fail	% Marg.	% Pass	N	Score Range	% Fail	% Marg.	% Pass	N
≧ 132	3	9	88	34	≧ 175	4	18	78	28	≧ 1895	3	17	80	40
115-131	6	28	66	32	156-174	4	25	71	24	1695-1894	5	21	74	19
98-114	8	23	69	13	139-155		11	89	18	1415-1694	18	9	73	22
≦ 97	23	35	42	26	≦ 138	23	31	46	35	≦ 1414	17	46	37	24

volving verbal abstract reasoning. It is most highly represented in the CTMM-Inference Test. The question of why verbal abstract reasoning is of more importance in Terminal training performance than for Enroute performance remains a topic for additional research.

Comparison of Equations in Terms of Predictive Effectiveness:—The data of Table IV illustrate the effectiveness with which the composite scores of Terminal equations 15 and 16, and of the Enroute equation, may be used to forecast performance in the Terminal ATC Training Course. Based on data of the Experimental Sample, cutting scores (or score range groupings) were established to represent the upper, two intermediate, and lower fourths of the distribution of composite scores derived by each equation. For each quarter of the composite score distribution, the percentages of failing, "marginal," and "non-marginal" passing cases are shown. "Marginal" trainees are those who passed the course but with grades so low that they constitute the bottom 23 per cent of the distribution of Combined Academic-Laboratory Grade Averages after exclusion of training-course failures. For validation of each equation, the cutting scores (established on the Experimental Sample) were imposed, without modification, on the distribution of composite scores computed for cases of the second sample. Results for this Validation Sample, which are presented in the lower portion of Table IV, illustrate the effectiveness with which these score range groupings may be used to forecast actual performance for the training course.

A comparison of results based on this particular type of validation procedure indicates slightly superior potential for equation 15 over 16, and that both Terminal equations are better than the Enroute equation for forecasting purposes. Of the trainees having the higher scores predicted by the four-test composite of equation 15, only 3 per cent were fails and only 9 per cent were "marginals." The remaining 88 per cent were passes. In contrast, 23 per cent of the cases comprising the lowest category of aptitude scores were failures; 35 per cent were "marginals," and only 42 per cent were "non-marginal" passes. In other words, only 12 per cent of the trainees represented in the top category failed to attain satisfactory "non-marginal" status for the training course, whereas 58 per cent of those in the lowest aptitude index grouping were unable to do so.

The three graphs of Figure 1, which are based on a second and more liberal validation procedure, permit a further comparison of the equations in terms of performance-forecasting efficiency. By this method, the cutting scores for each equation were established by simply dividing the predicted score distribution of the Validation Sample into approximate fourths, *independently* of Experimental Sample data. The prediction results shown in the top graph are based on the composite of four aptitude tests and the ATC Experience variable used in equation 16; the next indicates the highly comparable effectiveness of equation 15, which excluded ATC Experience, and the lower graph presents results obtained with the Enroute equation. The correlation of

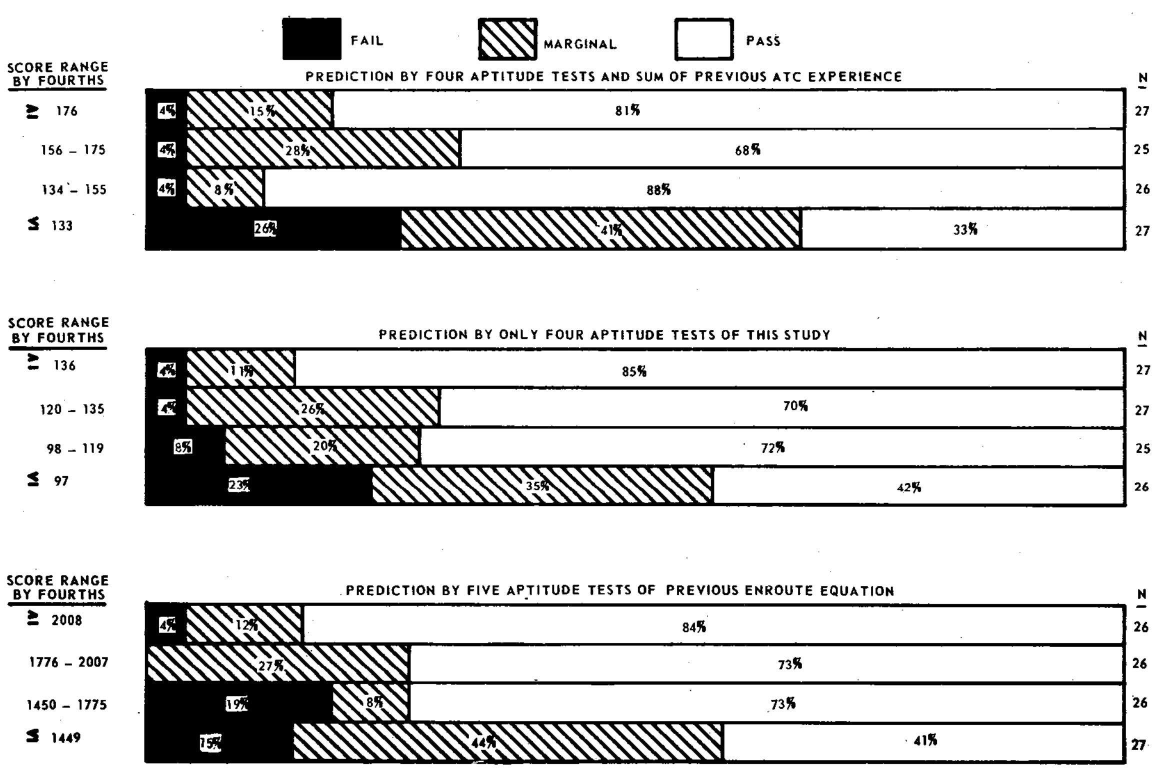

Fig. 1. Comparative Effectiveness of the three equations in predicting Terminal ATC course performance for cases of the second sample only.

each equation's predicted scores versus the Combined Academic-Laboratory Grade Average for this sample may be found in Table III. They were: .46 for equation 16; .44 for number 15, and .39 for the Enroute. A comparison of these correlations and of the graphically illustrated results in Figure 1 (for this latter and more liberal type of validation procedure), indicates that the five tests of the Enroute equation are somewhat less effective than the four "core" tests embodied in the two Terminal equations. The graphically presented data also illustrate that the inclusion of specific types of experience, such as represented in the ATC Experience variable of equation 16, contribute toward a slight improvement over the prediction accuracy achieved with only the four "core" tests of equation 15.

DISCUSSION AND CONCLUSIONS

The four aptitude tests, which the analyses have identified as most predictive of Terminal ATC Training Course performance measures, are different from, and are factorially more complex than, the five tests which previous research demonstrated as most predictive of Enroute training. The primary difference in test content appears to be the presence of a factor related to verbal abstract reasoning in the composite of tests selected for the Terminal trainees.

At the present time a battery of Civil Service Commission (CSC) tests is being used operationally to select applicants for all types of air traffic control specialist training. It is known that this group of tests incorporates those aptitude areas previously found to be of importance for Enroute trainees and it has also been shown that composite scores based on the battery are highly valid for the prediction of performance for Terminal trainees.[2, 5, 7] However, the degree to which the additional aptitude area of verbal abstract reasoning is represented in the operational screening battery and whether or not it is of sufficient significance in discriminating between Enroute and Terminal trainees to warrant a possible change in selection standards are questions which would require additional research and a study of monetary feasibility.

For the present study, it should be emphasized that the predictor composites and their indices of predictive potential have been derived on performance data of trainees who had been recruited under former CSC selection standards. Qualification and selection was based solely on medical and pre-employment experience requirements, with the latter including a consideration of not only those types of experience directly related to air traffic control, but also of various types of communications work and of pilot experience. In this sense, the samples represent a pre-screened input of trainees and thus contribute toward additional significance of predictive results obtained with composite scores based only on aptitude measures.

However, these screening procedures have also tended to prohibit a true evaluation of the Sum of ATC Experience variable. In view of the fact that the present analyses were based on data of only those applicants who qualified for training on the basis of previous experience, and inasmuch as the Sum of ATC Experience variable incorporated a major portion of those types and amounts of experience assessed for qualification purposes, the data for this variable are undoubtedly attenuated. The correlations and other indices are probably of lesser magnitude than would have been obtained with non-restricted samples. In other words, if the preceding hypotheses are correct, the results obtained in the present study are not indicative of the true value with which ATC Experience could be used to predict performance measures for a non-screened trainee input.

The contribution of ATC Experience in the prediction of training course performance is recognized in the newly established CSC standards for selection of ATCS trainees. Under the new standards, although everyone applying for training must achieve a qualifying score on the test battery, that score which is considered qualifying is lower for those applicants having this type of ATC experience.

REFERENCES

1. BOTTENBERG, R. A.: An iterative multiple regression procedure using the IBM 650. WCLL-TM-60-3, Personnel Laboratory, WADD, USAF, Lackland, AFB, Texas, 1959.
2. COBB, B. B.: Problems in air traffic management: II. Prediction of success in air traffic controller school. *Aerospace Med.*, 33:702, 1962.
3. GREENBERGER, M. H., and WARD, J. H., JR.: An iterative technique for multiple correlation analysis. IBM Tech. Newsletter No. 12, International Business Machines Corporation, New York, N. Y., 1956.
4. TRITES, D. K.: Problems in air traffic management: I. Longitudinal prediction of effectiveness of air traffic controllers. *Aerospace Med.*, 32:1112, 1961.
5. TRITES, D. K.: Ground Support Personnel, *Aerospace Med.*, 34:359, 1963.
6. TRITES, D. K., and COBB, B. B.: Problems in air traffic management: III. Implications of age for training and job performance of air traffic controllers. *Aerospace Med.*, 35: 336, 1964.
7. TRITES, D. K., and COBB, B. B.: CARI research on air traffic control specialists: age, aptitude, and experience as predictors of performance. A non-technical report, Civil Aeromedical Research Institute, FAA, Oklahoma City, Oklahoma, 1964.