UNCLASSIFIED

AD 263 916

Reproduced
by the

ARMED SERVICES TECHNICAL INFORMATION AGENCY
ARLINGTON HALL STATION
ARLINGTON 12, VIRGINIA

UNCLASSIFIED

Best
Available
Copy

NOTICE: When government or other drawings, specifications or other data are used for any purpose other than in connection with a definitely related government procurement operation, the U. S. Government thereby incurs no responsibility, nor any obligation whatsoever; and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use or sell any patented invention that may in any way be related thereto.

263916

U. S. NAVAL SCHOOL OF AVIATION MEDICINE
U. S. NAVAL AVIATION MEDICAL CENTER
PENSACOLA, FLORIDA

IMPROVING THE PREDICTIVE EFFECTIVENESS OF PEER RATINGS

CATALOGED BY ASTIA
AS AD No.

SPECIAL REPORT NO. 61-7

by

Ensign R. E. Doll, MSC, USNR

and

Ensign A. A. Longo, MSC, USNR

Approved by

Captain Ashton Graybiel, MC, USN
Director of Research

Released by

Captain Clifford P. Phoebus, MC, USN
Commanding Officer

61-4-6
XEROX

14 July 1961

Opinions or conclusions contained in this report are those of the authors. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department. Reference may be made to this report in the same way as to published articles noting author, title, source, date, and report number.

SUMMARY PAGE

THE PROBLEM

Low scores on peer ratings of leadership potential have been shown to predict failure to complete the Naval Air Training Program. It is hypothesized here, however, that low peer ratings awarded by reason of "anti-social" personality traits are invalid as predictors of training failure and that the predictiveness of low peer ratings would be improved by removing such cases from the sample.

FINDINGS

Of 278 low-rated cadets, 73 were identified as having been low-rated for "anti-social" reasons. There were significantly fewer failures to complete training among these men than among men who were low-rated for other reasons.

RECOMMENDATIONS

On the basis of these findings it is recommended that any administrator utilizing a student's low peer rating score as an aid in deciding his probability of successfully completing the flight program consider the reasons given for the low rating.

If a large proportion (about one-third or more) of the total reasons given are of the kind listed on page 2 of this report, the predictive validity of a low peer rating score should be doubted.

INTRODUCTION

Peer ratings, i.e., evaluations of the individuals in a group by one or more other individuals in that group, have proved to be useful instruments. Such ratings, even though made by untrained and relatively unsophisticated observers, have been shown to be good predictors of relative success or failure in several areas of endeavor. Studies have indicated that such ratings have substantial validity in predicting flight failure (1), officer efficiency ratings (5), military grades in Officer Candidate School (4), leadership performance in combat (3), and on-the-job performance (2).

Such peer ratings are among the measures used in the Naval Air Training Program to appraise the potential of individual cadets. During the eighth week of training each man in a class of cadets is asked to name the three most promising prospective officers and the three least promising in his class. The raters may give their reasons for each high rating and must give their reasons for each low rating. It has been shown (1) that these ratings typically have a bi-serial correlation of about .35 with subsequent failure to complete the training program and that, when combined with other measures, they have considerable administrative usefulness.

Inspection of the reasons given for the award of low ratings shows that these ratings, superficially at least, would appear to fall into several unrelated categories. Further, some of these categories, such as low ratings given for poor motivation or low ability level, would seem to be much more obviously related to subsequent training failure than would low ratings given for "anti-social" personality traits. If low ratings given for differing reasons are differentially valid, it follows that the validity of the peer ratings could be improved by modifying the scoring of low ratings that are given for nonvalid reasons.

This study investigates the extent to which one such category of low ratings, "anti-socialness," can be reliably identified and compares the subsequent training success of men rated low for "anti-social" reasons and those rated low for other reasons.

THE CLASSIFICATION PROCEDURE

Twenty-two statements describing behavior that could be considered "anti-social" were chosen from 300 previously executed peer ratings which were not to be used otherwise in the study. These were incorporated in the following instructions:

"Please take time to become familiar with the following descriptions:

1. High strung
2. Lacks self-control
3. Uncooperative
4. Strong-headed
5. Cannot get along
6. Self-centered
7. Overbearing
8. Doesn't mix well
9. Disagreeable
10. Not well-liked
11. Gripes too much
12. Selfish
13. Argues a lot
14. Doesn't care about others
15. Lusts for power
16. Dominant
17. Tends to create friction
18. Quick-tempered
19. Thoughtless or unkind
20. Boastful
21. Lacking in teamwork
22. Poor personality, per se

If a reason for a low rating reflects the exact meaning of one of the above statements, classify it under the A (anti-social) category. It must require no interpretation. If it requires interpretation or is different from the above statements place it in the O (other) category."

Using these trait descriptions, four judges (other research psychologists) were asked to categorize independently 100 reasons given for low ratings. There was complete agreement on 93 items, three judges agreed on 6 items, and on one item there was a 2 - 2 split. The classification procedure was therefore judged to be sufficiently reliable for this study.

The data for the analysis consisted of approximately 3000 individual peer ratings that had been given to 278 cadets, each of whose average peer rating score was at least one-half standard deviation below the mean rating score of all cadets.

Using the above statements and procedure and working in concert, the authors reviewed all of the statements made about these men by their raters. If all of the statements made about a man by a rater were A statements, or if all were O statements, a single appropriate tally was made. However, if a rater made mixed statements, a tally was made for the man both in A and in O. This procedure resulted in from 3 to 20 tally marks per cadet. Those cadets for whom one third or more of the statements were classified as A statements were (arbitrarily) identified as the "anti-social" rated group. Seventy-three cadets fell into this group.

RESULTS AND DISCUSSION

Table I compares the proportions of the A and O groups that completed training. The percentage of low rated cadets in A group who completed training is significantly larger than that in the O group. (Actually, the completion rate among the A group is practically the same as that for all cadets, regardless of peer rating.) It thus appears that the hypothesis of the lack of validity of low peer ratings based on "anti-social" personality traits is supported by the data. However, it seemed prudent to check the aptitudes, training grades, and mean peer ratings within the A and the O groups to make certain that there was not some other critical distinction between the groups. Available for this purpose were the individuals' scores on the Aviation Qualification Test (a general intelligence test given before entry into training); scores on the Flight Aptitude Rating (a composite of scores on mechanical comprehension; spatial apperception and biographical items--correlation with training success .35 - .40); and pre-flight school grade (reflecting performance in academics, physical education, and military training).

Table I

Comparative Proportions of A and O Groups Completing Training

	A Group		O Group		
Peer Rating	N	% Completing	N	% Completing	p*
- 0.5 σ	73	63	205	39	.001
- 1.0 σ	44	55	125	34	.001
- 1.5 σ	24	63	80	35	.02

*chi-square test for significance

Table II shows the mean scores and SDs of the successful and unsuccessful A and O groups on these variables.

Table II

Mean Scores and Standard Deviations of Completing and Noncompleting A and O Groups on Aptitudes, Grades, and Ratings

	A - Complete		O - Complete		
	Mean	SD	Mean	SD	p*
AQT	76.24	13.48	78.28	12.74	n.s.
FAR	6.58	0.93	6.41	1.23	n.s.
PF Grade	46.64	3.27	46.30	3.45	n.s.
Peer Rating	37.15	8.02	37.11	7.49	n.s.
	A - Noncomplete		O - Noncomplete		
AQT	68.88	12.63	75.00	12.70	.02
FAR	5.92	0.92	5.98	1.10	n.s.
PF Grade	46.59	3.23	44.63	3.61	.01
Peer Rating	36.44	5.84	35.00	8.51	n.s.

* t test for significance

Comparison of all A's with all O's shows that the only significant differences are between the noncompleting A's and O's on general intelligence and on pre-flight grades. Here the mean intelligence score of the noncompleting O's is significantly higher, but their mean pre-flight grade is significantly lower than that of the A's. Since AQT normally correlates about .50 with pre-flight grades, these data provide evidence of real differences in training behavior, and a strong inference that these are related to motivational differences specified by A and O categories.

Table III presents the results of the requirement of a more strict criterion for inclusion of a case in the A group--50 per cent versus the 33 1/3 per cent shown above. No important difference results from using the more rigorous criterion.

Table III

Comparison of the Effects of Requiring 33 1/3% versus 50% "A" Statements for Inclusion in "A" Category

	A Group		O Group	
	N	% Completing	N	% Completing
33 1/3% A Statements	73	63	205	39
50% A Statements	40	68	238	42

CONCLUSIONS

Low peer ratings of leadership potential when awarded because of "anti-social" personality traits were not valid for the prediction of failure to complete aviation training. If such cases are removed from the data, the accuracy of the remaining low ratings as predictors of training failure is improved.

It is possible, of course, that low peer ratings for "anti-social" reasons have validity for other criteria such as leadership or success as an officer, but this remains to be established.

REFERENCES

1. Berkshire, J.R., and Nelson, P.D., Leadership peer ratings related to subsequent proficiency in training and in the fleet. Special Report 58-20. Pensacola, Fla.: Naval School of Aviation Medicine, 1958.

2. Fuchs, E.R., Woods, I.A., and Harper, B.P., Prediction of job success in eight career ladders. PRB Research Report No. 997. Washington, D.C.: The Adjutant General's Office, 1953.

3. Lindzey, G., and Borgatta, E.F., Sociometric measurement. In Lindzey, G. (Ed.) Handbook of Social Psychology, Vol. I. Theory and Method. Cambridge, Mass.: Addison-Wesley Publishing Co., Inc., 1954, 405-448.

4. Suci, G.J., and Vallance, T.R., An analysis of peer ratings: II. Their validity as predictors of military aptitude and other measures in the Naval Officer Candidate School. Technical Bulletin No. 54-10. Washington, D.C.: Bureau of Naval Personnel, 1954.

5. Wherry, R.J., and Fryer, D.H., Buddy Ratings: Popularity contest or leadership criteria? Personn. Psychol., 2, 147-159, 1949.