

PROBLEMS IN AIR TRAFFIC MANAGEMENT:
I. LONGITUDINAL PREDICTION OF
EFFECTIVENESS OF AIR TRAFFIC CONTROLLERS

DAVID K. TRITES, Ph.D.*

**Chief, Selection Section
Psychology Branch*

61-1

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

December 1961

PROBLEMS IN AIR TRAFFIC MANAGEMENT:
I. LONGITUDINAL PREDICTION OF EFFECTIVENESS OF
AIR TRAFFIC CONTROLLERS

DAVID K. TRITES, Ph.D.

Current (1961) job performance evaluations and medical history data were obtained for 149 of 197 men who were trained in air traffic control work in 1956. Evaluations of psychological test and biographical data collected at the time they went through training indicate that: (1) Psychological tests can make a useful contribution to screening applicants for air traffic control work; (2) Instructors in the air traffic control school can make exceptionally valid predictions of job performance evaluations some years later; (3) Older trainees tended to receive poorer job performance ratings some years later than did their younger classmates; (4) Medical history information of the kind collected in this study is not predictable by the psychological tests which were used.

In 1956, representatives of the Aeronautical Center of the Civil Aeronautics Administration (now the Federal Aviation Agency Aeronautical Center), together with members of the Personnel Laboratory of the United States Air Force, discussed procedures for the selection of air traffic management personnel. A joint research project was arranged and experimental selection testing of air traffic controller (ATC) trainees was begun later in the year at the Aeronautical Center in Oklahoma City. The results of this project have been reported by Brokaw (1). His findings indicated that academic grades and instructor ratings, obtained immediately after training, and on-the-job performance ratings made by supervisors within the first year after training could be successfully predicted by psychological tests.

The present investigation represents a five year follow-up of the subjects tested in 1956 to determine if current job performance evaluations, retention in ATC work, incidents of unsatisfactory ATC work, and medical history information could be predicted by the data collected in 1956 and 1957.

PROCEDURE

Predictor Variables. — The Personnel Laboratory, USAF, supplied the basic data for all measures described by Brokaw (1). A listing of these is given in Table 1.

Sample. — Regional Offices of the Federal Aviation Agency (FAA) were able to supply current FAA facility addresses, or other information, on all but 10 of the original 197 subjects. Since no information was available for these 10 men, their data were excluded from all regression analyses. Of the remaining 187 subjects, 16 had failed the training course and left the FAA early in 1957, 15 who had passed the training course had left the FAA, 2 were deceased, replies were not received for 2, and 3 were with the FAA but no other information was available. This left 149 subjects (including 4 training course failures still with the FAA) for whom relatively complete criterion data were obtained.

Criterion Data Collection Method. — A letter describing the project and data collection forms was sent to the Chiefs of the facilities at which

TABLE 1**Listing of Predictor Variables¹***Numerical and Reasoning Tests*

1. Dial & Table Reading (AF)
2. Test 6, Cal. Capacity Quest.
3. Number Series, Cal. Test of Mental Maturity
4. Numerical Quantity, Cal. Test of Mental Maturity
5. Arithmetic, Pers. Selection and Classification Test
6. Numerical Quantity, DAT
7. Air Traffic Problems I
8. Arithmetic Reasoning, AC 2A

Abstract Reasoning & Perceptual Tests

9. Test 5, Cal. Capacity Quest.
10. Abstract Reasoning, DAT
11. Space Relations, DAT
12. Aerial Landmarks, AFOQT
13. Spatial Orientation, AFOQT
14. Instrument Comprehension AFOQT

Verbal Tests

15. Test 7, Cal. Capacity Quest.
16. Reading, Pers. Selection and Classification Test
17. Language Usage, Sentences, DAT
18. Verbal Test, AC 2A

Clerical Speed and Accuracy Tests

19. Code Translation, Survey of Working Speed & Accuracy
20. Counting, Survey of Working Speed & Accuracy

Temperament Tests

21. Family Relations, CTB, Cal. Test of Personality
22. Nervous Manifestations, CTB, Cal. Test of Personality

Biographical and Experience Variables

23. Age
24. Education
25. Marital Status
26. Previous Flying Experience
27. Airport Traffic Control Experience
28. Ground Control Approach Experience
29. Any Air Traffic Experience
30. Senior FAA Rating
31. FAA Certification in any status

¹ Complete descriptions and sources of all psychological test variables available upon request.

the men were located. For each of the original men at his facility the Chief was asked to supply promotional and job information, an indication of any disciplinary actions taken as a result of violations of air traffic rules or procedures, and medical history information. In addition, each Chief was asked to have four of his supervisors rate each subject on a job performance evaluation form. The form contained items related to job performance, ability as a controller, judgment, and personality characteristics.

From these two sources of information, six criterion measures were developed. These

were: (1) Average Supervisor Rating; (2) Active vs. Inactive Controller; (3) With the FAA vs. Not With the FAA; (4) Mean Hours of Sick Leave; (5) No Symptoms vs. Symptoms; and (6) No Disciplinary Action vs. Disciplinary Action. More complete descriptions of the criterion variables are given in Table 2, together with their means and standard deviations for the maximum number of subjects on whom the criterion variable had been obtained.

Table 2 also contains similar information for the criterion variables collected in 1956 and 1957, and used in the original study.

TABLE 2

Description of the 1957 and 1961 Criterion Measures

<i>Criterion Variable</i>	<i>N</i>	<i>Mean</i>	<i>Std. Dev.</i>
1957—Average Lecture Grade: The lecture grade attained by each student in the training course.	195	92.4	4.20
1957—Composite Instructor Rating: A numerical transformation of the consensus of ATC instructors relative to the student's standing in nine topical areas relating to work habits, ability, emotional stability, and the exercise of judgment and reasoning. Ratings were obtained after briefing instructors on rating theory, halo effect, and objectivity.	188	9.93	5.42
1957—Average Supervisor Rating: A numerical transformation of the consensus of supervisors relative to the individual's standing on eleven areas relating to work habits, ability, and the exercise of judgment and reasoning. Form was almost identical to one used by instructors except for two additional items concerned with demonstrated aptitude for ATC work and potential ability. Forms and rating instructions were supplied by mail.	170	27.1	5.95
1961—Average Supervisor Rating: A numerical transformation of ratings collected from 1 to 4 supervisors of each individual on a form containing the same items used for the 1957 Average Supervisor Rating. Three new items related to the individual's emotional stability and relationships to others were added. Forms and rating instructions were supplied by mail. (Using individuals with 2 or more forms, a corrected split-half reliability of .75 was obtained for the derived scores.)	149	33.9	8.37
1961—Active vs. Inactive Controller: Individuals were dichotomized as being either still active in controller work or not active.	169	.822	.382
1961—With the FAA vs. Not with the FAA: Individuals were dichotomized as being either with the Federal Aviation Agency or not with the FAA.	169	.882	.323
1961—Mean Hours of Sick Leave: The mean number of hours of sick leave taken in the years 1957 through 1960 was computed for each individual.	142	41.1	34.9
1961—No Symptoms vs. Symptoms: A dichotomy representing medical complaints of individuals as known to and reported by their facility Chiefs.	146	.829	.377
1961—No Disciplinary Action vs. Disciplinary Action: A dichotomy representing the presence of or absence of disciplinary actions taken as a result of control errors. Reported by facility Chiefs for individuals.	148	.824	.380

Statistical Methodology.— Because the report by Brokaw did not contain data sufficiently detailed to permit synthesizing proper raw-score regression equation weights and because of a minor, but potentially important, limitation in his analyses, new regression equations were computed. Using an iterative multiple regression procedure described by Greenberger and Ward (2), and Bottenberg (3), two regression equations and their corresponding multiple correlations, beta weights, raw-score weights and constants were computed for each of the 1956-1957 criterion variables described in Table 2. For each criterion, one equation was based only on psychological tests (Variables 1 to 22 in Table 1,) whereas the second equation was based on the psychological tests plus the three biographical variables, age, education, and marital status. Since two of the six equations were found to be identical, a subsequent regression analysis was considered as Number 6. This sixth equation utilized all psychological tests, the three biographical variables, and the 1956-1957 Average Lecture Grade and the Composite Instructor Rating as predictors of the 1957-Average Supervisor Rating.

The correlation matrix for computation of all regression equations was based on a common N of 135 subjects who had data for all variables and had successfully completed the training course. The inclusion of variables in each equation was terminated when no further statistically significant increase in the magnitude of the multiple correlation could be obtained.

Making several assumptions regarding the parameters of the data used in Brokaw's analyses, approximate raw-score regression weights were computed for the psychological test variables which he recommended as predictors. This has been named Regression Equation

Number 7. Seven predicted scores were then computed for each subject using the raw score regression weights for the seven equations; and all variables described in Tables 1 and 2 and the seven predicted values were intercorrelated by a computational routine permitting each correlation to be determined on the maximum possible number of cases having the appropriate data.

Finally, a number of partial correlations were computed in order to assess the relationships between the 1961 criterion measures and other variables independent of the influence of subjects' age.

Since several matrices of first order correlations based upon varying numbers of subjects were available, the correlations used in the computation of each partial were those which seemed to be most nearly comparable.

RESULTS

Regression Analyses.— The results of regression analyses 1 through 6 are presented in Table 3. Equations 1, 2, and 3 represent the psychological test variables plus the three biographical variables; Equations 4 and 5 are for the psychological variables alone; and Equation 6 is for the psychological test variables, three biographical variables, and the two training school criterion measures. Table 4 contains the regression data given by Brokaw (1). In all instances, highly significant multiple correlations were obtained.

The overlap in the variables selected for each of the seven equations against the three different criteria suggested that there is a considerable amount of variance common to all three criteria. However, the crucial question concerns the relationships between the seven predicted values and the six 1961 criterion measures.

TABLE 3

Most Efficient Combinations of Psychological Tests, Biographical Variables, and the 1957 School Criterion Measures for Prediction of Average Lecture Grade (Equations 1 & 4), Composite Instructor Rating (Equations 2 & 5), and the 1957 Average Supervisor Rating (Equations 3 & 6)

Variable		Regression Equations ¹											
		Psych. Tests & Biog.			Psych. Tests Only		Psych. Tests, Biog., & 1957 School Criteria						
		1	2	3	4	5	6						
No.	Description	B†	r‡	B	r	B	r	B	r	B	r		
<i>Test Variables</i>													
1	Dial & Table Reading (AF)	20	40**	27	37**	22	33**	14	40**	25	37**		
2	Verbal Knowledge & Reasoning (Test 6, Cal. Capacity Quest.)					23	30**					28	30**
7	Air Traffic Problems I	17	25**	25	32**			16	25**	24	32**		
8	Arithmetic Reasoning, AC 2A	24	39**					21	39**				
9	Symbolic Reasoning & Perceptual Speed (Test 5, Cal. Capacity Quest.)	21	35**	20	27**			19	35**	19	27**		
10	Abstract Reasoning, DAT					22	27**					23	27**
12	Aerial Landmarks, AFOQT					-18	07					-17	07
16	Reading, Pers. Select. & Class. Test							16	28**				
21	Family Relations, CTB, Cal. Test of Pers.					19	22**					50	22**
22	Nervous Manifestations, CTB, Mental Health Analysis	21	22**					26	22**				
<i>Biographical Variables</i>													
25	Marital Status	16	09*	16	09*								
<i>1957 Criterion Variables</i>													
	Composite Instructor Rating											60	56**
<i>Multiple Correlations</i>			59**		50**		48**		59**		47**		65**

¹ Intercorrelation matrix for equations was based on a common N of 135 trainees who completed the training course. All decimal points have been omitted from table entries.

† Regression Coefficients (Betas) for standardized scores.

‡ Validity Coefficients.

* Significant at less than the .05 level.

** Significant at less than the .01 level.

* Point biserial correlation

TABLE 4

**Regression Coefficients for Standardized Scores (Betas) Reported by
Brokaw and Used to Derive Equation Number 7***

<i>Variable</i>		
<i>No.</i>	<i>Description</i>	<i>Regression Coefficients†</i>
7	Air Traffic Problems I	20
8	Arithmetic Reasoning, AC 2A	11
9	Symbolic Reasoning and Perceptual Speed (Test 5, Cal. Capacity Quest.)	16
19	Code Translation, Survey of Working Speed and Accuracy	05

* Table entries were taken from: Brokaw, L. D. School and job validation of selection measures for air traffic control training. WADC-TN-39, Pers. Lab., WADD, USAF, Lackland AFB, Texas, 1959.
† Decimal points omitted.

Table 5 contains the correlations between those regression equations which were significantly related to the 1961 criterion measures together with partial correlations eliminating the effect of age. Only Regression No. 2 was significantly related to the Disciplinary Action criterion and this at a relatively low level. Regression Nos. 5, 6 and 7 were all significantly related to the 1961 Average Supervisor Rating. If we note the correlations between the 1956-1957 criterion measures and the 1961 criteria, also contained in Table 5, it is not surprising that Regression No. 6, which included the Composite Instructor Rating as a predictor variable, had the highest correlation with the 1961 Average Supervisor Rating. What was surprising, was the remarkably high correlation between the Composite Instructor Rating and the 1961 Average Supervisor Rating. Obviously, the instructors in 1956 were making exceptionally valid judgments concerning a trainee's potential for air traffic control work.

The correlation between Regression No. 5, newly computed using only psychological tests as predictors, and the 1961 Average Supervisor Rating became insignificant when age was partialled out. However, Regression No. 7, which utilized tests originally recommended by Brokaw (1), remained significant when age was eliminated. The significance of the latter

regression leaves no doubt that a valuable contribution can be made to the selection of air traffic controllers by the use of psychological tests.

A comparison of the correlations between the 1956-1957 criterion measures and the 1961 criterion measures indicated that all three of the earlier measures could predict the 1961 Average Supervisor Rating and Active vs. Inactive Controller status, whereas retention in the FAA was predicted only by the 1956-1957 Composite Instructor Rating and the 1957 Average Supervisor Rating. These relationships remained essentially unchanged, as indicated by the partial correlation coefficients in Table 5, when the influence of age was removed statistically.

Individual psychological test and biographical measures which were significantly related to any of the 1961 criteria are presented in Table 6. Two tests of the abstract reasoning type and two of the space relations or orientation type were predictive of the 1961 Average Supervisor Rating. When the effect of age was eliminated from the correlation, one of the abstract reasoning tests no longer was significantly related to the criterion. However, the present findings, and the fact that Brokaw found these four tests significantly related to one or more of the ATC school criterion meas-

TABLE 5

First Order and Partial Correlations Between the 1961 Criterion Measures, the 1957 Criterion Measures, and Scores Predicted from Regression Equations Where the First Order Correlations with the 1961 Criteria are Significant

Criterion	Variable	Criterion		Age		Criterion <i>r</i>
		<i>r</i> ¹	N	<i>r</i>	N	with Age Partialed Out
1961 Average Supervisor Rating	Age	-23**	149			
	Average Lecture Grade	24**	148	-05	135	23**
	Composite Instructor Rating	45**	149	-18*	135	43**
	1957 Average Supervisor Rating	33**	143	-07	135	32**
	Regression Equation 5	17*	143	-14	133	14
	Regression Equation 6	44**	127	-09	133	43**
	Regression Equation 7	23**	143	-11	133	21*
Active vs. Inactive Controller	Age	-15**	169			
	Average Lecture Grade	26***	168	-14	195	24**
	Composite Instructor Rating	24***	169	-24	188	21**
	1957 Average Supervisor Rating	24***	162	-11	170	23**
With FAA vs. Not With FAA	Age	-03*	169			
	Composite Instructor Rating	16**	169	-24	188	16*
	1957 Average Supervisor Rating	20**	162	-11	170	20*
No Disci- plinary Action vs. Discipli- nary Action	Age	-02*	133			
	1957 Average Supervisor Rating	28***	142	-07	135	28**
	Regression Equation 2	17**	138	-06	133	17*

- ¹ Decimal points have been omitted.
- * Significant at less than the .05 level.
- ** Significant at less than the .01 level.
- Point-biserial correlations.

TABLE 6

First Order and Partial Correlations Between the 1961 Criterion Measures and Psychological Test and Biographical Measures Where the First Order Correlations with the 1961 Criteria are Statistically Significant

Criterion No.	Variable Description	Criterion		Age		Criterion <i>r</i> with Age
		<i>r</i> ¹	N	<i>r</i>	N	Partialed Out
1961 Average Supervisor Rating	23 Age	-23**	149			
	9 Symbolic Reasoning & Perceptual Speed (Test 5, Cal. Capacity Quest)	21**	149	-17*	135	18*
	10 Abstract Reasoning, DAT	18**	149	-26**	135	13
	11 Space Relations, DAT	18*	148	-06	135	17*
	13 Spatial Orientation, AFOQT	23**	143	-10	135	21*
	26 Previous Flying Exper.	-19**	139	67***	135	-05
	29 Any Air Traffic Exper.	18**	139	-57***	135	06
Active vs. Inactive Controller	23 Age	-15**	169			
	21 Family Relations, CTB Cal. Test of Personality	21***	150	-03	175	21**
With FAA vs. Not With FAA	23 Age	-03*	169			
	24 Education	-16**	168	44**	196	-16*
Mean Sick Leave	23 Age	-18**	142			

- ¹ Decimal points have been omitted.
- * Significant at less than the .05 level.
- ** Significant at less than the .01 level.
- Point-biserial correlations.

ures, indicates tests related to these aptitude areas should be included in any predictor test battery.

The only other individual psychological test found to be significantly related to a 1961 criterion was the Family Relations Scale with Active vs. Inactive Controller status. This relationship was independent of age. Although this suggests that personality scales may be of some potential use, further refinement of these measures will be necessary before they can be recommended for operational use in a selection test battery.

From the negative relationship between age and the 1961 Average Supervisor Rating, it may be inferred that older trainees are not as likely as their younger classmates to be considered by supervisors some years later as being among the better controllers. Age, also, was negatively related to Active vs. Inactive Controller status which corroborates the inference that older trainees do not tend to be as successful as younger men.

It was surprising, however, to find age negatively related to Mean Hours of Sick Leave. This implies that older men are less often ill than younger men. Obviously, this requires further exploration of the factors behind the relationship.

Two biographical variables, Previous Flying Experience and Any Air Traffic Experience, were significantly related to the 1961 Average Supervisor Rating; but when the effect of age was removed, these relationships became negligible. Thus, none of the biographical variables representing previous experience was related to subsequent job performance. It may be that these experience variables which, then and now, are being used as partial requirements for entry into ATC training may keep some potentially unsatisfactory individuals out of ATC work; but they certainly have little relationship to subsequent performance as a controller.

Finally, Education was found to be negatively related to retention in the FAA. The reasons for this are not at all clear and further research is needed to clarify the causative factors.

DISCUSSION

The present investigation has indicated that psychological tests can make a useful contri-

bution to personnel selection for air traffic control work. In addition, it has demonstrated that the most effective prediction of current supervisory opinion of a controller's proficiency can be obtained from the controller's ATC school instructors, assuming that the same instructor-curriculum-student relationships have continued to exist as prevailed in 1956-57.

The relative effectiveness of psychological tests and the instructor's evaluation of a student as predictors of job performance ratings are presented in Figures 1, 2, 3, and 4. The first two figures were derived by identifying those individuals in the low 25 percent of the scores predicted from Regression Equation 7 using psychological tests only, in the case of Figure 1, and the low 25 per cent of the Instructor Ratings made in 1956 in the case of Figure 2. The per cent of the individuals in each quarter of the 1961 Average (Mean) Supervisor Ratings was then determined and plotted.

The second two figures were derived by identifying those individuals who were either in the lowest 25 per cent of the 1961 Average Supervisor Ratings or who had been cited for violations of air traffic rules and regulations. The per cent of these 54 marginal individuals in each quarter of the scores from Regression Equation 7 was determined and plotted in Figure 3; similarly, the per cent of marginal individuals in each quarter of the 1956 Instructor Ratings has been plotted in Figure 4.

Considering the correlations previously discussed, the superior predictive efficiency of the Composite Instructor Rating would be expected. But, it must be remembered, these predictions were made after trainees had been recruited, hired by the organization, and put through a lengthy and expensive training program. Some loss in efficiency of prediction by using psychological tests instead of instructors could well be tolerated on economic grounds.

REFERENCES

1. Brokaw, L. D. School and job validation of selection measures for air traffic control training. WADC-TN-59-39, Pers. Laboratory, WADD, USAF, Lackland AFB, Texas, 1959.
2. Greenberger, M. H. and Ward, J. H. Jr. An iterative technique for multiple correlation analysis. IBM Tech. Newsletter No. 12, IBM Co., New York, N. Y., 1956.
3. Bottenberg, R. A. An iterative multiple regression procedure using the IBM 650. WCLL-TM-60-3, Pers. Laboratory, WADD, USAF, Lackland AFB, Texas, 1960.

FIGURE 1

Percent distribution (by fourths) of mean supervisors' ratings of controllers in the lowest quarter of predicted scores from regression equation.

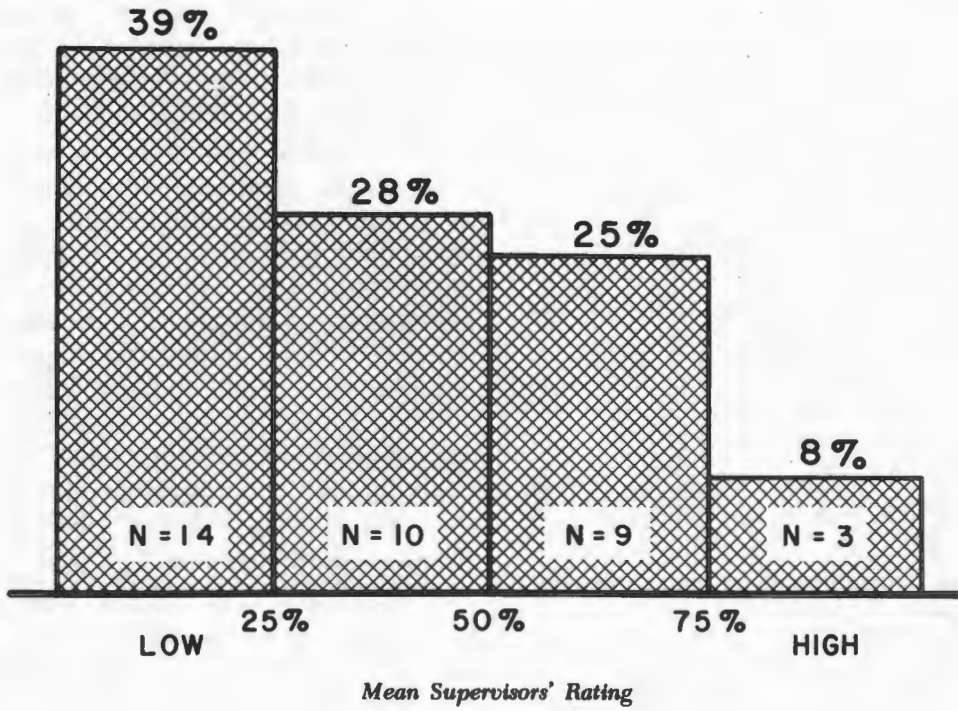


FIGURE 2

Percent distribution (by fourths) of mean supervisors' ratings of controllers in the lowest quarter of the Instructors' Ratings.

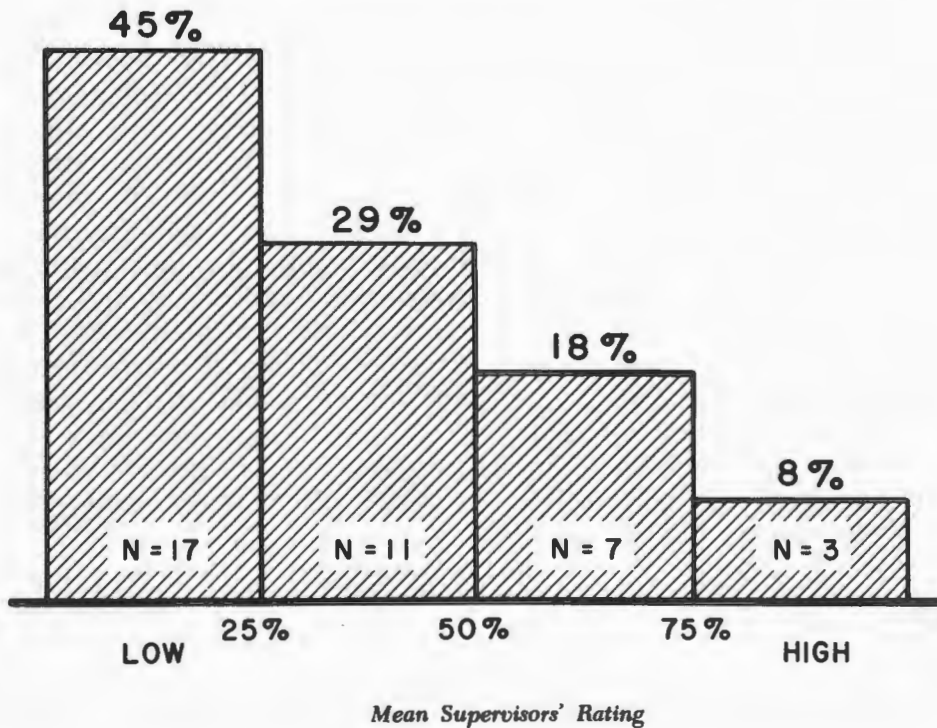


FIGURE 3

Percent of marginal and satisfactory controllers in approximate fourths of scores predicted from regression equation.

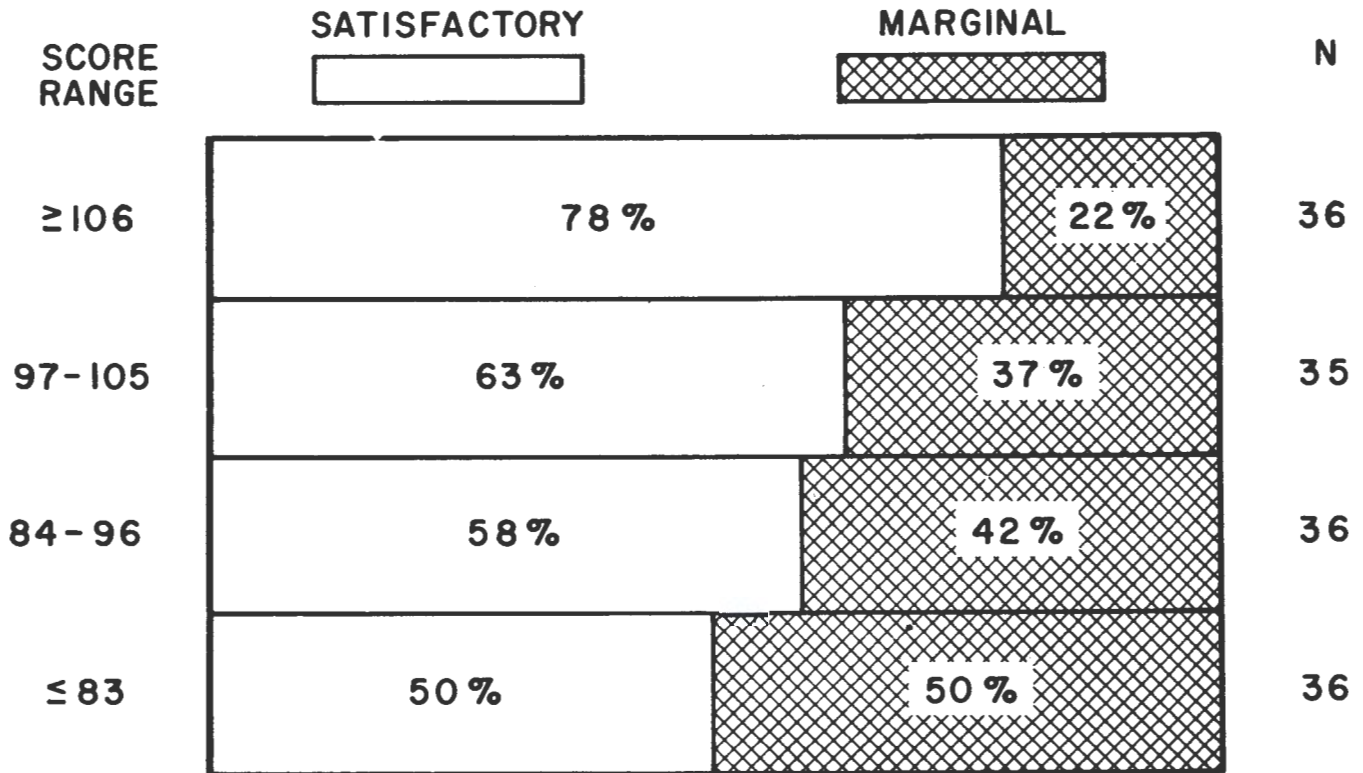


FIGURE 4

Percent of marginal and satisfactory controllers in approximate fourths of instructors' ratings.

