

1. Report No. FAA-AM-74- 10		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle MULTIPLE TASK PERFORMANCE AS A PREDICTOR OF THE POTENTIAL OF AIR TRAFFIC CONTROLLER TRAINEES: A FOLLOWUP STUDY				5. Report Date November 1974	
				6. Performing Organization Code	
7. Author(s) W. Dean Chiles, Ph. D., and Georgetta West, M.S.				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. (TRAI5)	
				11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Office of Aviation Medicine Federal Aviation Administration 800 Independence Avenue, SW. Washington, D.C. 20591				13. Type of Report and Period Covered OAM Report	
				14. Sponsoring Agency Code	
15. Supplementary Notes This research was conducted under Tasks AM-A-71-PSY-29 and AM-C-75-PSY-40.					
16. Abstract The current professional status of 229 air traffic controller trainees tested in an earlier series of complex performance studies was determined by reference to the official FAA personnel roster some 2 to 2 1/2 years later. Point-biserial correlations between the previously obtained performance measures and the retention/termination status of the trainees were computed; correlations were also computed between the criterion used in the earlier study (instructor ratings of trainee potential) and the retention criterion and between the Civil Service Commission (CSC) air traffic control specialist (ATCS) aptitude screening battery scores (where available) and the retention criterion. For only one group of trainees was the correlation between the retention criterion and the performance measure significant; this was for a subset of subjects who had not been selected on the basis of their CSC battery scores. Twelve of the fourteen correlations between the instructor ratings and the retention criterion were significant at the 0.05 level of confidence or better, but none of the correlations between the CSC test scores and the retention criterion was significant. A final determination of the potential of the complex-performance-battery approach to ATCS selection will require further validating research. Should FAA selection policies for ATCS's require an unequivocal assessment of the predictive efficiency of performance tests, the results of this study and those of the earlier studies provide a basis for optimism regarding the expected outcome of such research.					
17. Key Words Selection Complex Performance			18. Distribution Statement Document is available to the public through the National Technical Information Service, Springfield, Virginia 22151		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 4	22. Price \$3.00

MULTIPLE TASK PERFORMANCE AS A PREDICTOR OF THE POTENTIAL OF AIR TRAFFIC CONTROLLER TRAINEES: A FOLLOWUP STUDY

I. Introduction.

During the period from May 1970 to January 1971, exploratory studies were carried out to examine the potential usefulness of a performance measurement device as a predictor of the future performance of air traffic control specialists (ATCS). Five groups comprising a total of 229 air traffic controller trainees were tested on the CAMI Multiple Task Performance Battery (MTPB), and the predictor scores that were derived from the performance battery were correlated against a criterion provided by FAA Academy instructors. (The MTPB requires the simultaneous performance of various combinations of six tasks: warning lights monitoring, meter monitoring, mental arithmetic, pattern discrimination, group problem solving, and two-dimensional compensatory tracking.) The major conclusion drawn from that study was "... that the MTPB-approach to selection offers promise as a screening device for air traffic control specialist applicants, but further research is required to establish this as a fact. . . ."² This report concerns a followup study of the trainees tested in that program to examine the extent to which the various previously obtained measures were predictors of whether a trainee became and continued to work as an ATCS.

The official FAA personnel roster (dated January 1, 1973) was screened to determine whether each trainee who participated in the earlier study was still listed as having the job code of an ATCS. Thus, the elapsed time from initial testing to the followup ranged from 2 to 2½ years. Since about 90 of the trainees were tested well along in their training program, their tenures in the ATCS program ranged up to more than 3 years on the followup date.

Three main questions are of interest in this report: (1) How well do the MTPB measures obtained in the earlier study predict the retention

of trainees as controllers? (2) How well does the criterion (instructors' ratings) used in the earlier study predict retention? (3) How well do scores on the Civil Service Commission (CSC) ATCS selection battery (Examination for Air Traffic Control Specialists) predict retention of those trainees for whom scores were available?

The criterion referred to in the second question above was the measure used in the earlier study to evaluate the MTPB as an ATCS selection device; the criterion consisted of ratings given by FAA Academy instructors as to their estimates of each trainee's potential to become a fully rated ATCS. These ratings are very easy to obtain, and they appear to possess satisfactory reliability (estimated to be 0.84).² As noted in the earlier report, previous research¹ had shown that instructor ratings of this general sort are of substantial predictive value with regard to the on-the-job performance of ATCS's. Further corroboration of those findings would be clearly desirable.

It is most important to note that, in general, these trainees represented a *very* select population. At the time of testing, two of the five groups (Studies II and III) had successfully completed the FAA Academy training program, which, at that time, served as a secondary screening mechanism. In addition, 28 of the 61 subjects in Study II and 60 of the 89 subjects in Study V had been selected originally on the basis of having achieved satisfactory scores on the CSC screening battery (the number of trainees who entered the program on the basis of CSC scores was not determined for subjects in Studies III and IV). Thus, it would be expected that the predictive validity of all of the various measures used would be substantially attenuated

and perhaps nullified because of the range restriction produced by the processes previously at work in eliminating most of the trainees with marginal aptitudes.

II. Method.

A. Subjects.

1. Study I. This was a preliminary study of 19 en route students, and the results were not considered to be usable for purposes of the followup.

2. Study II. The 60 subjects in this study had all completed the 9-week en route training course at the Academy and had returned to the Academy for an interim radar course.

3. Study III. This study involved 31 subjects who had the same general background as the subjects in Study II. However, several systematic changes in the task configuration of the MTPB were introduced in this study so that direct comparability of the data between Studies II and III was not afforded.

4. Study IV. This study involved 30 trainees from the Academy terminal training program. These were entry-level trainees who were tested during the first week of their 9-week course. The study task configuration for this group was the same as that used in Study III.

5. Study V. This study involved an entire class of 89 entering trainees who were tested during the first week of the 9-week terminal training course. The task configuration used was intentionally different from that used in each of the other studies.

B. *Procedure.* Using our roster of those trainees who had been tested in the earlier study, we searched the January 1, 1973, FAA personnel roster for each participant to see if he was listed as an ATCS; identities of those trainees who were found on the roster were verified by social security number.

The statistical evaluation was performed by computing point-biserial correlations between the retention/termination criterion and specific measures from the earlier study.

The MTPB measures consisted of an equal variance composite of all the individual task measures. The composite was constructed by converting each individual measure to a standard score with a mean of 5 and a variance of 1; a

higher score reflected better performance. These derived measures were then combined by summation.

The instructor ratings were based on a simple form in which the instructor was asked to assign a number from 0 to 100 (70 being passing) indicating his estimate of a given trainee's potential of becoming a fully rated ATCS. For most trainees, ratings were obtained from three different instructors.

The CSC ATCS screening battery score, when available, was obtained from the official records for each trainee as provided by the appropriate personnel office. (Not all trainees entered the program through the CSC test route; i.e., some were accepted on the basis of qualifying experience. For this reason, CSC scores were not available on a number of trainees.)

III. Results.

A. *MTPB Measures.* Of the 13 correlations computed between the MTPB composite measures and the retention/termination criterion (Table 1, column 3), only one correlation was significant at the 0.05 level of confidence. This was the correlation ($r=0.39$) for subjects in Study II for whom CSC scores were not available, and, hence, it is assumed that these were trainees who were accepted into the program on the basis of qualifying experience and not their CSC scores. The subjects in Study II for whom CSC scores were available yielded a nonsignificant negative correlation with the MTPB composite score ($r=-0.09$). Thus, the magnitude of the marginal correlation ($r=0.23$; $p<0.09$) for all subjects in Study II should presumably be attributed to the non-CSC subgroup.

B. *Instructor Ratings.* All but 2 of the 14 correlations between instructor ratings and the retention/termination criterion were significant at the 0.05 level or better; for those 2, $p<0.10$. Of the 12 significant correlations, 3 were significant at the 0.05 level and 9 were significant at the 0.01 level (Table 1, column 5).

C. *CSC Scores.* None of the three correlations between the available CSC screening battery scores and the retention criterion was significant. However, one of these correlations was of moderate magnitude ($r=0.22$; $p<0.11$).

Table 1. Correlations obtained for trainees in Studies II through V (a) between retention criteria and scores obtained from the CSC test, the MTPB, and ratings by Academy instructors and (b) between MTPB scores and ratings by instructors

Study	Trainees N	Terminations N	Predictors						
			MTPB Scores		Instructors Ratings		CSC Scores		Instructor Ratings vs. MTPB
			r	p	r	p	r	p	
(1) II through V	214	55	0.06		0.41	<0.01			
(2) II through V (Less Academy failures)	204	45			0.29	<0.01			
(3) II through IV	125	33	0.10		0.38	<0.01			
(4) II (All trainees)	61	16	0.23	<0.09	0.26	<0.05			0.50 <0.01
(5) II (Trainees with CSC scores)	28	7	-0.09		0.31	<0.10	-0.09		0.31
(6) II (Trainees with no CSC scores)	33	9	0.39	<0.05	0.27	<0.10			0.59 <0.01
(7) III	31	7	-0.11		0.39	<0.05			0.53 <0.05
(8) IV	28	5	0.09		0.50	<0.01			0.07
(9) V (All trainees)	89	22	0.00		0.53	<0.01			0.24
(10) V (Less Academy failures)	79	12	0.02		0.28	<0.05			
(11) V (Trainees with CSC scores)	61	14	0.08		0.47	<0.01	0.22 <0.11		0.36 <0.05
(12) V (Trainees with no CSC scores)	28	7	-0.07		0.55	<0.01			0.47 <0.05
(13) II + V (Trainees with CSC scores)	88	21	-0.13		0.40	<0.01	0.11		
(14) II + V (Trainees with no CSC scores)	61	16	0.24	<0.08	0.39	<0.01			

IV. Discussion.

This followup study posed questions concerning the effectiveness of predicting trainee retention in the ATC system by MTPB performance measures, Academy instructor ratings, and CSC aptitude test scores.

The most impressive result obtained was the statistically demonstrated ability of the FAA Academy instructors to predict the retention/termination criterion. Even the nonsignificant correlations were of substantial magnitude (0.31 and 0.27). Apparently, the instructors were able to tap sources of variance in their association with the trainees that were of considerable validity in estimating whether a trainee would remain in the ATC specialty. Probably the major factor underlying the obtained validity was that the potential of the trainee was foremost in the minds of the instructors as they observed the trainees in various phases of the Academy program. In addition, the instructors no doubt had opportunities to hear comments and observe behavior that might be reflective of attitudes and motivations of relevance to the future performance of the trainees. Whatever may have been the source of the validity, it was substantial, and these results are clearly in direct support of the earlier findings of Brokaw¹ and Trites.³

The finding of predictive validity for the instructor ratings also lends support to the appropriateness of their use as a criterion in the earlier study with the MTPB. Specifically, the present

results indicate that ratings of this kind are the "best available criterion" for short term (non-longitudinal) studies.

The interpretation of the data with respect to the CSC scores as a predictor of the retention criterion should give proper attention to two factors. First, on the assumption that those trainees who had CSC scores were selected on the basis of the aptitude test and therefore had suitably high scores on that test, it can be argued that the bulk of the predictive validity of the CSC battery would have been "used up" in the initial screening process. Second, the trainees in Study II had gone through additional screening both at the Academy (i.e., had graduated successfully from the Academy) and on the job through the normal performance appraisal process. Thus, a strong defense is available for the argument that the retention or termination of the Study II trainees would be dependent on attitudinal, motivational, and other factors that the CSC screening battery does not purport to measure. For these reasons, the small negative correlation between the CSC scores and the retention criterion is not surprising and should not be interpreted as reflecting unfavorably on the value of the CSC battery as a screening device; the battery would have already done its job as an initial screening instrument by eliminating those who were least likely to succeed in ATC work. With respect to Study V, the trainees having CSC scores in that study had gone

through, at the time at which we tested them, only the screening associated with the selection process. Thus, the small positive correlation ($r=0.22$; $p<0.11$) is a somewhat less ambiguous—though still incomplete—index of the value of the CSC scores as a predictor of retention on the job.

The interpretation of the results with respect to the MTPB measures must take into account not only the screening factors that were outlined in connection with the CSC scores but also the characteristics of the studies in which the MTPB data were collected. As noted in the introduction, this was a series of essentially exploratory studies; in general, the question being attacked was, Does the MTPB approach to ATCS selection offer sufficient promise to warrant *further* research? Thus, as is common in exploratory research, changes in the task configurations were made from study to study. The reasons for the changes were discussed in the earlier report,² and, therefore, they will not be repeated here. However, it is relevant to note that the changes were in the general nature of increasing the overall difficulty of the task complex in Studies III, IV, and V. Although there is no good rationale as to why the increased difficulty *per se* would be expected to have resulted in poorer predictive validity of the MTPB measures for the retention criterion, it is quite possible that a byproduct of the changes would have had an effect. The reasoning is as follows: In all but Study I, the trainees were fully informed as to the intent of the study; *viz*, to evaluate the MTPB as an ATCS selection device. In any such study, it is very important that the test device and test situation be, at least in a vague sense, “believable” to the trainees. In retrospect, although we did not fully appreciate it at the

time the studies were being carried out, there is reason to suspect the difficulty of the test was pushed to, and perhaps beyond, the believability limit. Unfortunately, no record was made of the best source of evidence on this point—the comments of the trainees during and after the testing. But, to the extent that the subjects did react to the test situation as being unrealistically difficult, unknown but clearly undesirable effects on motivation would be expected.

At this time, we can marshal no factual evidence to support the hypothesis that the lower level of difficulty of the task complex in Study II was the reason for the significant correlation between the MTPB measures and the retention criterion. However, on purely rational grounds, the hypothesis is not unreasonable.

With respect to the conclusions to be drawn in regard to the MTPB measures as a predictor of retention of trainees as ATCS's, this followup study inherited the features introduced by the exploratory nature of the original study. The most significant of these features was that changes were made in the task complex from study to study. Thus, direct support for arguing that the MTPB has predictive power with respect to the potential of ATCS trainees derives from the results with the subjects of Study II. However, indirect support is also provided in that (1) the instructor ratings served as a good predictor of the retention criterion and (2) the MTPB was a good predictor of the instructor ratings in the earlier study. Thus, although it must be concluded that final answers to the main question posed in the study can be obtained only through further validating research, these results and those of the earlier studies provide a basis for optimism with respect to the expected outcome of such research.

REFERENCES

1. Brokaw, L. D.: School and Job Validation of Selection Measures for Air Traffic Control Training. USAF: WADC-TN-59-39 (AD 214-884), 1959.
2. Chiles, W. D., A. E. Jennings, and Georgetta West: Multiple Task Performance as a Predictor of the Potential of Air Traffic Controller Trainees. FAA Office of Aviation Medicine Report No. AM-72-5, 1972.
3. Trites, D. K.: Problems in Air Traffic Management: I. Longitudinal Prediction of Effectiveness of Air Traffic Controllers. Civil Aeromedical Research Institute Report 61-1, Oklahoma City, Oklahoma, 1961.