USE OF WRONG SCORE TO INCREASE THE PREDICTIVE VALIDITY OF APTITUDE TESTS

S. CHATTERJI, MANJULA MUKERJEE & S. N. CHAKRABORTY
Indian Statistical Institute, Calcutta

ABSTRACT: In an objective type of test, we generally consider the right scores and the number of omissions. We ignore the wrong scores. So naturally the question arises regarding the use of these wrong scores.

In this study, the authors have studied on the basis of empirical data, the possibility of using the wrong scores to improve the predictive validity of objective type of test.

However, in this study, with the specific criterion used, the predictive ability of the tests could not be increased by considering both right and wrong scores. But it does not rule out the possibility of profitably using the wrong scores to predict some other criterion.

Introduction: An objective type of test yields several independent scores viz., the right score, wrong score and omission score. These scores are not surely independent of one another but are not so related that one can be safely substituted by others. The most common use of the wrong score or the omission score is in the case of correction for guessing, when a fraction (depending upon the number of alternatives only) of the wrong score or the omission score is either subtracted from or added to the right score respectively. In correction for guessing either the wrong score or the omission score is used because these two methods yield equivalent results. The weight assigned to the wrong/omission score is independent of statistics like reliability, validity and difficulty level of the test. So it is a trite question whether such a composite obtained from right and the wrong/omission score is the best way of combining the two scores or whether they can be utilised in a different way. Can the right and the wrong/omission scores be treated as if they are obtained from two different tests and in predicting some external criterion can these scores be combined on the basis of suitable regression weights for increasing the test validity?

In this study, however, efforts have been made to study on the basis of empirical data the possibility of using the wrong score as separate measurement and how far prediction of a criterion can be improved from the use of this score along with the right score.

Some studies have been made with the wrong scores by several investigators. Frutcher and Thurstone studied the nature of the wrong scores. Frutcher by factorially analysing the right and the wrong scores observed that "the two may measure different common factors as well as the same factor to different extent" (2). Researches in the Army Air Forces (2) also considered the possibility of utilising the wrong scores and it was observed that the wrong scores in Clerical Aptitude Tests measured carefulness whereas the right scores which was supposed to indicate carefulness failed to do

so. Thurstone (2) considered the situation when the aim was to predict certain criterion. He calculated the optimal weight for the wrong score when a weight of + 1 was assigned to the right score. It was found that weight assigned to the wrong score was usually negative but it might become positive when speed was the important parameter in the test.

Sample : The sample utilised in this investigation consisted of three groups of students of Business Management in an institution in the years 1965, 1966 and 1967. Before being admitted to the course, these students had to qualify themselves through an admission test. The purpose of this admission test was to predict the criterion i.e., the cumulative grade point average (CGPA) obtained by the students at the end of their course. There were 52, 74 and 65 students in first, second and third groups respectively. For these students both the admission test scores and the final CGPA were available. The admission test battery consisted of the following five tests viz.,

1. Breadth of Knowledge
2. General Ability
3. Mathematics
4. English Comprehension
5. Data Interpretation

For each test both the right and the wrong scores were available for each individual.

Analysis Of The Data

(a) Preliminary Analysis : As mentioned earlier the admission test battery was divided into several parts. At first, the means, standard deviations and intercorrelations among the scores (both right and wrong) were calculated separately for three groups of students. The obtained values are presented in Tables 1, 2 and 3. All the calculations were, however, done with the help of electronic computer HONEY-WELL 400

From Tables 1, 2 and 3 it follows that though the means of the wrong scores were much less than the corresponding means of the right scores yet the corresponding standard deviations were almost of equal magnitude. This shows that the groups were more homogenous with respect to the factor or factors measured by the right score than with respect to the factor or factors measured by the wrong scores. Scrutinising the nature of the correlations (ignoring signs) among right scores in different tests it can be said that these were not much different from those observed among the right and wrong scores either in the same test or in separate tests. For example from Table 1. it follows that the correlation between the right scores in test 1 and test 2 was as high as .54 whereas the correlation between the right and the wrong scores in test 1 and test 2 were —.61 and —.45 respectively which indicated that the magnitudes of the correlations were not very much different. Again for test 4 and test 5 these correlations between the right and the wrong scores were insignificant. Similar results were also obtained with respect to the second and third groups.

Considering the correlations among the right and the wrong scores in different tests the situation appears to be encouraging because most of the correlations were insignificant.

High degree negative correlation indicates that the wrong score is more or less dependent (though in opposite way) upon the right score ; insignificant correlations indicate that the wrong score is independent

TABLE 1 Showing the means, standard deviations and intercorrelations among different scores for 1965 groups (N=52)

Scores	Right scores in different tests					Wrong scores in different tests					Criterion
	1	R_2	R_3	R_4	R_5	W_1	W_2	W_3	W_4	W_5	Y
R_1	—	.54**	—.04	.23	.12	—.61**	—.35*	.13	.06	—.15	—.14
R_2		—	.27	.28*	.23	—.40**	—.45**	.06	—.01	—.05	.06
R_3			—	.45**	.18	—.05	—.09	—.01	.00	.05	.38**
R_4				—	.17	—.07	—.08	.09	.19	—.15	.35**
R_5					—	—.13	—.36**	—.18	—.11	—.37**	.34**
W_1						—	.64**	—.08	.16	.19	.09
W_2							—	.21	.32*	.50**	.00
W_3								—	.21	.39**	—.08
W_4									—	.33*	—.06
W_5										—	—.08
Y											—
Mean	22.86	37.40	17.15	13.31	12.90	10.85	21.73	3.44	6.83	12.67	5.78
Standard deviation	3.76	7.72	4.76	3.93	3.14	3.78	6.37	2.11	2.31	3.78	1.09

R_i = right score in the ith test. i=1, 2, 3, 4, 5.
W_i = wrong score in the ith test. i=1, 2, 3, 4, 5.
Y = criterion score (CGPA)

TABLE 2 Showing the means, standard deviations and intercorrelations among different scores for 1966 groups (N = 74)

Scores	Right scores in different tests					Wrong scores in different tests					Criterion
	R_1	R_2	R_3	R_4	R_5	W_1	W_2	W_3	W_4	W_5	Y
R_1	—	.10	.09	.14	.16	—.75**	—.20	—.17	—.24*	—.14	—.32*
R_2		—	.67**	.31**	.15	—.03	—.76**	—.46**	.07	—.12	—.01
R_3			—	.43**	.27*	.00	—.50**	—.68**	—.03	—.15	.08
R_4				—	.32**	—.09	—.18	—.18	.23	.08	.05
R_5					—	—.12	.05	—.02	.00	—.53**	.27*
W_1						—	.25*	.15	.39**	.29*	—.29*
W_2							—	.63**	.19	.15	.07
W_3								—	.19	.14	.02
W_4									—	.43**	—.08
W_5										—	—.16
Y											—
Mean	20.97	26.96	39.70	32.68	20.49	8.28	9.41	12.69	8.38	6.88	5.56
Standard deviation	4.39	5.30	8.24	6.49	3.09	3.41	4.23	5.27	4.18	2.88	0.50

R_i = right score in the ith test. $i = 1, 2, 3, 4, 5$.
W_i = wrong score in the ith test. $i = 1, 2, 3, 4, 5$.
Y = criterion score (CGPA)

TABLE 3 Showing the means, standard deviations and intercorrelations among different scores for 1967 group (N=65)

Scores	Right scores in different tests					Wrong scores in different tests					Criterion
	R_1	R_2	R_3	4	R_5	W_1	W_2	W_3	W_4	W_5	Y
R_1	—	—.49**	—.43**	.49**	.47**	—.74**	.34**	.38**	—.47**	—.20	.37**
R_2		—	.60**	.49**	.40**	—.26*	—.58**	—.37**	—.19	—.05	.26*
R_3			—	.50**	.42**	—.25	—.29*	—.68**	—.23	—.04	.21
R_4				—	.59**	—.20	—.05	—.13	—.18	.12	.29*
R_5					—	—.08	.06	—.16	—.07	—.25*	.21
W_1						—	.42**	.41**	.54**	.33**	—.30*
W_2							—	.55**	.39**	.22	—.18
W_3								—	.48**	.41**	—.13
W_4									—	.41**	—.09
W_5										—	—.01
Y											—
Mean	20.46	25.42	33.60	31.38	17.63	9.43	11.25	13.95	9.92	8.75	5.74
Standard deviation	5.32	4.78	9.59	8.87	4.13	4.02	4.24	7.00	5.38	3.67	1.03

R_i = right score in the ith test. i = 1, 2, 3, 4, 5.
W_i = wrong score in the ith test. i = 1, 2, 3, 4, 5.
Y = criterion score (CGPA)

of the right score. It may be due to the fact that either all the wrong answers are due to random responses and hence the wrong score is uncorrelated with the right score or the wrong score is measuring some other independent dimension. Not very high but significant negative correlation on the other hand, indicates that the wrong score is not measuring something which is already measured by the right score but it is measuring something in the opposite direction and also it indicates that the wrong answers are not due to random responses and hence are not uncorrelated with the right scores.

Considering the magnitude of the validity coefficients obtained with the wrong score, the results seemed to be not much encouraging because most of the correlations were insignificant which indicated that there was no common factor present between the criterion measured and the wrong score. Only the wrong scores in Mathematics had significant negative correlations with the criterion in two successive years which indicated that those who had high wrong scores in Mathematics were likely to get low CGPA.

Investigating the nature of the wrong score in the three groups it can be concluded that though the wrong scores were measuring something different from those factors measured by the right score yet these did not measure any factor the presence or absence of which could influence the relation of the test scores with the CGPA.

Next step was to obtain the regression equation using both the right scores and the wrong scores and to see whether any increment in multiple correlation could be brought about by considering the wrong score along with the right score to predict the CGPA.

(b) Multiple Regression analysis by utilising the wrong scores along with the right scores : The regression equation with the test scores as independent variables and the criterion score (CGPA) as dependent variable was obtained along with the corresponding multiple correlation. These were obtained separately for the three different groups of students as before. Moreover, these regression equations were first obtained by utilising the right scores only ; then these were obtained just on the basis of the wrong scores and finally considering both the right and the wrong scores. Hence for each group, there were three different regression equations and the corresponding multiple correlations. These regression coefficients and the multiple correlations are presented in tables 4 and 5.

From the results presented in tables 4 and 5 it follows that the multiple correlations obtained with the right scores only as independent variables are all significant but the other multiple correlations obtained by utilising the wrong scores separately and along with the right scores are all insignificant. This proves that though the wrong scores might be measuring some factors which were not measured by the right scores yet such factors were not present in the criterion. Almost similar results were obtained with the three groups of data, though the regression coefficients of the tests varied from group to group.

Conclusion : By analysing the wrong scores obtained in different tests included in a battery of selection tests used for screening three groups of students in three successive years in a management training course, it was observed that

(i) The wrong scores in different tests measured some factors which were neither completely

TABLE 4 Showing the regression coefficients for three regressione quations and the corrseponding multiple correlations (1965 group)

Variables	Reg. coefficient with right scores as independent variables.	Reg. coefficient with wrong scores as independent variables.	Reg. coefficient with right&wrong scoresas independent variables
Breadh of Knowledge (R_1)	-.0621	—	-.0555
English Comprehension (R_2)	-.0025	—	-.0025
Quantitative Reasoning (R_3)	.0481	—	.0449
Verbal Reasoning (R_4)	.0715	—	.0785
Graph & Table Reading (R_5)	.0996		.1068
Breadth of Knowledge (W_1)		.0404	-.0067
English Comprehension (W_2)		-.0079	.0229
Quantitative Reasoning (W_3)		-.0136	-.0238
Verbal Reasoning (W_4)		-.0215	-.0516
Graph & Table Reading (W_5)		-.0190	.0062
Constant of the regression equ.	4.2080	5.9433	3.6897
Multiple Corre.	.55**	.16	.56
No. of cases	52	52	52

** indicates significance at the 1 percent level

dependent nor completely independent of the right scores. The magnitude of the correlations between the right and the wrong scores of the same tests was not much different from that between the right scores on different tests.

(ii) the criterion, however, was not much related with the wrong scores. Except in two cases, all the other correlations were insignificant. This proves that the wrong scores of the tests included in the selection battery did not measure any factor which was present in the criterion in question.

TABLE 5 Showing the regression coefficients for three regression equations and the corresponding multiple correlations (1966 and 1967 groups)

Variables	Reg. coefficient with right scores as independent variables		Reg. coefficient with wrong scores as independent variables		Reg. coefficient with right&wrong scores as independent variables	
	66 Gr.	67 Gr.	66 Gr.	67 Gr.	66 Gr.	67 Gr.
Mathematics (R_1)	0.0342	.0571	—	—	.0225	.0364
Breadth of Knowledge (R_2)	-.0118	.0151	—	—	-.0007	-.0010
English Comprehension (R_3)	.0067	-.0021	—	—	.0138	.0005
General Ability (R_4)	-.0053	.0165	—	—	-.0075	.0191
Graph & Table Reading (R_5)	.0373	-.0079	—	—	.0299	.0093
Mathematics (W_1)	—	—	-.0473	-.0883	-.0293	-.0471
Breadth of Knowledge (W_2)	—	—	-.0211	-.0176	.0192	-.0261
English Comprehension (W_3)	—	—	-.0033	-.0050	.0112	.0033
General Ability (W_4)	—	—	.0081	.0192	.0015	.0278
Graph & Table Reading (W_5)	—	—	-.0215	.0245	.0022	.0081
Constant of the regression equ.	4.3024	3.8377	5.8716	6.4234	4.0790	4.5195
Multiple Corre.	.41*	.40*	.35	.33	.45	.43
No. of cases	74	65	74	65	74	65

* indicates significance at the 5 percent level

(iii) the predictive ability of the battery of tests could not be increased by considering the wrong scores in the tests along with the corresponding right scores.

It should be remembered, however, that the factors which were measured by the wrong scores though proved to be not so useful in predicting the criterion in question, might be useful in predicting some other criterion. Thepresent investigation, however didnot try to throw any light to this point i.e.,

it did not try to find out the other scores could be utilised with respect possible ways in which the wrong to other criterion.

References

1. Fisher R. A. *Statistical Methods for research workers*, Oliver and Boyd. Edinburgh, Tweedale Court. London 1954.
2. Guilford J. P. *Psychometric Methods*, McGraw Hill