Technical Report Documentation Page

1. Report No. FAA-AM-79-21	2. Government Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle TOWARD THE DEVELOPMENT OF A NEW SELECTION BATTERY FOR AIR TRAFFIC CONTROL SPECIALISTS		5. Report Date SEPTEMBER 1979 6. Performing Organization Code
7. Author(s) JAMES O. BOONE		8. Performing Organization Report No.
9. Performing Organization Name and Address FAA Civil Aeromedical Institute P.O. Box 25082 Oklahoma City, Oklahoma 73125		10. Work Unit No. (TRAIS) 11. Contract or Grant No.
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, S.W. Washington, D.C. 20591		13. Type of Report and Period Covered 14. Sponsoring Agency Code

15. Supplementary Notes

Work was performed under task AM-C-79-PSY-70.

16. Abstract

In an effort to update and refine the selection battery for air traffic controllers, five experimental tests measuring aptitudes and skills considered important in air traffic work were administered to newly selected Air Traffic Control Specialist (ATCS) trainees on their first day of training at the FAA Academy in Oklahoma City. The testing covered a 21-month period from July 1976 to April 1978 and involved a final sample of approximately 2,500 new trainees.

The five experimental tests and the five tests presently used by the Civil Service Commission (CSC) for selecting ATCS trainees were correlated with the averaged laboratory scores from FAA Academy training. These correlations were then employed in an iterative stepwise regression (stepdown procedure). The tests that made a significant contribution in predicting Academy scores were then used to form a composite and the multiple correlation was computed for the old test battery and the new battery.

The new composite demonstrated a statistically significant increase in the multiple correlation over the old test battery. Use of the new test battery could result in a savings to the FAA in terms of Academy attrition due to failures. It could also aid in upgrading the quality of ATCS selectees and aid in minimizing human error in air traffic control work.

17. Key Words Selection, Regression Testing, Validity Air Traffic Control	18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22161

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	37	

Form DOT F 1700.7 (8-72) Reproduction of completed page authorized

TOWARD THE DEVELOPMENT OF A NEW SELECTION BATTERY FOR AIR TRAFFIC CONTROL SPECIALISTS

James O. Boone

Introduction.

In August 1960, the Civil Aeromedical Institute (then the Civil Aeromedical Research Institute) began administering a heterogeneous battery of commercially available aptitude tests on an experimental basis to newly selected air traffic control specialist (ATCS) students at the Federal Aviation Administration (FAA) Academy in Oklahoma City. After the 9-week training course at the Academy, the student's average academic training test scores and average laboratory scores were summed to form a composite, and this composite was correlated (Pearson product-moment formula) with the composite of the aptitude test scores. The coefficients ranged from .35 to .54. Based on this evidence, it was decided that aptitude tests could enhance the selection process for air traffic control specialists (26).

Since commercially available tests were considered more susceptible to compromise than tests under rigid governmental control, the commercially available tests that showed the most promise were used to identify Civil Service Commission (CSC) tests that appeared similar in factor content. The CSC tests and an additional Air Traffic Problems test (ATP) were then employed, beginning August 1961, in another series of testing sessions at the Academy. Subsequent regression analysis resulted in five best predictors. These are listed and described in Table 1 (5).

Beginning July 1962, the new test battery served as the major selection method for applicants with no previous experience related to air traffic control (ATC). The Civil Aeromedical Institute continued to collect data on the new test, and in January 1964, the CSC battery was introduced as a means to determine if the applicants were qualified for placement on the register regardless of their previous experience. Experience related to air traffic was then used as additional information in ranking applicants on the register (5).

In October 1968, a new means was introduced to select air traffic controllers, aimed at relieving the critical shortage of air traffic personnel due to the expanding airline industry. Under the new method applicants with previous air traffic experience, especially radar experience, were hired at a higher pay grade and without taking the CSC battery (6,8,29,31).

The methods and standards for establishing rankings on the register based on prior related air traffic experience has varied from time to time since the beginning use of a test battery in 1962, although the total selection procedure has remained essentially the same (8,9,12,13,16,17). The test

Table 1. Civil Service Commission Tests

51. CSC Spatial Patterns: Identify solid figures that can be made from an unfolded pattern or, from various views of an object, identify the object in a series of alternatives.

24. CSC Computations: Test of arithmetic computational skill.

157. CSC Abstract Reasoning: Indicate which of a series of choices (figures) properly carries out of a principle of logical development exhibited by a sequence of figures.

135. CSC Oral Directions: From orally presented information, decisions must be made regarding performance of simple tasks.

540. CSC Air Traffic Problems, Part I: Determine whether aircraft may be permitted to change altitude without violating a specified time-separation rule.

Empirical Validities

CIVIL SERVICE COMMISSION TESTS	Course Grade N_t r	P-F N_p-N_f r_{pb}
CIVIL SERVICE COMMISSION	N=183	143-40
CSC 51-Spatial Patterns	.37**	.27**
CSC 24-Computations	.28**	.16*
CSC 157-Abstract Reasoning	.28**	.18*
CSC 135-Oral Directions	.23**	.23**
CSC 540-ATP I+II	.41**	.29**

**$p < .01$
*$p < .05$

battery is used to qualify applicants for the register and prior experience is weighted and used either directly or indirectly to select air traffic personnel. The same general procedure (with the exception of the maximum eligible age level which was established at age 30 for En Route and Terminal options in 1973) has continued until the present (7,10,14,22,24,26,27,28,30).

In a continuing effort to update and improve air traffic controller selection procedures, a task force was commissioned in December 1974 to review the agency's selection policies. The task force identified the following areas of concern (19):

1. The testing and screening of applicants for air traffic control work.

2. The CSC rating guide used to grant additional points for certain types of related prior experience.

3. The evaluation of current recruitment and testing practices for cultural bias against women and racial minorities.

As a result of the task force review of air traffic controller selection procedures, several activities were initiated, including the collection of data on already existing tests (see Table 2 for a description of the tests) and on two newly developed tests, and two major studies were performed. These activities were primarily under the auspices of the FAA Office of Personnel and Training in Washington, D.C.

The two newly developed tests were the Directional Headings Test (DHT) and the Multiplex Controller Aptitude Test (MCAT). The DHT is a highly speeded and rather novel paper-and-pencil aptitude test. The test is in two parts. In each item the subject is presented one, two, or three pieces of information reflecting the cardinal points on a mariner's compass. As an example, N, Λ, and 360° all denote North. In Part I of the test the examinee must determine very swiftly if the information conflicts or agrees. The item is followed by one of five questions: North?, East?, West?, South?, or Conflict?, to which the examinee must respond yes or no. Part II is similar to Part I except the examinee answers whether the data presented represents opposite directions. A complete review of the test and various statistics can be found in Cobb and Mathews (11).

The MCAT consists of job sample items from controller activities. The test comprises two homogeneous areas: (i) air traffic aptitude and (ii) the ability to recognize potential conflicts, and contains subcategories under these two areas. The items are sequenced in increasing difficulty. With each item an air route map is presented with various identified aircraft on the routes. Tabular information is given for each aircraft, such as altitude and speed. Various questions are then posed related to this information. A description of the MCAT and various statistics on the reliability and validity of the test are given in Dailey and Pickrel (20). The MCAT as used here was varied in format and in length. These variations were a function of its developmental phases. Further developments have occurred since accomplishment of this study.

The first of the two major ATCS selection studies was performed by Education and Public Affairs (EPA), a private research organization located in Washington, D.C. One of the major objectives of this FAA-contracted study (18, 19,23) was to determine the potential of an experimental test battery to predict ATCS success. An aggregate "success" criterion was employed in the study, based on a composite of supervisory assessment and career progression. The experimental tests considered were:

Multiplex Controller Aptitude Test (MCAT)
Directional Headings Test (DHT)
Dial Reading Test (DRT)
Arithmetic Reasoning Test (ART)

Table 2. Description of Already Existing Tests Used in the Studies

Other Than CSC Tests*

Dial Reading (Part I of the Dial and Table Reading Test)

- USAF Air Training Command, Lackland Air Force Base, San Antonio, Texas.

- 57 items; 11½ minutes; 3 practice items.

- The examinee is presented with seven dials for each set of six questions and is required to read the correct value on the correct dial in order to select the answer from among five given alternatives.

  Validity (Dial and Table Tests combined): .41 against success in navigation training (final composite grade) with nearly 2,000 students; validities of .20 to .28 (p < .01) against performance in pilot training. (Communication from Jay Bowles, AFHRL to EPA.) Task I re-analysis produced a validity of .17 (p < .05) against progression for 180 new hires in 1971.

- Reliability: Mean phi coefficients (Dial and Table Tests combined): of .20 with a range from .04 to .42, using upper and lower 25 percent of group of 800 unclassified aviation students.

Arithmetic Reasoning

- Army Air Force Aviation Psychological Research Unit No. 3. Chief contributors: Capt. Lloyd G. Humphreys, Lt. David H. Jenkins, and Jean R. Lyons. Authorization for FAA to use this test was obtained from the Air Force Human Resources Laboratory at Lackland Air Force Base.

- 20 items selected from among the easier of the original 30 items; 25 minutes; no practice items.

- "Arithmetic reasoning problems that can be solved with minimal formal mathematical training . . . The items of the test are arranged roughly in order of increasing difficulty. They are formulated in aviation terms in the interest of face validity. All problems are presented simply and concisely in an attempt to minimize verbal variance"(Guilford and Lacey, 1947).

- Reliability: Using samples of unclassified aviation students, odd-even reliability was .77 (N=500); equivalent-halves reliability was .84 (N=1,000) (Guilford and Lacey, 1947).

- Validity: Comparable validity was inferred for the present study population based on the similarity of items in this test to those in the Dailey Technical and Scholastic Test Arithmetic Reasoning. The Task I re-analysis revealed validities of .07 (p < .05) against both progression and supervisory ratings for 596 journeyman controllers in the 1971 research.

*Taken from a study by Education & Public Affairs, Washington, D.C.,(23).

ATC General Information Test (GIT)
ATC Occupational Knowledge Test (OKT)
The present ATC selection battery (Table 1), Office of Personnel Management

The EPA study (23) found that the MCAT and the OKT clearly had value in predicting ATCS success. The DHT and the DRT showed some value in the study, but their value was not as clearly demonstrated as that of the MCAT and OKT. The EPA study was not able to demonstrate the relative value of the experimental tests and the present battery since no information was available on the experimental tests from the applicant group. To evaluate the relative value of all the tests that demonstrated potential, the Office of Personnel Management administered two of the experimental tests in conjunction with the regular battery. These data were then employed in the second major study which was conducted by the Aviation Psychology Laboratory at the Civil Aeromedical Institute (CAMI). The CAMI study is the subject of this paper. The purpose of the CAMI study was to determine which of the selected experimental tests, either independently or in combination with present CSC tests, were the best predictors of success at the FAA Academy. Final decisions regarding the choice of tests to be included in the battery were the prerogatives of FAA's Office of Aviation Medicine and Office of Personnel and Training. The basic questions to be studied are illustrated in Figure 1.

ATC SCREENING TEST

Should the present ATC CSC test be changed?

Evaluate experimental tests under consideration.
Should any of these be in the ATC Selection Battery?

Is the MCAT and/or DHT more predictive of ATC "success" than present CSC tests or individual test parts?

Decision to MODIFY/CHANGE CSC ATC Test Battery.

Should battery parts be differentially weighted?
If so, how?

Figure 1. Basic questions to be studied.

Methods.

Subjects. The subjects came from two sources. In 1977, the CSC in cooperation with the FAA administered the MCAT, the DHT, and the regular ATCS battery to approximately 7,000 ATCS applicants. The second source of subjects was persons selected for ATCS work beginning May 1976 through April 1978. Only subjects who had a complete data set were included in the sample. The final sample contained 1,828 subjects. Newly selected ATCSs were given 1 week of orientation at their regional office prior to coming to the FAA Academy for ATCS training. During the first day at the Academy, new trainees were tested with the experimental test battery. The battery included the following instruments:

Multiplex Controller Aptitude Test (MCAT)
Directional Headings Test (DHT)
Dial Reading Test (DRT)
Biographical Questionnaire (BQ) (See Appendix 1 for example items.)

The testing sessions were conducted in a large auditorium. The administrative procedures were standardized in written form. Timing of the tests was done by two separate devices in case one failed; many of the testing sessions were also recorded by tape recorder and the timing and procedures verified later.

Criterion. Performance scores were maintained on the trainees throughout their Academy training. Training scores were obtained from performance in academic phases and a lab phase where academics were applied. Previous studies demonstrated that the laboratory scores were the most reliable predictors of ATCS success (5,6,26,31). Consequently, the laboratory average was used as the criterion of success for the present study. Several adjustments, including a change in the weighting of score components in January 1979, were made in the lab grading procedures during the data collection phase. In order to compensate for these possible instabilities across inputs, the laboratory scores were standardized within each input by converting the scores to a common metric, having a mean of 0 and a standard deviation of 1. This variable was termed ZLAB. Appendix 2 shows the Academy scoring procedures.

Analyses. The first step in the analyses was the calculation of descriptive statistics on the CSC applicant group and the CAMI trainee group. Descriptive statistics consisted of sample sizes, means, standard deviations, distributions, and intercorrelations. Distributions were graphed. The descriptive statistics for each test being considered were reviewed for their value in predicting successful ATCSs.

The remaining analyses presented several rather unique problems. First, several different experimental forms of the MCAT were employed in the CAMI testing, and the order of administration was varied for each form (see Dailey and Pickrel (20) and Appendix 3 of this report for order and form effects). However, since the MCAT 706 was used in testing the applicant group, the MCAT scores were converted to the same metric as MCAT 706 by the following linear conversion:

$$X_{ba} = \frac{\sigma a}{\sigma b} X_b - \frac{\sigma a}{\sigma b} M_b - M_a ,$$

where X_{ba} = transformed score, M_a = mean of distribution a, σa = standard deviation of distribution a, X_b = a value in distribution b, M_b = mean of distribution b, and σb = standard deviation of distribution b. The order effect problem was handled by using the scores from the MCAT 706 given first, since an applicant would be taking the test for the first time. Since the MCATs used in this study are a miscellaneous collection of early prototypes, converting MCAT scores by this method could have some restricting effects; however, without the conversion the smaller sample size on any given form would be a more marked restriction.

The second problem involved the well-known restriction-in-range effect (25). Since criterion information is available only on those persons who were selected, correlations of test scores with the criterion were spuriously low. This situation is illustrated in Figure 2.

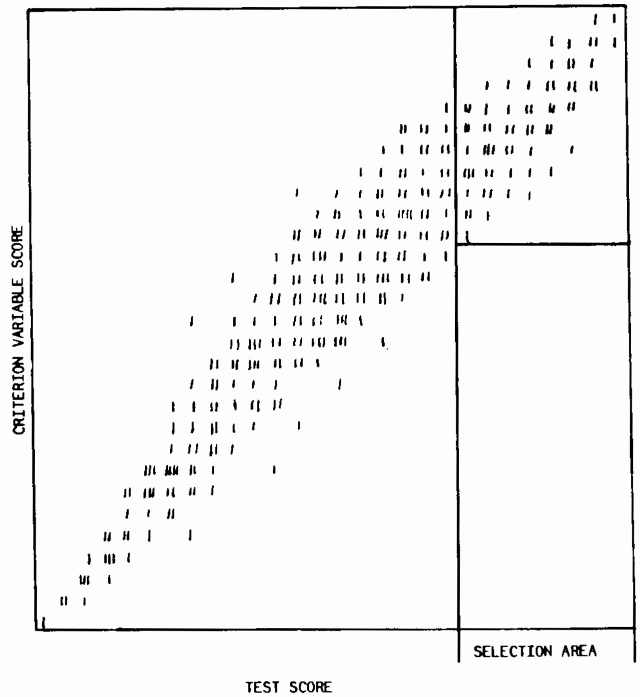

Figure 2. The effect of restricted range on a correlation coefficient. Subjects in the smaller box represent the selected group. The unrestricted correlation of the two variables is .88, and the restricted is .15.

To adjust the restricted correlations so they would reflect the relationship between the tests and the criterion for the applicant group, the correlations were corrected for their restriction in range. The usual methods for correcting correlations for restriction in range in the three-variable case are based on the assumption that unrestricted information is available only on the variable used for selection or the third incidental variable but not on both. In the present situation unrestricted information was available on both variables. A modified procedure to include this information was developed by returning to the assumptions usually made in developing the correction formula and deriving a new set of equations based on a modified set of assumptions which use all the available information. Full details of the procedure can be obtained elsewhere (2); see Appendix 4 for derivations.

The corrected correlation coefficients were input into a stepwise multiple regression computer program, REG. REGR is a modified version of REGRAN (32), adapted by the author for use on the PDP 11 computer system. Since multicolinearity could be a problem in multiple regression when evaluating weights, a stepwise procedure was employed and several different combinations of variables were examined. This is not considered to be a complete solution; however, based on administrative policies requiring the interpretation of the relative magnitude of regression coefficients, this was considered a viable approach.

Various models were examined to determine which subset of the tests in a linear weighted composite produced the maximally efficient prediction of the success criterion. When this weighted subset was identified, the beta weights from the multiple regression analysis were converted to raw score weights via

$$W_p = \frac{\sigma p}{\sigma c} B_p \qquad (3)$$

where Wp = raw score weight, σp = standard deviation of the predictor, σc = standard deviation of the criterion, and Bp = the Beta for the predictor. Unit weights were then assigned since they are much easier for field testing personnel to use in forming a composite score. The multiple R and R^2 were compared using the beta weights and the unit weights to calculate any shrinkage in prediction between the two weighting systems (21).

Crossvalidation. Crossvalidation of the weighting system was reviewed in the following manner. Random numbers ranging from 1 to 2,000 were assigned to each data record from a population of uniformly distributed random numbers. The data records were then sorted into ascending order based on their random number. The sample was then divided into two equal groups. Subsequent multiple regressions were calculated on the first group and unit weights developed. These weights were then applied to the data in the second sample. The multiple Rs and R^2s for each group were then compared for stability based on using the unit weights.

Results.

In Tables 3 and 4 and Figures 3 through 13, the descriptive statistics for the unrestricted applicant group are given. The earned rating is the final compilation of test scores, experience, and education points. There are some interesting results shown in the distribution graphs (Figures 12 and 13). The distributions for CSC 135 and CSC 51 are markedly skewed left. The selection ratio for applicant to selectee is about 5 to 6 percent for air traffic control. Viewed from the graphs there is very little variation among the applicants at the extreme end of the distribution. Consequently, it is evident that CSC 135 and CSC 51 discriminate very poorly between applicants with high scores. Further, the disparity between the applicant group variance and the selected group variance creates a spuriously high corrected correlation. This problem is discussed in detail in the Discussion section of this paper. Based on this information, CSC 135 and CSC 51 are not included in subsequent analyses.

Table 3. Unrestricted Means and Standard Deviations and Sample Sizes From the CSC

	VARIABLE	MEAN	S.D.	MAX N	MIN N
1	CSC 24 Score	39.6641	9.6026	7412	6821
2	CSC 51 Score	26.6541	6.6892	7412	6821
3	CSC 540 Score	28.9810	13.1209	7412	6821
4	CSC 157 Score	29.2903	10.3189	7412	6821
5	CSC 135 Score	21.4935	8.5843	7412	6821
6	MCAT A	16.5932	5.7064	6822	6605
7	MCAT C	9.1665	4.1516	6822	6605
8	MCAT TOTAL	25.7597	9.0909	6822	6605
9	DHT A	24.1950	11.9715	7073	6605
10	DHT B	22.7973	11.7144	7073	6605
11	DHT TOTAL	46.9867	22.5647	7073	6605
12*	EARNED RATE	70.0255	18.5049	7412	6821

*The CSC composite score plus extra credit for experience and education.

Table 4. Unrestricted Correlation Matrix From the CSC

	VARIABLE	1 T 24.S CORE	2 T 51.S CORE	3 T5 540.S CORE	4 T1 157.S CORE	5 T1 135.S CORE	9 T5 .A	10 T5 .C	11 T5 10.T TOTAL	12 DH.B	13 DH.B	14 DH.T TOTAL	19 EAR NED. RATE
1	T24.Score	100	33	56	50	51	52	44	53	46	47	49	66
2	T51.Score	33	100	46	59	54	60	42	57	54	54	57	75
3	T540.Score	56	46	100	59	58	63	56	65	55	56	59	78
4	T157.Score	50	59	59	100	63	63	49	62	53	56	57	85
5	T135.Score	51	54	58	63	100	66	52	65	54	58	59	77
6	T510.A	52	60	63	63	66	100	69	94	65	67	69	77
7	T510.C	44	42	56	49	52	69	100	89	54	55	57	60
8	T510.TOTAL	53	57	65	62	65	94	89	100	66	67	70	76
9	DH.A	46	54	55	53	54	65	54	66	100	82	95	68
10	DH.B	47	54	56	56	58	67	55	67	82	100	95	69
11	DH.TOTAL	49	57	59	57	59	69	57	70	95	95	100	72
12	EARNED RATE	66	75	78	85	77	77	60	76	68	69	72	1.00

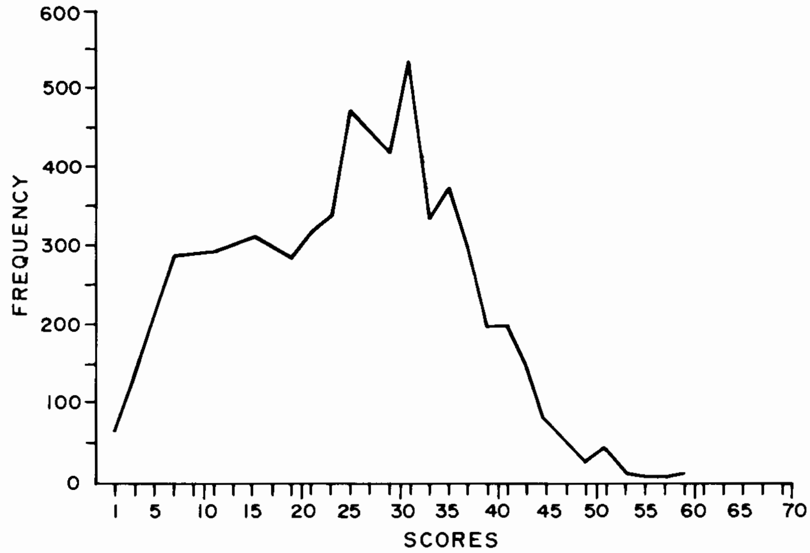

Figure 3. Distribution for DHT Part A.

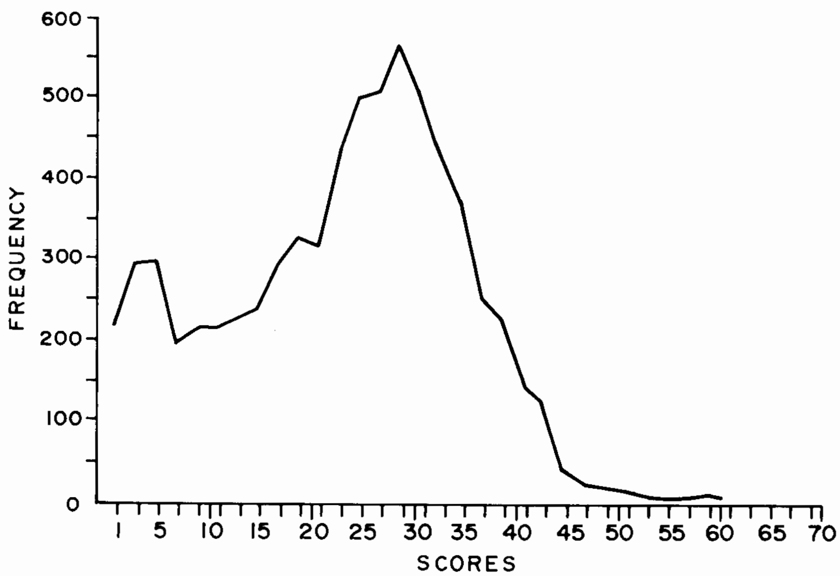

Figure 4. Distribution for DHT Part B.

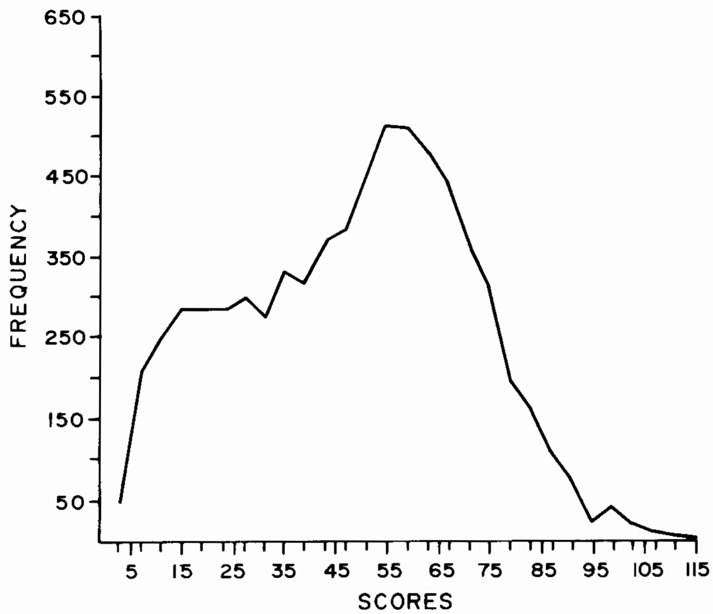

Figure 5. Distribution for DHT Total.

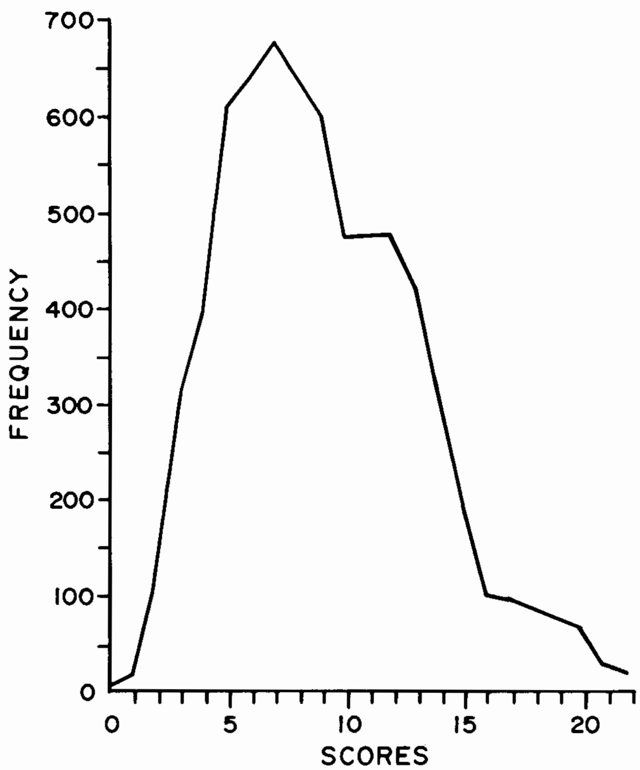

Figure 6. Distribution for MCAT Conflicts.

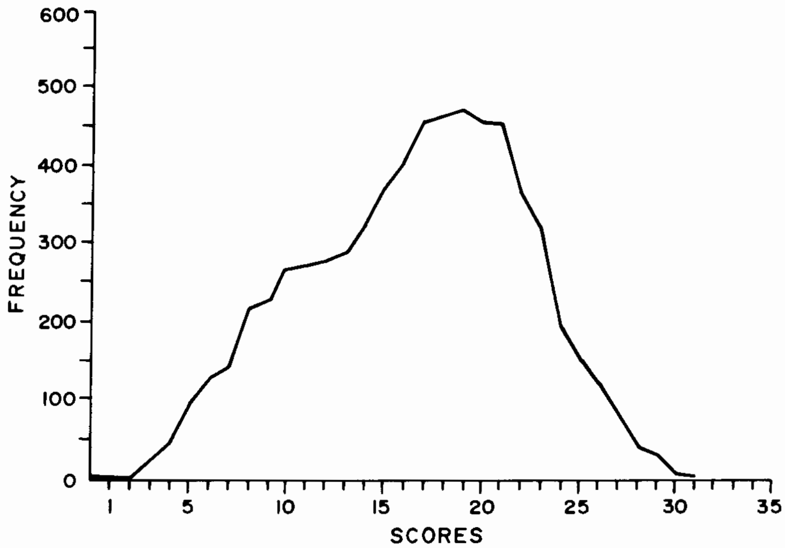

Figure 7. Distribution for MCAT Apitude.

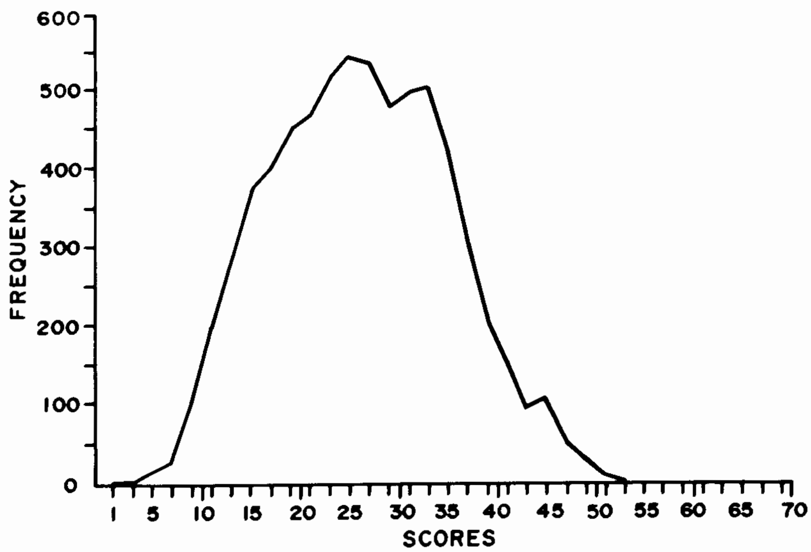

Figure 8. Distribution for MCAT Total.

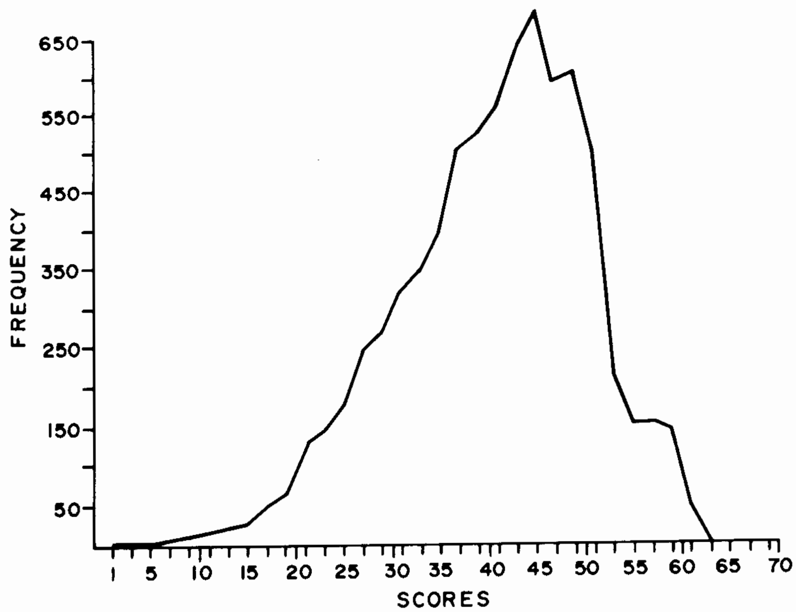

Figure 9. Distribution for CSC 24.

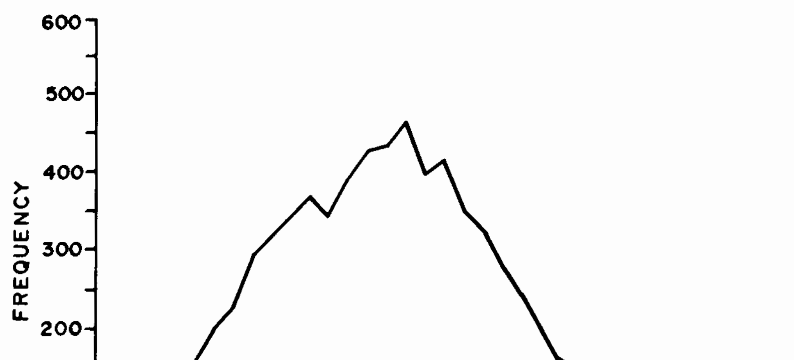

Figure 10. Distribution for CSC 540.

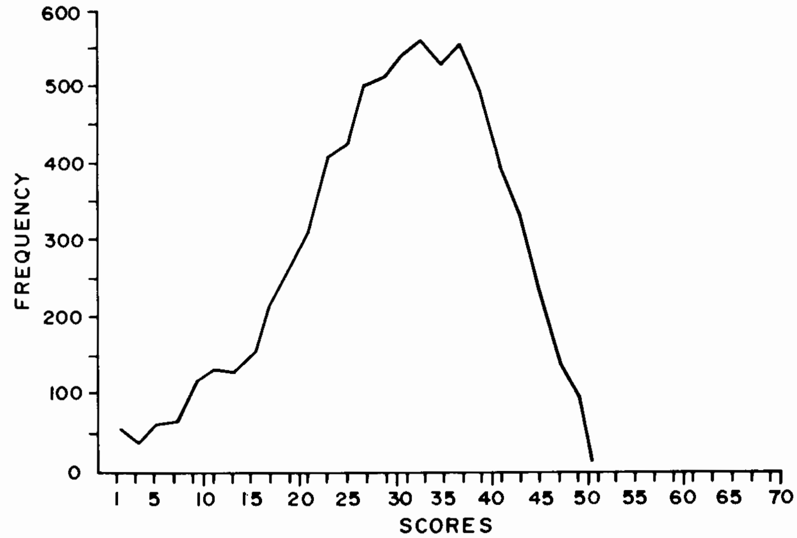

Figure 11. Distribution for CSC 157.

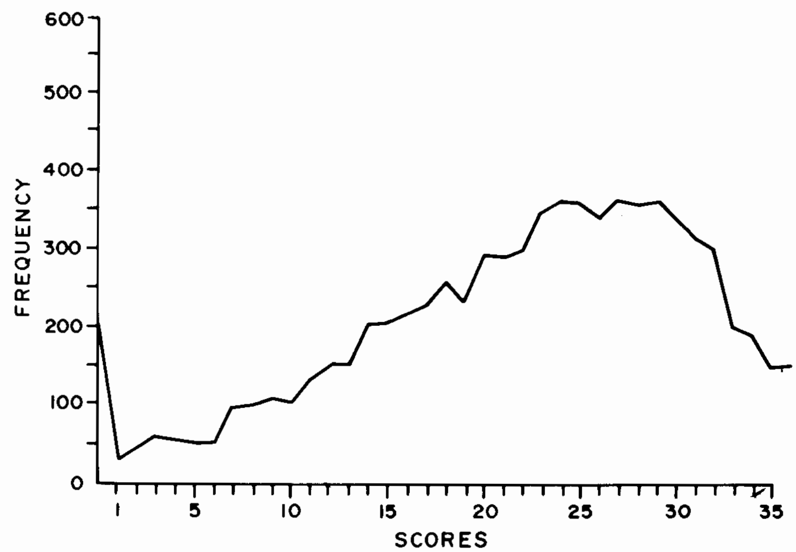

Figure 12. Distribution for CSC 135.

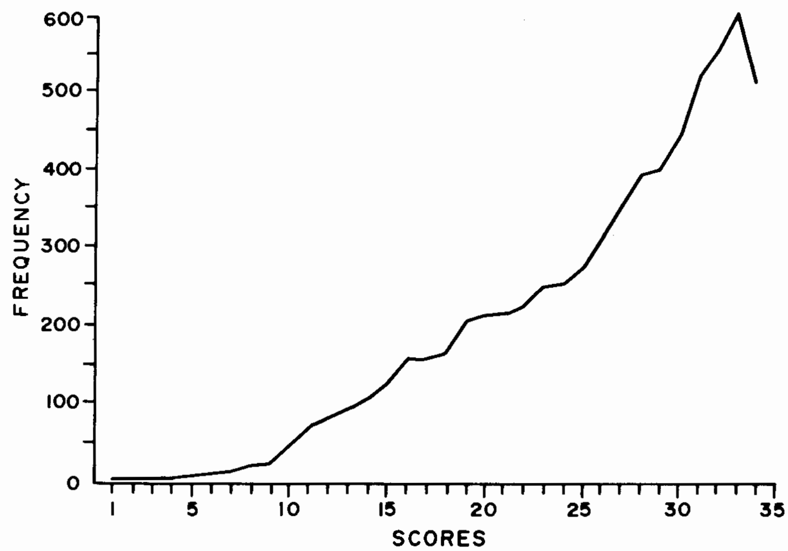

Figure 13. Distribution for CSC 51.

Table 5 contains the means, standard deviations, and intercorrelations for the selected (FAA Academy students) group. The correlations of particular interest are the correlations between the various tests and ZLAB. These are the zero order validity coefficients. The effects of restriction are immediately apparent in the low correlations between the tests used for selection and ZLAB. The two highest zero order restricted validity coefficients are for MCAT total score at .277 and DHT at .227. It must be noted, however, that neither of these tests was restricted by direct selection. As previously noted, ZLAB is in "Z" score form and consequently has a mean of 0.000 and a standard deviation of 0.994 which is very near 1.000.

Table 6 contains the estimated unrestricted correlations (as well as the actual unrestricted correlation from the CSC sample). Again the correlations of primary interest are the correlations of the tests with ZLAB. After corrections, as in Table 4, the MCAT at .531 and DHT at .461 have the highest zero order validity coefficients.

The next step in the analyses was to employ the unrestricted and corrected correlations in a stepwise multiple regression procedure. Tables 7 through 12 contain the results of several different models. The models were executed in a series of steps. Each model was a refinement of the previous model based on information from the previous model. The test scores were regressed on ZLAB.

Table 5. Restricted Correlation Matrix Used to Correct Correlations (N = 1,828)

	MEANS	STANDARD DEVIATIONS
CSC 24	47.07	6.68
CSC 540	42.88	9.66
CSC 157	76.35	12.87
DHT 1	32.68	8.95
DHT 2	31.63	8.96
DHT TOT	64.34	17.02
MCAT A	23.14	4.46
MCAT C	15.74	4.06
MCAT TOT	38.87	7.63
ZLAB	00.00	0.99
DL·RD·	40.97	9.16

CSC 24	CSC 540	CSC 157	DHT 1	DHT 2	DHT TOT	MCAT A	MCAT C	MCAT TOT	ZLAB	DL·RD
1.000	0.333	0.140	0.245	0.255	0.271	0.214	0.146	0.204	0.097	0.343
0.333	1.000	0.145	0.199	0.222	0.227	0.204	0.190	0.221	0.096	0.289
0.140	0.145	1.000	0.090	0.140	0.126	0.241	0.164	0.229	0.071	0.281
0.245	0.199	0.090	1.000	0.803	0.950	0.258	0.268	0.296	0.207	0.338
0.255	0.222	0.140	0.803	1.000	0.949	0.284	0.268	0.308	0.223	0.341
0.271	0.227	0.126	0.950	0.949	1.000	0.282	0.282	0.316	0.227	0.356
0.214	0.204	0.241	0.258	0.284	0.282	1.000	0.596	0.903	0.246	0.481
0.146	0.190	0.164	0.268	0.268	0.282	0.596	1.000	0.882	0.250	0.445
0.204	0.221	0.229	0.296	0.308	0.316	0.903	0.882	1.000	0.277	0.518
0.097	0.096	0.071	0.207	0.223	0.227	0.246	0.250	0.277	1.000	0.272
0.343	0.289	0.281	0.338	0.341	0.356	0.481	0.445	0.518	0.272	1.000

Table 6. Unrestricted and Corrected Correlations Used in Regression Analyses

CSC 24	CSC 540	CSC 157	DHT 1	DHT 2	DHT TOT	MCAT A	MCAT C	MCAT TOT	ZLAB	DL·RD
1.000	.560	.500	.460	.470	.490	.520	.440	.530	.342	.515
.560	1.000	.590	.550	.560	.590	.630	.560	.650	.386	.501
.500	.590	1.000	.530	.560	.570	.630	.490	.620	.399	.502
.460	.550	.530	1.000	.820	.950	.650	.540	.660	.432	.498
.470	.560	.560	.820	1.000	.950	.670	.550	.670	.446	.510
.490	.590	.570	.950	.950	1.000	.690	.570	.700	.461	.522
.520	.630	.630	.650	.670	.690	1.000	.690	.940	.503	.611
.440	.560	.490	.540	.550	.570	.690	1.000	.890	.475	.551
.530	.650	.620	.660	.670	.700	.940	.890	1.000	.531	.635
.342	.386	.399	.432	.446	.461	.503	.475	.531	1.000	.466
.515	.501	.502	.498	.510	.522	.611	.551	.635	.466	1.000

Model 1 contained CSC 24, CSC 540, CSC 157, DHT Part I, DHT Part II, MCAT aptitude, MCAT conflicts, and DRT scores. The total scores for DHT and MCAT were not included in the model since the total scores are the sum of the part scores and, as such, would introduce direct multicolinearity into the regression producing spurious results (21). Model 1 contained a negative beta for CSC 540. This could be interpreted as a suppressor variable; however, since the magnitude of the beta is very small, it appears more reasonable to assume that it is sampling error in the distribution of beta and that the actual beta is 0 (21). The part scores for MCAT aptitude and conflicts have betas of about the same magnitude. Part II in the DHT has a beta somewhat larger than that of Part I. The DRT, when taken with CSC total scores, DHT, and MCAT part scores, has a comparatively larger beta. The CSC 24 beta is quite small, and CSC 157 is about equal with DHT Part I. The multiple "R" for this model was .5689 with a significant "F" ($p < .0001$).

Table 7. Regression Model 1

R = 0.5689 RSQ = 0.3236

V	BETA	B
CSC 24	0.0071	0.0011
CSC 540	-0.0066	-0.0007
CSC 157	0.0555	0.0043
DHT A	0.0513	0.0057
DHT B	0.0912	0.0101
MCAT A	0.1452	0.0322
MCAT C	0.1668	0.0407
DL·RD	0.1856	0.0201
REG. CONST.	=	-3.0582

F-TEST 1	TOTAL MODEL WITH PART SCORES		
RSQ FULL	= 0.3236	Model 1	
RSQ REDUCED	= 0.0000	Model 0	
DIFFERENCE	= 0.3236		
DFN = 7.	DFD = 1800.	F-RATIO = 123.020	P < 0.0001

Model 2 was the same as Model 1 except that total scores for DHT and MCAT were used instead of part scores. Again, CSC 540 has a negative beta. The CSC 24 and CSC 157 betas remain small, while MCAT total, DRT, and DHT total have comparatively larger betas, in that order. The multiple "R" remains essentially unchanged at .5673.

Model 3 demonstrates the effect of eliminating CSC 540 from the equation. Removing CSC 540 has little effect on the betas of the other tests and creates only a negligible impact on the multiple "R" at .5672.

Table 8. Regression Model 2

R = 0.5673 RSQ = 0.3218

V	BETA	B
CSC 24	0.0055	0.0008
CSC 540	-0.0052	-0.0005
CSC 157	0.0472	0.0036
DHT T	0.1332	0.0077
MCAT T	0.2904	0.0377
DL·RD	0.1877	0.0203
REG. CONST.	=	-3.0868

F-TEST 2 FULL MODEL WITH TOTAL SCORES
RSQ FULL = 0.3218 Model 2
RSQ REDUCED = 0.0000 Model 0
DIFFERENCE = 0.3218
DFN = 5. DFD = 1800. F-RATIO = 170.802 $P < 0.0001$

Table 9. Regression Model 3

R = 0.5672 RSQ = 0.3218

V	BETA	B
CSC 24	0.0043	0.0006
CSC 157	0.0470	0.0036
DHT T	0.1322	0.0077
MCAT T	0.2891	0.0375
DL·RD	0.1869	0.0202
REG. CONST.	=	-3.0860

F-TEST 3 24, 157, DHTT, MCATT, DLRD
RSQ FULL = 0.3218 Model 3
RSQ REDUCED = 0.0000 Model 0
DIFFERENCE = 0.3218
DFN = 4. DFD = 1800. F-RATIO = 213.478 $P < 0.0001$

The DRT was considered of marginal value by the Education and Public Affairs study (23) and consequently was not included in the CSC applicant group testing (a more detailed explanation of the DRT is in the Discussion section of the present paper). Model 4 considers the equation without the DRT. When DRT is dropped from the regression, the beta for DHT increases slightly, and the betas for CSC 24, CSC 157, and MCAT increase somewhat. The largest proportional increase is in CSC 24. There is only a slight decrease in the multiple "R" to .5500. It appears that the other tests, especially the CSC 24 and CSC 157, may be measuring attributes similar to those measured by Dial Reading, and the variance shared by the DRT and the other tests is accounted for in the equation by the other tests when Dial Reading is dropped.

Table 10. Regression Model 4

R = 0.5500 RSQ = 0.3025

V	BETA	B
CSC 24	0.0407	0.0060
CSC 157	0.0725	0.0056
DHT T	0.1455	0.0085
MCAT T	0.3262	0.0470
REG. CONST.	=	-3.0831

F-TEST	24, 157, DHTT, MCATT		
RSQ FULL	= 0.3025	Model 4	
RSQ REDUCED	= 0.0000	Model 0	
DIFFERENCE	= 0.3025		
DFN = 3.	DFD = 1800.	F-RATIO = 260.192	P < 0.0001

The DHT is a highly speeded test (90 seconds for each part) and is considered difficult to administer due to the need for strict timing controls. Model 5 considers the equation minus the DHT. Again, the betas for the other tests increase, though not as much as when the DRT was dropped. The betas for the CSC 24 and CSC 157 are still comparatively small.

Table 11. Regression Model 5

R = 0.5407 RSQ = 0.2924

V	BETA	B
CSC 24	0.0608	0.0090
CSC 157	0.0964	0.0074
MCAT T	0.4391	0.0570
REG. CONST.	=	03.2045

F-TEST	24, 157, MCAT TOTAL		
RSQ FULL	= 0.2924	Model 5	
RSQ REDUCED	= 0.0000	Model 0	
DIFFERENCE	= 0.2924		
DFN = 2.	DFD = 1800.	F-RATIO = 371.890	P < 0.0001

The last model (Model 6) is the resultant equation when the DHT and DRT were substituted for the CSC 24 and CSC 157. The betas for this model are more evenly distributed across the tests. A reasonable explanation would be that these three tests measure a similar factor but measure different aspects of that factor. Since we have regressed the tests on ATC Academy success, we are assuming that factor to be "potential success in air traffic control." The multiple "R" (.5659) is slightly higher for this combination of tests than in the previous model. Model 3 contains CSC 24, CSC 157, DHT, MCAT, and DRT and has a multiple "R" of .5672.

Table 12. Regression Model 6

R = 0.5659 RSQ = 0.3203

V	BETA	B
DHT T	0.1446	0.0084
MCAT T	0.3071	0.0398
DL·RD	0.1944	0.0210
REG. CONST.	=	-2.9506

F-TEST	DHT, MCAT, DR		
RSQ FULL	= 0.3203	Model 1	
RSQ REDUCED	= 0.0000	Model 0	
DIFFERENCE	= 0.3203		
DFN = 2.	DFD = 1800.	F-RATIO = 424.105	P < 0.0001

To further explore the characteristics of the test scores, a factor analysis (principal axis analysis with varimax rotation) was performed (Table 13). There appear to be two rather clear structures underlying the data with the orthogonal rotation. Factor 1 and factor 5 account for 22.72 and 42.99 percent of the variance, respectively. Factor 5 contains the largest loadings for all the tests and ZLAB with the exception of the CSC 24 test. It is also notable that a division seems to occur on both factors between the three CSC tests (numbers 24, 540, and 157) and the remaining test (MCAT, DHT, and DRT) and ZLAB. On factor 1 the CSC tests load highest, while on factor 5 the remaining tests and ZLAB load highest.

Based on the models and the outlined constraints the tests in Model 5 were selected to employ in the updated selection battery (see Discussion section). The beta weights were converted to raw score weights via the previously presented formula and then assigned unit weights. The following equation constitutes the composite score:

$$Y_c = 1(\text{CSC } 24) + 2(\text{CSC } 157) + 4(\text{MCAT})$$

where Y_c = composite score. Using unit weights produces the following change in the multiple R and R^2.

Table 13. Factor Analysis on Test Scores (Principal Axis Analysis - Varimax Rotation)

Variables	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
CSC 24	0.9731	-0.0550	-0.0261	-0.0181	0.2583	-0.0211
CSC 540	0.4396	0.3568	-0.1129	-0.5921	0.5849	-0.0564
CSC 157	0.3578	0.0785	-0.7302	0.0079	0.5883	-0.0350
DHT T	0.2638	-0.1068	0.0238	-0.0347	0.7255	0.0543
MCAT T	0.2715	-0.1042	-0.0066	-0.0305	0.7357	0.0522
ZLAB	0.2834	-0.1036	0.0143	-0.0415	0.7666	0.0563
DL·RD	0.3137	-0.0885	-0.0788	-0.0704	0.7740	0.0315
% variance accounted for	22.7188	2.5335	7.9083	5.1392	42.9871	0.2097

Condition	Multiple R	R^2
Using betas	.5407	.2924
Using unit weights	.5354	.2867

To investigate the stability of the results of the regression Model 5, a crossvalidation study was performed. The study was done by randomly dividing the sample and applying the weights derived from the first sample to the second sample and determining what shrinkage occurred in the multiple R. Given a sample size of 900 in each group, little difference was anticipated. New data will provide the ultimate test. The results of the crossvalidation are presented in Tables 14, 15, 16, and 17.

The descriptive statistics, means, standard deviations, intercorrelations, and distributions by sex and race, show the characteristics of the data sets. (See Appendix 5 for a description of the "quick" method employed for stratified random sampling.) These are shown in Tables 14 and 15.

Table 14. Crossvalidation Sample Number 1
(N = 914)

	MEAN	S.D.
ZLAB	0.028	1.007
CSC 24	46.998	6.871
CSC 157	38.490	6.538
MCAT	35.608	7.451

CORRELATIONS

ZLAB	1.000	0.328	0.402	0.537
CSC 24		1.000	0.500	0.530
CSC 157			1.000	0.620
MCAT				1.000

DISTRIBUTION BY SEX AND RACE

	Men	Women	Total
BLACK	47	17	64
HISPANIC	14	3	17
AM. INDIAN	0	1	1
ORIENTAL	6	1	7
ESKIMO	1	0	1
OTHER	730	94	824
TOTAL	798	116	914

Table 15. Crossvalidation Sample 2
(N = 914)

	MEAN	S.D.
ZLAB	-0.020	0.990
CSC 24	47.026	6.853
CSC 157	38.252	6.244
MCAT	35.686	7.307

CORRELATIONS

ZLAB	1.000	0.326	0.396	0.527
CSC 24		1.000	0.500	0.530
CSC 157			1.000	0.620
MCAT				1.000

DISTRIBUTION BY SEX AND RACE

	Men	Women	Total
BLACK	45	16	61
HISPANIC	15	4	19
AM. INDIAN	0	0	0
ORIENTAL	7	1	8
ESKIMO	1	1	2
OTHER	723	101	824
TOTAL	791	123	914

Table 16. Crossvalidation CSC Selection Study, Sample 1

MODEL 1 CRITERION = 4
PREDICTORS = 1- 3

R = 0.5450 RSQ = 0.2970

V	BETA	B
1	0.0354	0.0052
2	0.1023	0.0158
3	0.4548	0.0615

REG. CONST. = -3.0113

F-TEST CROSSVALIDATION
RSQ FULL = 0.2970 MODEL 1
RSQ REDUCED = 0.0000 MODEL 0
DIFFERENCE = 0.2970
DFN = 2. DFD = 913. F-RATIO = 192.847 P < 0.0001

In Table 16 the regression equation is given for the first sample with the multiple R, R^2, and an F test. Unit weights were computed using Formula 1, and the multiple Rs and R^2s were computed on groups 1 and 2 using the unit weights. Table 17 contains the results. Very little shrinkage occurred in the multiple Rs, from .5381 to .5292.

Table 17. Crossvalidation Sample 1

CALCULATED MULTIPLE R USING THE FOLLOWING BETA WEIGHTS

VAR#	WEIGHT
1	1
2	2
3	4

RXY = 0.5381 R SQUARED = 0.2895

Crossvalidation Sample 2

VAR#	WEIGHT
1	1
2	2
3	4

RXY = 0.5292 R SQUARED = 0.2801

Discussion.

As illustrated in the introduction on Figure 1, the basic questions to be resolved by the validation studies were:

1) Should the present CSC battery be changed?

2) Should any of the experimental tests under consideration be used in the ATC selection battery?

3) Specifically, should the MCAT and/or DHT be used in the selection battery?

4) If the battery is changed, how should the tests in the new battery be weighted?

Essentially, these questions can be summarized as follows: Using the tests listed, what is the most efficient and maximally predictive set of tests that can be used to form a composite score for selecting air traffic controllers? To answer the question, each test will be considered independently, and then a composite formed. Information from the Education and Public Affairs study (18,19,23) and this study constituted the statistical and analytical evaluation, while ease of administration, scoring, and length of time required for the test constituted practical criteria. Tests in the present CSC battery that are eliminated in the new battery are discussed first (Table 1), the already existing tests (Table 2) are discussed second (the arithmetic reasoning test was eliminated in the EPA study), the newly developed tests (MCAT and DHT) are discussed third, and composites are discussed last.

CSC 51 and CSC 135. CSC 51 and CSC 135 were eliminated from the battery based on their descriptive statistics. Figures 12 and 13 show CSC 51 and CSC 135 to be negatively skewed, -1.30 and -1.80, respectively. Extreme selection results in a sharp reduction of the variance in the selected group. This effect is accentuated when negative skew is also present, causing the scores of persons in the selected group to be closely clustered. This causes the correlation of the variable with a criterion in the selected group to be very small. When correcting for the restriction in range, the difference between the applicant group variance and the selected group variance is employed as a measure of the amount of curtailment that has occurred due to selection. It was not determined if the skew resulted in a violation of the linearity assumption; however, the extreme disparity between the two variances for CSC 51 and 135 resulted in a corrected correlation that was much higher compared to the other corrected correlations (1,3). In our case if CSC 51 and CSC 135 were corrected and input with the other test correlations into a multiple regression, none of the other tests either independently or in combination added anything significant beyond CSC 51 and CSC 135 to the multiple R. These results were considered spurious; consequently, CSC 51 and CSC 135 were eliminated from the battery.

CSC 540. Models 2 and 3 demonstrate why CSC 540 was eliminated from the battery. In Model 2, CSC 540 had a very small negative beta. Negative betas may indicate that a variable is a suppressor variable and makes a significant contribution to the prediction equation. However, in this case the beta is very near 0, and as shown in Model 3, there is essentially no loss in multiple R by eliminating the test from the battery. Further, the test was designed to measure air traffic controller aptitude which is a duplication of one of the aims of the MCAT test.

Dial Reading Test. The results on this test are puzzling. In the EPA study (23), the Dial Reading Test received a 0 weighting for the VFR, IFR, and all options combined. In the CAMI study (Model 3), the DRT has the second highest beta in the equation. In Model 4 when dial reading is dropped, the betas for CSC 24, CSC 157, and MCAT increase somewhat. In Model 6 when dial reading and the DHT are substituted for the two CSC tests used in Model 5, dial reading again has a substantial beta, and the multiple R is slightly higher than Model 5. The different results obtained in the two studies could be due to a difference in the criterion variable employed. The CAMI study employs training success as a criterion while the EPA study also contains criterion information on field success. Also, in the EPA study the sum of two MCAT forms was employed in the equation and the MCAT made a larger contribution to predictive variance. An administrative decision was made to drop the DRT from the battery. However, it is suggested that further consideration should be given the test as more information becomes available on field success.

Directional Headings Test. Consistently in Models 1-4 and 6, the DHT appears to make a substantial contribution to the regression equation. In Model 4, the DHT has a beta higher than the CSC 24 or CSC 157 beta. In the EPA study (23) the DHT received a comparatively large weight for VFR option, IFR, and for all options combined. Considering that the test requires less than 5 minutes to administer it, it appears to produce substantial information in an efficient manner. Unfortunately, the highly speeded nature of the test requires strict timing and controls. The parts are timed for a 90-s interval. At present, strict controls on timing are not available at field testing facilities. Lack of strict controls makes administration of the DHT very difficult. The lack of strict timing could have resulted in a larger unrestricted variance estimate even in this study and an effect on the corrected correlation. Lengthening the DHT to even 10 min would require several answer sheets. For these reasons an administrative decision was made to drop the DHT from the battery. The test should be pursued further though, to determine if the concept of the test can be extended to a form requiring less administration difficulty in timing. CAMI researchers are presently in the process of reviewing the test.

CSC 24 and CSC 157. The CSC 24 and CSC 157 demonstrate the most potential of the five present CSC tests. Their betas in the equation (Models 1-3) are quite small. However, when the DHT and DRT are dropped (Model 5), the CSC 24 and CSC 157 betas have a comparatively substantial increase. Consistently, CSC 157 appears to have a larger beta than does CSC 24. Given that the DHT and DRT are not included in the battery, it is suggested that CSC 24 and CSC 157 be retained as part of the battery.

Multiplex Controller Aptitude Test. Throughout the EPA reports (18,19,23) and in this study, the MCAT appears to be the most promising test to be included in the battery. In the EPA study the MCAT aptitude and conflict portion was the highest weighted of the experimental tests. In the CAMI study, again, the MCAT was the highest weighted test (Models 1-6). The lowest comparative betas for the MCAT occur when it is combined with the DRT and DHT, both of which also show promise. It is recommended that the MCAT be included in the selection battery. The MCATs employed in this study contain a single set of air traffic samples, consequently the exact forms used in the study may not be the most appropriate to implement. Further development with more traffic samples would be desirable.

The Weighted Composite. Based on the analyses and decisions outlined above, Model 5 is suggested at the present time to represent the air traffic controller selection battery. If unit weights are to be employed, it is suggested that CSC 24 be weighted 1, CSC 157 be weighted 2, and MCAT be weighted 4. As shown in the Results section, the unit weights result in a multiple R of .5354 as opposed to .5407 shown in Model 5. This multiple R leaves room for improvement. However, when compared to data in the general literature on validity studies, a multiple R of .5407 represents a good predictive battery.

Crossvalidation. In order to investigate the stability of the results, the sample was divided into random parts. The first sample was employed in a regression to develop weights and these weights were applied to the second sample. The multiple Rs were compared to determine any shrinkage. As expected, little shrinkage occurred, .5381 to .5292. It should be noted that a crossvalidation study with a large sample and random division of the sample is not as accurate as collecting data on a totally new group of subjects to perform crossvalidation. It is suggested that, as new information becomes available, the crossvalidation be performed on the new sample.

Future Considerations. The EPA study (23) and Model 6 in this study offer evidence that the DHT and possibly the DRT could enhance the selection process for air traffic controllers. A comparison of Model 3 and Model 6 indicates that if the DHT and DRT were included in the battery in place of the CSC 24 and CSC 157, a more efficient and well-rounded battery might result. In Model 6 it appears that MCAT, DHT, and DRT are measuring a similar ability but perhaps different aspects of that ability. The factor analysis in Table 13 further substantiates this idea and also indicates that the CSC 24 and CSC 157 may be measuring a different factor than MCAT, DHT, DRT, and the criterion, ZLAB. At this point it seems advisable to continue study on revising the DHT and collecting further field success data to compare with DRT scores.

REFERENCES

1. Boldt, R. F.: Robustness of Range Restriction in Court. Paper presented at the annual meeting of the American Psychological Association, Toronto, Canada, 1978.

2. Boone, J. O., and M. A. Lewis: The Development of a New Procedure to Make Maximum Use of Available Information When Correcting Correlations for Restriction in Range Due to Selection. FAA Office of Aviation Medicine Report No. AM-78-36, 1978.

3. Brewer, J. K., and J. R. Hills: Univariate Selection: The Effect of Size of Correlation, Degree of Skew, and Degree of Restriction. PSYCHOMETRIKA, 34, 1969.

4. Buckley, E. P., W. E. O'Connor, and T. Beebe: A Comparative Analysis of Individual and System Performance Indices for the Air Traffic Control System. NAFEC, FAA, Atlantic City, New Jersey, September 1969.

5. Cobb, B. B.: Problems in Air Traffic Management: II. Prediction of Success in Air Traffic Controller School. FAA Civil Aeromedical Research Institute Report No. 62-2, 1962.

6. Cobb, B. B.: Problems in Air Traffic Management: V. Identification and Potential of Aptitude Test Measures for Selection of Tower Air Traffic Controller Trainees. FAA Office of Aviation Medicine Report No. AM-65-19, 1965.

7. Cobb, B. B.: The Relationships Between Chronological Age, Length of Experience, and Job Performance Ratings of Air Route Traffic Control Specialists. FAA Office of Aviation Medicine Report No. AM-67-1, 1967.

8. Cobb, B. B.: A Comparative Study of Air Traffic Trainee Aptitude-Test Measures Involving Navy, Marine Corps, and FAA Controllers. FAA Office of Aviation Medicine Report No. AM-68-14, 1968.

9. Cobb, B. B.: Air Traffic Aptitude Test Measures of Military and FAA Controller Trainees. FAA Office of Aviation Medicine Report No. AM-71-40, 1971.

10. Cobb, B. B., C. D. Lay, and N. M. Bourdet: The Relationship Between Chronological Age and Aptitude Test Measures of Advanced-Level Air Traffic Control Trainees. FAA Office of Aviation Medicine Report No. AM-71-36, 1971.

11. Cobb, B. B., and J. J. Mathews: A Proposed New Test for Aptitude Screening of Air Traffic Controller Applicants. FAA Office of Aviation Medicine Report No. AM-72-18, 1972.

12. Cobb, B. B., J. J. Mathews, and C. D. Lay: A Comparative Study of Female and Male Air Traffic Controller Trainees. FAA Office of Aviation Medicine Report No. AM-72-22, 1972.

13. Cobb, B. B., and P. L. Nelson: Aircraft-Pilot and Other Pre-Employment Experience as Factors in the Selection of Air Traffic Controller Trainees. FAA Office of Aviation Medicine Report No. AM-74-8, 1974.

14. Cobb, B. B., P. L. Nelson, and J. J. Mathews: The Relationship of Age and ATC Experience to Job Performance Ratings of Terminal Area Traffic Controllers. FAA Office of Aviation Medicine Report No. AM-73-7, 1973.

15. Cobb, B. B., and J. J. Mathews: A Proposed New Test for Aptitude Screening of Air Traffic Controller Applicants. FAA Office of Aviation Medicine Report No. AM-72-18, 1972.

16. Colmen, J. G.: Review and Evaluation of Present Systems for Selection of Air Traffic Controllers. Report of Phase 1, Task 1, FAA Contract #DOT-FA-70WA-2371, Education and Public Affairs, Washington, D.C., July 1970.

17. Colmen, J. G.: Worker Attributes for Air Traffic Control Jobs and Their Measurement. Report of Phase 1, Task 2, FAA Contract #DOT-FA-70WA-2371, Education and Public Affairs, Washington, D.C., 1970.

18. Colmen, J. G., and J. Marshall-Mies: Design of Field Validation: Report of Task III. Education and Public Affairs, Washington, D.C., January 1976a.

19. Colmen, J. G., and J. Marshall-Mies: Selection of Air Traffic Controllers. Report of Task II, Design of Field Validation. Education and Public Affairs, Washington, D.C., January 1976b.

20. Dailey, J. T., and E. W. Pickrel: Development of New Selection Tests for Air Traffic Controllers. FAA Office of Aviation Medicine Report No. AM-77-25, 1977.

21. Kerlinger, F. N.: Multiple Regression in Behavioral Research. Holt, Rinehart, and Winston: New York, 1973.

22. Lewis, M. A.: Use of the Occupational Knowledge Test to Assign Extra Credit in Selection of Air Traffic Controllers. FAA Office of Aviation Medicine Report No. 78-7, 1978.

23. Marshall-Mies, J., and J. G. Colmen: Development of Recommendations for ATCS Selection Tests. Report of Task 1. Education and Public Affairs, Washington, D.C., 1976c.

24. Milne, A. M., and J. G. Colmen: Selection of Air Traffic Controllers for Federal Aviation Administration. Education and Public Affairs, Washington, D.C., January 1972.

25. Thorndike, R. L.: Personnel Selection. Wiley and Sons: New York, 1949.

26. Trites, D. K.: Problems in Air Traffic Management: I. Longitudinal Prediction of Effectiveness in Air Traffic Controllers. FAA Civil Aeromedical Research Institute Report No. 61-1, 1961.

27. Trites, D. K.: Problems in Air Traffic Management: VI. Interaction of Training-Entry Age With Intellectual and Personality Characteristics of Air Traffic Control Specialists. FAA Office of Aviation Medicine Report No. AM-65-21, 1965.

28. Trites, D. K., and B. B. Cobb: Problems in Air Traffic Management: III. Implications of Age for Training and Job Performance of Air Traffic Controllers. FAA Civil Aeromedical Research Institute Report No. 62-3, 1962.

29. Trites, D. K., and B. B. Cobb: Problems in Air Traffic Management: IV. Comparison of Pre-Employment, Job-Related Experience With Aptitude Tests as Predictors of Training and Job Performance of Air Traffic Control Specialists. FAA Civil Aeromedical Research Institute Report No. 63-31, 1963.

30. Trites, D. K., and B. B. Cobb: CARI Research on Air Traffic Control Specialists: Age, Aptitude, and Experience as Predictors of Performance. FAA Civil Aeromedical Research Institute Unnumbered Report, 1964.

31. Trites, D. K., M. C. Miller, and B. B. Cobb: Problems in Air Traffic Management: VII. Job and Training Performance of Air Traffic Control Specialists--Measurement, Structure, and Prediction. FAA Office of Aviation Medicine Report No. AM-65-22, 1965.

32. Veldman, D. J.: Fortran Programming for the Behavioral Sciences. Holt, Rinehart, and Winston: New York, 1967.

Appendix 1

BIOGRAPHICAL QUESTIONNAIRE
(Example Items)

All the items which follow are in the familiar multiple choice format. Answer each one by blackening the circle in the appropriate column (A, B, C, D, or E) on your answer sheet. Choose the response that best fits you and only make one response per question.

HIGH SCHOOL EDUCATION

1. Which of the following best describes your high school career?

   A. Did not attend high school
   B. Did not complete high school
   C. High school diploma granted by school
   D. High school diploma granted by G.E.D.

2. How old were you when you left high school?

   A. 15 or younger
   B. 16
   C. 17
   D. 18
   E. 19 or older

What grades, on the average, did you get in the following high school courses? Fill in the letter corresponding to the grade for each subject.

   A. About "A-" to "A+"
   B. About "B-" to "B+"
   C. About "C-" to "C+"
   D. Lower than "C-"
   E. Did not have course

3. Arithmetic, Math

4. Physical Science

5. Biological Science

6. English

7. Social Studies

Appendix 2

AN EXAMPLE OF THE COMPONENTS AND WEIGHTS USED IN COMPUTING THE FAA ACADEMY TRAINING TOTAL SCORE

Lab Average *65.0%	**Extra Credit 13.00%	Sixth Problem	Extra Credit	2.60%
			Instructor Assessment	3.90%
			Problem Errors	6.50%
	**Instructor Assessment 19.50%	Fifth Problem	Extra Credit	2.60%
			Instructor Assessment	3.90%
			Problem Errors	6.50%
	Problem Errors 32.50%	Fourth Problem	Extra Credit	2.60%
			Instructor Assessment	3.90%
			Problem Errors	6.50%
		Third Problem	Extra Credit	2.60%
			Instructor Assessment	3.90%
			Problem Errors	6.50%
		Second Problem	Extra Credit	1.30%
			Instructor Assessment	1.95%
			Problem Errors	3.25%
		First Problem	Extra Credit	1.30%
			Instructor Assessment	1.95%
			Problem Errors	3.25%

Controller Skills Test 25.00%

Comprehensive Phase Test 8.00%

Block Average 2.00%

*The lab average constitutes 65% of the total training score. ZLAB is based on this average.

**On each lab problem the instructor gives a performance rating for that problem that is averaged with the student's problem performance. Since the rating is not allowed to be below 40, essentially the student is given a certain amount of extra credit in the computation of the problem average.

Appendix 3

EFFECT OF ADMINISTRATION ORDER ON MCATS

	1ST ADMINISTRATION			2ND ADMINISTRATION			TOTALS		
MCAT FORM	MEAN	S.D.	N	MEAN	S.D.	N	MEAN	S.D.	N
606-A A	19.75	2.88	398	20.66	2.46	308	20.33	2.72	706
606-A C	12.67	3.15	398	14.61	2.36	308	13.52	2.99	706
606-A T	32.74	5.29	398	35.28	4.00	308	33.85	4.93	706
606-B A	18.62	3.12	487	19.51	3.00	454	19.03	3.09	941
606-B C	12.61	2.75	487	13.63	2.54	454	13.15	2.69	941
606-B T	31.22	5.12	487	33.12	4.83	454	32.17	5.05	941
706-A A	22.14	4.27	595	24.70	3.50	516	23.33	4.13	1111
706-A C	13.49	4.14	595	17.08	3.70	516	15.16	3.89	1111
706-A T	35.64	7.56	595	41.77	6.33	516	38.49	7.27	1111
706-B A	20.79	4.90	335	23.71	4.11	434	22.25	4.70	769
706-B C	16.27	2.88	335	17.64	2.58	434	16.96	2.74	769
706-B T	37.06	7.08	335	41.36	5.84	434	39.21	6.83	769
607 A	22.45	3.70	516	23.88	4.14	362	23.04	3.95	878
607 C	14.76	3.99	516	16.41	3.66	362	15.44	3.94	878
607 T	37.21	7.16	516	40.30	7.05	362	38.49	7.12	878
707 A	21.81	4.56	247	25.01	4.22	398	23.78	4.38	645
707 C	12.90	4.14	247	16.85	4.05	398	15.33	4.08	645
707 T	34.71	7.82	247	41.86	7.48	398	39.11	7.54	645

Appendix 4

DERIVATION FOR CORRECTING CORRELATIONS FOR RESTRICTION IN RANGE

Following Gulliksen's (4) schema for derivation of the correction formulas, three assumptions were employed, where upper case and lower case letters represent unrestricted and restricted variables respectively and x = the test used for selection, y = the new test being assessed and z = the success criterion.

Assumption 1. The slopes of the regressions of the new test and the criterion used for selection are not affected by selection.

$$Rxy\,\frac{Sy}{Sx} = RRxy\,\frac{SSy}{SSx}$$

$$Rxz\,\frac{Sz}{Sx} = RRxz\,\frac{SSz}{SSx} \qquad (1)$$

Assumption 2. The error made in estimating the new test scores and the criterion from the selection test scores is not affected by selection.

$$Sy^2\,(1-Rxy^2) = SSy^2\,(1-RRxy^2)$$

$$Sz^2\,(1-Rxz^2) = SSz^2\,(1-RRxz^2) \qquad (2)$$

Assumption 3. The partial correlation between the new test and the criterion is not affected by selection.

$$\frac{Ryz-RxyRxz}{\sqrt{(1-Rxy^2)\,(1-Rxz^2)}} = \frac{RRyz-RRxyRRxz}{\sqrt{(1-RRxy^2)\,(1-RRxz^2)}} \qquad (3)$$

Based on assumptions 1 through 3, derivation of the root formulas proceed as follows.

Equation (1) is solved for RRxy,

$$RRxy = Rxy\,\frac{SySSx}{SSySx} \qquad (4)$$

and RRxy is substituted in equation (2),

$$Sy^2\,(1-Rxy^2) = SSy^2\left(1-Rxy^2\,\frac{Sy^2SSx^2}{SSy^2Sx^2}\right). \qquad (5)$$

Multiplying the right side through by SSy^2,

$$Sy^2\,(1-Rxy^2) = SSy^2-Rxy^2Sy^2\,\frac{SSx^2}{Sx^2} \qquad (6)$$

and solving for SSy^2,

$$SSy^2 = Sy^2\left[(1-Rxy^2) + \left(Rxy^2\,\frac{SSx^2}{Sx^2}\right)\right]. \qquad (7)$$

Substituting SSy^2 in equation (4),

$$RRxy = \frac{Rxy\,\frac{SSx}{Sx}}{\sqrt{1-Rxy^2 + Rxy^2\left(\frac{SSx}{Sx}\right)^2}}. \qquad (8)$$

The same method can be used to derive SSz^2 and $RRxz^2$.

$$SSz^2 = Sz^2\left[1-Rxz^2 + Rxz^2\left(\frac{SSx}{Sx}\right)\right]^2 \qquad (9)$$

$$RRxz = \frac{Rxz\,\frac{SSx}{Sx}}{\sqrt{1-Rxz^2 + Rxz^2\left(\frac{SSx}{Sx}\right)^2}} \qquad (10)$$

Solving for RRyz in equation (3), we algebraically change equation (2), dividing first by SSy^2 and taking the square root,

$$\sqrt{(1-RRxy^2)} = \frac{Sy}{SSy}\sqrt{(1-Rxy^2)} \qquad (11)$$

and dividing by SSz^2 and taking the square root,

$$\sqrt{(1-RRxz^2)} = \frac{Sz}{SSz}\sqrt{(1-Rxz^2)} \qquad (12)$$

Substituting (11) and (12) in the denominator of (3)

$$\frac{Ryz-RxyRxz}{\sqrt{1-Rxy^2}\,\sqrt{1-Rxz^2}} = \frac{(RRyz-RRxyRRxz)\;SSySSz}{SySz\sqrt{1-Rxy^2}\sqrt{1-Rxz^2}} \qquad (13)$$

and solving for RRyz

$$RRyz = \frac{(Ryz-RxyRxz)\;SySz}{SSySSz} + RRxyRRxz. \qquad (14)$$

Substituting estimates for SSz (9) and RRxz (10) in the root formula (14) and simplifying gives the correction formula,

$$RRyz = \left[\frac{Sy(Ryz-RxyRxz}{SSy\sqrt{(1-Rxz^2)} + Rxz^2\left(\frac{SSx^2}{Sx^2}\right)}\right] + \left[\frac{Rxz\,\frac{SSx}{Sx}}{(1-Rxz^2) + \left(Rxz^2\,\frac{SSx^2}{Sx^2}\right)}\right]RRxy. \qquad (15)$$

Appendix 5

A "QUICK" METHOD FOR STRATIFIED RANDOM SAMPLING*

Discrete Description Stratified Sampling. This procedure is employed with discrete data. Under this data form the variable is either naturally discrete or, if not, is converted into discrete categories. Some of the variables may already be in discrete form, such as sex (e.g., 1 = female, 0 = male), race (e.g., 0 = white, 1 = black), or socioeconomic status (e.g., 1 = very high, 2 = high, 3 = average, 4 = low, and very low = 5; or any amount of discrimination desired). It becomes obvious, the finer the discrimination the less advantage there is in using this method. Thus, one should balance the fineness of discrimination against the advantage of simplicity.

Accordingly, if the variables were sex (male = 1, female = 2), race (white = 1, nonwhite = 2), achievement (high = 1, medium = 2, low = 3), and socioeconomic class (high = 1, medium = 2, low = 3), the notation 1123 would be the description of a male, white, medium achievement scorer, from a low socioeconomic background. In this example there are 2 x 2 x 3 x 3 = 36 possible descriptor sets. In order to form stratified random samples, discrete descriptor sets are first listed. Then each subject who fits each description is listed under that descriptor set. The last step is the random and equal assignment of subjects from each descriptor set into matched groups.

*Taken from an unpublished university paper by James Boone and James K. Brewer, Florida State University, Tallahassee, Florida, 1975.