DOCUMENT RESUME

ED 074 127 TM 002 490

AUTHOR	Fincher, Cameron
TITLE	The Advantageous Uses of Part-Whole Correlations for the Reduction of Standardized Test Batteries.
PUB DATE	26 Feb 73
NOTE	12p.; Paper presented at annual meeting of the Natl. Council on Measurement in Education, AERA (New Orleans, La., Feb. 25-March 1, 1973)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Achievement Tests; *College Entrance Examinations; Computer Programs; Educational Research; *Predictive Validity; Speeches; *Standardized Tests; Testing; *Test Reliability
IDENTIFIERS	GAT; *General Achievement Tests

ABSTRACT

To find an acceptable way of reducing testing time without altering the administrative use of admissions tests, a study was conducted to test the prediction that scores on subtests of the General Achievement Tests (GAT) in Social Studies, Natural Sciences, and Mathematics could be used to predict the total test score. All answer sheets for 1000 subjects (250 male and 250 female) who had previously taken the tests were rescored for part scores on the two sections of each test. Pearson product-moment coefficients of correlation were then computed by sex for the two subtests and the total score on each of the three GAT tests. Regression equations were then derived from the correlations and used for the prediction of total scores made by subjects in the cross-validation sample. Correlations were then run between the predicted total scores and obtained scores previously recorded. Results showed that none of the original part-whole correlations exceeded +.95. The study findings resulted in discontinuance of the complete GAT tests at a four-year urban college. Each college applicant took only the 15-minute subtests on the three tests, and a total score in scale-form was predicted thereby for use in the admissions process. This reduced the testing time for the GAT from 120 to 45 minutes. The conversion of subscores into predicted total scores was part of the computer-scoring operations for the admissions test battery. It is concluded that the disadvantages of reduced reliability were offset by the advantages of reduced testing time. (DB)

The Advantageous Uses of Part-Whole Correlations for the Reduction of Standardized Test Batteries

Cameron Fincher
Institute of Higher Education
University of Georgia

Paper presented at the annual meeting of
National Council on Measurement in Education/American Educational Research Association
in New Orleans, February 26, 1973.

The Advantageous Uses of Part-whole Correlations for the Reduction of Standardized Test Batteries

Cameron Fincher
Institute of Higher Education
University of Georgia

The purpose of this paper is two-fold. It seeks first to report a study involving part-whole correlations for the reduction of a test battery and second, to discuss some of the reasons why the study was conducted. In the latter sense, the paper should lend further credence to the belief that too many decisions in psychological and educational testing are made for reasons which are exterior to the professional and technical aspects of testing.

The purpose of the study itself was to capitalize on the well-known spurious effects of part-whole correlations for the reduction of a lengthy test battery used in admissions testing. The problem was imposed by conditions that prohibited the elimination of any tests from the battery and by the necessity of finding some practical way of reducing the total amount of testing time. In brief, the problem was simply one of finding an acceptable way of reducing testing time without altering the administrative use of specific tests involved in the overall admissions process.

Reasons for the Study

The necessity of the study arose from administrative-legal-political decisions that had little to do with testing principles and even less to do with education. The study is, to no small extent, historical in its content but not in its implications. At

the time it was conducted in 1963, it was not possible, or advisable, to publicize the findings or their implications.

The institution in which the study was conducted was a public, four-year urban college with degree programs in arts, business, and science. The college had just recently undergone a major shift in admissions policies, moving rapidly from an open-door institution to one with unusually stringent admission requirements.* At the time of the study, it was under an injunction from the Federal courts not to discriminate against minority group members.

The battery of tests used for admission purposes was originally chosen for its "shot-gun" validity. It was expected that as the separate tests were analyzed, they would be eliminated if they did not contribute significantly to the prediction of academic grades and to the admissions decision itself. Much to the chagrin of those in charge of the testing program, it was discovered later that the use of specific tests as an admissions battery had created a legal precedence. This implied, in turn, that tests should not be removed from the battery until such time as it was legally permissible.

The tests used in the battery included the: (1) Otis Quick-Scoring Test of Mental Ability: Gamma Form, (2) Cooperative General Achievement Tests, (3) Cooperative English Expression Test, and (4) Nelson-Denny Reading Test: Revised Edition. In addition to this four-hour battery of tests, beginning freshmen applicants were

*Those desiring to hear this particular story should read Thomas F. McDonald, An Investigation of the Effects of a Rapid Transition from "Open-door" to "Selective" Admissions. Unpublished doctoral dissertation, Michigan State University, 1966 and Cameron Fincher, "Changes in Institutional Characteristics as a Function of Selective Admissions" in Clarence H. Bagley (Ed.) Research on Academic Input. Association of Institutional Research, 1966, pp. 177-183.

required to present scores on the CEEB-Scholastic Aptitude Test (SAT). To say then that the total amount of testing time was appreciable is to understate the case.

Rationale and Procedure

The rationale for capitalizing on the spurious effects of part-whole correlations was based on the belief that a high degree of correlation between a subtest and a total score may be spurious but not necessarily meaningless. If the correlation was high enough and proved out in a cross-validation, it could justify the substitution of the part for its larger whole.

Three of the tests in the admissions battery suggested by virtue of their composition that a subtest score might be used to predict the total score, thereby eliminating the administration of the longer subtest following the first. The tests in question were the General Achievement Tests (GAT) in Social Studies, Natural Sciences, and Mathematics. The first section of each test required 15 minutes for administration and dealt with terms, concepts, and general information; the second section had a time limit of 25 minutes and measured the subject's ability to comprehend and interpret written materials in that particular field.

The thought had occurred earlier that a correlational analysis of the GAT could throw light on the relative importance of factual subject matter tests as opposed to those purporting to measure developed abilities or critical reading skills. Educational Testing Service had at that time pushed rapidly ahead with tests of developed ability, with some critics left believing that the majority of variance

might be accounted for in terms of general reading ability. This possibility was related further to the belief that differential prediction was more likely through tests of fairly specific variance and that the GAT subtests dealing with concepts and terms might be combined in multiple regression equations to a better advantage than their total scores. Indeed, the specific prediction was made that the subtests dealing with concepts would correlate lower among themselves than the subtests dealing with the comprehension and interpretation of written materials.

To test the rationale, two samples were drawn from the pool of subjects who had previously taken the tests. Each sample consisted of 250 male applicants to the college and 250 female applicants, giving a total of 1000 subjects in the study. All answer sheets were rescored for part scores on the two sections of each test. Pearson product-moment coefficients of correlation were then computed by sex for the two subtests and the total score on each of the three GAT tests. Regression equations were then derived from the correlations and used for the prediction of total scores made by subjects in the cross-validation sample. Correlations were then run between the predicted total scores and obtained total scores previously recorded for the subjects. If the correlations between predicted and obtained scores proved sufficiently high, it could then be determined whether the part scores could be used in substitution for the total scores.

Analysis of the Data

The inter-correlation of subtest and total scores on the GAT produced a 9 by 9 correlational matrix for each sex. The correlation

coefficients, means, and standard deviations are presented for male students in Table 1 and for female students in Table 2. As will be quickly seen, the part-whole correlations in the tables are quite substantial but do not always exceed +.90.

The inter-correlations among the concepts subtests range from +.42 to +.64 and may be contrasted with the range of +.53 to +.67 for the reading or comprehension and interpretation subtests. The range for total scores is slightly higher, running from +.53 to +.78. As would be expected, the three concepts subtests correlate substantially with reading subtests but not as high as they do with their own total scores.

From the coefficients reported in Tables 1 and 2, regression equations were derived for the prediction of total scores for subjects in the cross-validation sample. The correlations between predicted scores and obtained scores are reported in Table 3, along with the means and standard deviations of the predicted and obtained scores. The coefficients for the three tests by sex range from +.912 to +.957 and may be compared favorably with the original part-whole correlations on which they are based. None of the original part-whole correlations exceed +.95 and for females on the Natural Science Test, the original part-whole correlation was as low as +.91. This suggests that for purposes of prediction the derived equations are remarkably accurate in predicting total scores from the concepts subtest, a finding further reflected in the means of predicted and obtained scores.

Results and Outcomes

The results of the study were sufficiently encouraging to discontinue the administration of the complete GAT tests. Each applicant

to the college was required to take only the 15 minute subtests on the three tests and a total score in scale-score form was predicted thereby for use in the admissions process. This reduced the testing time for the GAT from 120 minutes to 45 minutes. The other three tests in the admissions battery continued to be administered as they had been.

The decision to use a predicted total score was a silent one. The conversion of subscores into predicted total scores was built directly into the computer-scoring operations for the admissions test battery and no reference to subtests made. Scores for applicants taking the tests were reported without indicating in any way that the subject had not taken the complete GAT tests, and the practice became routine until the admissions test battery was discontinued. At no time was the practice questioned by applicants taking the test or detected by those using the test results for selection purposes. The most noticeable outcome was a compliment to the testing staff for becoming more efficient and for administering the tests more rapidly.

Implications and Discussion

It is readily conceded that the reduction of testing time can alter appreciably the reliability of tests administered separately and independently. It was believed, however, that since the subtests would continue to be administered as part of a battery, the disadvantages of reduced reliability would be offset by the practical advantages of reduced testing time. There was also the expectation that

the subtests would continue tc contribute to predictive efficiency even though their respective reliabilities were inadequate for purposes of differential diagnosis. Neither questions of predictive efficiency nor differential diagnosis were particularly acute, however. The test results were used primarily for a global assessment of the applicant's ability to successfully complete degree requirements.

More directly, the study indicates that part-whole correlations should not be dismissed simply because of their spurious effects. If part-whole correlations are consistently high, as they were in this study, they would suggest not only that there is considerable redundancy in testing effort but that there may be definite advantages in making use of those correlations. The implications, therefore, are that in test batteries not used for differential diagnosis, similar modifications might well be in order.

Less direct are certain implications that the inter-correlations among the sections of the GAT might have. The correlations between scores for the concepts and terms sections and those for the comprehension and interpretation sections are suggestive of the relationships that are found between general concepts on the one hand and general skills of critical reading on the other. The inter-correlations are low enough to suggest that comprehension and interpretation are fairly specific to the fields of social studies, natural sciences, and mathematics but high enough to indicate the high degree of redundancy when all three tests were used. Given the ease with which vocabulary and general concepts can be measured, plus the highly suggestive data

that they predict rather well performance on the comprehension and interpretation sections, it would follow that the testing of specific, factual content is not necessarily undesirable. There may well be occasions when it would be the better part of wisdom to test what is so readily available.

The most important implication, however, is that when tests are used in conjunction, their inter-relations should be studied carefully. Multiple regression techniques emphasize the importance of inter-correlations among predictor and criterion variables, but may place too heavy a burden on the criterion variables. Not only should the combination of predictor variables be based on empirical relationships but they should be subjected to more intensive logical analysis than they have usually received.

Finally, no rationalization is offered for the "secretive" way in which the predicted total scores were implemented. The way in which the matter was handled is to be neither condoned, condemned, nor recommended. Quite fortunately, the question of the "legal validity" of the predicted scores never arose. The administrative-legal-political conditions surrounding the study eventually dissolved and there was never a need to justify the action either legally, professionally, or technically. It is a striking reminder, nonetheless, that decisions concerning psychological and educational tests cannot be made on the basis of professional and technical considerations alone. The use of tests is subject not only to sociocultural constraints but to an increasing number of legal-political entanglements. Coping with these constraints and entanglements is not now as easy as it was in 1963.

TABLE 1

Correlational Matrix of General Achievement Test Scores for Male Students

	Social Studies:			Natural Science:			Mathematics:		
	Concepts	Reading	Total	Concepts	Reading	Total	Concepts	Reading	Total
Social Studies:									
Concepts		.73	.95	.55	.60	.62	.42	.48	.49
Reading			.91	.57	.67	.66	.41	.54	.51
Total				.60	.68	.68	.45	.54	.53
Natural Science:									
Concepts					.71	.95	.56	.53	.60
Reading						.89	.52	.60	.61
Total							.58	.61	.65
Mathematics									
Concepts								.64	.94
Reading									.86
Total									
MEAN	21.65	20.55	42.20	19.74	16.70	36.40	13.25	10.66	23.91
S.D.	6.77	5.01	10.99	6.37	4.25	9.85	6.74	4.35	10.09
NO.	250								

TABLE 2

Correlational Matrix of General Achievement Test Scores for Female Students

	Social Studies:			Natural Science:			Mathematics:		
	Concepts	Reading	Total	Concepts	Reading	Total	Concepts	Reading	Total
Social Studies:									
Concepts		.72	.94	.64	.65	.73	.45	.55	.53
Reading			.91	.62	.67	.72	.45	.59	.55
Total				.68	.71	.78	.48	.61	.58
Natural Science:									
Concepts					.59	.91	.47	.43	.50
Reading						.87	.41	.53	.50
Total							.50	.53	.56
Mathematics									
Concepts								.64	.95
Reading									.85
Total									
MEAN	20.61	19.82	40.44	16.85	15.49	32.30	11.16	9.39	20.85
S.D.	6.43	5.26	10.85	5.29	4.26	8.56	6.20	3.74	9.08
NO.	250								

TABLE 3

Comparison of Predicted Performance and Obtained Scores on General Achievement Tests

Test	Male Students: Predicted Scores	Male Students: Obtained Scores	Male Students: Correlation (r)	Female Students: Predicted Scores	Female Students: Obtained Scores	Female Students: Correlation (r)
Social Studies						
MEAN	42.16	42.38	.957	40.20	40.54	.912
S.D.	10.69	11.41		9.85	9.89	
Natural Science						
MEAN	37.32	37.32	.947	32.84	33.10	.924
S.D.	9.44	9.93		7.71	8.46	
Mathematics						
MEAN	24.44	24.80	.946	21.27	20.90	.920
S.D.	9.03	10.28		7.96	8.56	
NUMBER	250			250		