IMPROVEMENT THROUGH TRAINING: A CRITERION IN TEST VALIDATION

by

Atmananda Sharma

Psychologial Research Wing, New Delhi

ABSTRACT

This study was undertaken to decide about an adequate criteria which may be adopted in validating aptitude tests used in Naval Training Schools. Three criteria—final score, residual gain and crude gain, were compared. The final decision regarding the choice of criterion was primarily based on logical considerations. Residual gain, which measures that aspect of final score which is uncorrelated with initial status and thus removes some of the non-appropriate variance from the criterion and further overcomes the inconsistencies which crude gain presents was found to be a more realistic criterion.

Introduction

One of the most difficult of all aspects of the validity problem is that of obtaining adequate criteria of what we are measuring. A distinction is made between three criteria—immediate, intermediate and ultimate. In Naval Gunnery tasks these are respectively illustrated by such evidence as complete training in the gunnery school, accuracy of firing on the practical range in operational training and accuracy of firing in combat. Ultimate criterion is more complex than immediate and intermediate indices of success. For this reason, as well as for the pressure of time and other practical considerations, ultimate criterion or 'on the job performance' has been rarely used in test validation.

'Scholastic achievement' or 'final grade' in a training course has been the most commonly used criterion for validating a test. This immediate criterion which is most readily available and saves time is by no means an ideal one. The ability to complete the training in a gunnery school is not necessarily identical with the ability to destroy enemy ships or withstand the greater and more enduring stresses of battle. Such sailors who cannot complete training never get to combat, the criterion is however suitable in a negative way. It is by no means simple one too. It is affected by a variety of factors. Firstly, herein it is often assumed that scholastic achievement, like intelligence is a unitary attribute of each individual. But this is far from the truth. Although there is generally a positive correlation between achievement in different school subjects, there is sufficient disagreement to permit an individual to receive marks all the way from A to Z in different subjects. Where a certain test may possess only a moderate or low correlation with averages of school-marks it may correlate very high

with specific subjects or courses. An attempt to single out the most significant aspects as criteria is most desirable but, this is seldom done in practice. *Secondy*, there is the problem of securing dependable measurements. It is well known that school marks as ordinarily assigned by instructors are rather poor metric material. Most marks are neither very reliable nor very valid indications of achievement. *Thirdly*, the final result may not adequately represent the performance of trainees in a training course because the training course does not ordinarily result in mastery of the job but rather develops skills fundamental to efficient learning on-the-job. Further the trainees differ in their initial level of performance because of their differences in prior education and experience.

In all the Naval Training Schools, sailors are drafted for specialisation courses from Direct Entry sailors as well as from Boys. The wide gap in their prior education, experience and age leads to diversity in pre-training achievement and initial training examination results. In such a situation, the usually overlooked relevant criterion, namely, the '*Improvement through training*' may constitute a more significant dimension of trainee performance than does final score.

How to measure this improvement through training? In the past *crude gain* (the arithmetic difference between marks on pre-training and marks on post-training achievement) had been used as a criterion. But this crude gain appeared to be uncorrelated with aptitude measures and with other measures of gain as well[1]. It was suspected that this property of unrelatedness of crude gain scores was derived from a peculiarity of the crude gain measure itself. Dubois[2] had advocated another measure of improvement and calling it a '*residual gain*' had defined it as that portion of the measure of final status which was not correlated with initial status. A residual gain score represents the difference between actual final score and the final score predicted from initial score.

Aims

The preliminary study was taken with three main aims in view:—

(1) To determine the correlations between different psychological tests and each of the three criteria—crude gain, residual gain and the final score.

(2) To determine the regression coefficients of psychological tests on each of the three criteria and to determine the multiple correlations between each of three criteria and various combinations of psychological tests.

(3) To compare the predictability of three criteria—crude gain, residual gain and final score—of trainee proficiency.

Method

The subjects of this study were 118 sailors drawn from successive Third Specialist (UCII/WWIII) courses, V to XI of the Torpedo and Anti-Submarine

School, Naval Base, Cochin. Before admitting them to the specialisation course proper they were administered a comprehensive examination as the pre-specialisation proficiency measure. Following 12 weeks of specialisation training in UC III/UW III course, a regular final examination (alternate form) was administered. These two examinations, one administered at the commencement and another at the end of the training, served as basis for calculating two gain measures—crude gain and residual gain. Final examination marks were used as the measure of final standing in the course.

Prediction

The five psychological tests constituting the TAS Classification Battery served as the predictors. These included two tests of intelligence, namely MGIT and PGIT. The three mechanical aptitude tests included were MCT, MAT, and MASYT.

The product-moment correlations between these predictors were calculated from standard scores made on these tests. These correlations are set out in Table 1.[4]

TABLE 1

Inter-Correlations of Standard Scores on Predictors

(N=118)

Test	No.	1	2	3	4	5
PGIT	1	..				
VGIT	2	·63	..			
MCT	3	·22	·25	..		
MAT	4	·21	·27	·87	..	
MASYT	5	·32	·13	·29	·04	..

Criteria Correlations

The 'final score criterion' (X) was the score obtained by a trainee in the final examination out of a maximum of 100 marks. The pre-specialisation proficiency measure (Y) was the score of the trainee out of a maximum of 100 in the pre-admission examination. The difference $(X-Y)$ between these two scores, pre-admission and the post-training examination results, constituted the crude gain criterion.

The predicted final score (X_1) of a trainee was obtained from the Y score by employing the following regression equation:

$$X_1 = a + by$$

IMPROVEMENT THROUGH TRAINING

The product moment correlations between the five predictors and each of the three criteria—final score, residual gain and the crude gain are set out in Table 2.

TABLE 2

Correlations between the Predictors and the Three Criteria

(N=118)

Tests Criteria	PGIT	VGIT	MCT	MAT	MASYT
Final Score ..	·32*	·25*	·26*	·02	·16
Residual Gain ..	·21†	·20†	·23†	·03	·14
Crude Gain ..	—·02	—·03	·00	·11	—·10

* Significant at 1 percent level.
† Significant at 5 percent level.

The correlations of the predictors with the final score were all positive and significant at 1 percent level (except two), but the relationship was only slight. The correlations of the predictors with residual gain were lower than those with final score but were positive and significant at 5 percent level (except two). The correlations of crude gain with the predictors were almost zero for all practical purposes and showed the unrelatedness of crude gain with aptitude measures. This suggested that there were real differences in the predictability of the three criteria.

Regression Coefficients

The regression coefficients of psychological test scores on the three criterion scores—final score, crude gain and residual gain, were calculated by Aitken's modified method with each pivot covered to unity, as set out by Thomson[5]. These are given in Tables 3, 4 and 5.

TABLE 3

Regression coefficients of Tests on Final Score

PGIT	VGIT	MCT	MAT	MASYT
·320				
·270	·080			
·250	·045	·193		
·230	·106	·949	—·881	
·399	·067	1·588	—1·482	—·496

If only PGIT was used, its regression coefficient was ·320. If PGIT and VGIT were used their regression coefficients were ·270 and ·080. If PGIT, VGIT and MAT were used, their regression coefficients were ·250, ·045 and ·193. If PGIT, VGIT, MCT and MAT were used the regression coefficients were ·230, ·106, ·949 and —·881. And if all the five Tests viz. PGIT, VGIT, MCT, MAT and MASYT were used, the regression coefficients were ·399, ·067, 1·588, —1·482 and —·496.

TABLE 4

Regression coefficients of Tests on Residual Gain

PGIT	VGIT	MCT	MAT	MASYT
·21				
·139	·112			
·120	·078	·187		
·104	·130	·819	—·736	
·229	·100	1·292	—1·103	—·367

TABLE 5

Regression coefficients of Tests on Crude Gain

PGIT	VGIT	MCT	MAT	MASYT
—·02				
—·179	·029			
·006	—·015	—·78		
·023	—·067	—·0723	·752	
—·077	—·0044	—1·102	1·108	·294

(*N.B.*—The tables 4 and 5 would be read and interpreted in exactly the same way as table 3).

Multiple Correlations

The multiple correlations between each of the three criteria and various combinations of predictor tests were computed and the significance of these

multiple correlations then tested by means of the analysis of variance. The multiple correlations are set out in Table 6.

TABLE 6
Multiple Prediction of Three Criteria

Serial No.	Battery of Tests	Multiple Correlation		
		Final Score	Residual Gain	Crude Gain
1	PGIT	·32†	·21*	·02
2	PGIT & VGIT	·33†	·23*	·06
3	PGIT, VGIT & MCT	·38†	·29†	·08
4	PGIT, VGIT, MCT & MAT	·57†	·45†	·37†
5	PGIT, VGIT, MCT, MAT & MASYT ..	·57†	·52†	·43†

* Significant at 5 percent level.
† Significant at 1 percent level.

The three criteria can be further compared by a study of the multiple correlations set out in Table 6. None of the multiple correlations of crude gain were significant at 1 percent level; only two of them were significant at 5 percent level. All the multiple correlations with residual gain were higher than those with crude gain and were found to be significant at 5 percent level, three of them were significant at 1 percent level also. Of course the multiple correlations with the final score were all significant at 1 percent level.

Conclusion

(1) The correlations of final score with the predictors were all positive and significant at 1 percent level, those of residual gain were lower than those of the final score and only three were found to be significant at 5 percent level. Crude gain, however, was found to be unrelated with aptitude measures which had been the finding of Dubois and other research workers.

(2) A comparison of the three criteria through a study of multiple correlations showed that the correlations of Crude gain were lowest and none of them significant at 1 percent level. The multiple correlations of residual gain were higher than those of crude gain and all were significant at 5 percent level, three were significant at 1 percent level as well. This was also true, of course, for multiple correlations with final score, all of which were all significant at 1 percent level.

(3) The high correlation of aptitude measures with the final score, it is quite possible, may be due to overlap of non-valid variance such as verbal facility, method of taking the examination and test wiseness. On the other hand the residual gain measures that aspect of final score which is uncorrelated with initial status and thus removes some of the non-appropriate variance from the

criterion. And it is, therefore, a better and more realistic criterion than the commonly adopted one of final score. Further-more it overcomes the inconsistencies which crude gain presents as a measure of aptitude.

(4) Lastly, the selection of a criterion in a particular situation is primarily based on logical considerations and Naval Training situations demand the adoption of 'Residual Gain as a Criterion'.

Acknowledgement

References

1. Woodraw, H., *Psychol. Rev.*, **53**, 147, 1946.
2. Du Bois, P.H., *Multivariate correlation analysis*, New York: Harper, 1957.
3. Sharma, A., Naval Psychological Research Unit, Cochin. *Classification Series I—Manual for TAS Classification Battery.*
4. Sharma, A., *Def. Sci. J.*, **10**, 100, 1960.
5. Thomson, G. H., *The factorial analyses of Human Ability*, University of London Press, London, Edition—1951, pp. 205—209.