REPORT RESUMES

ED 012 082 CG 000 243

PREDICTING GRADES FROM BELOW CHANCE TEST SCORES.

BY- HILLS, JOHN R. GLADNEY, MARILYN B.
UNIVERSITY SYSTEM OF GEORGIA, ATLANTA

REPORT NUMBER USG-RB-3-66 PUB DATE 27 MAY 66
EDRS PRICE MF-$0.09 HC-$0.72 18P.

DESCRIPTORS- RESEARCH PROJECTS, NEGROES, *COLLEGE ENTRANCE EXAMINATIONS, COLLEGE STUDENTS, *PREDICTIVE VALIDITY, GRADE POINT AVERAGE, GULLIKSEN WILKS REGRESSION TESTS, SCHOOL AND COLLEGE ABILITY TEST, SCHOLASTIC APTITUDE TEST, ATLANTA

THIS STUDY IS AN ATTEMPT TO DETERMINE WHETHER THE USE OF BELOW-CHANCE SCORES CAN BE EXPECTED TO GIVE DIFFERENT RESULTS IN PREDICTION OF GRADES THAN THE USE OF ABOVE-CHANCE SCORES, THAT IS, WHETHER IT IS SOUND TO USE BELOW-CHANCE SCORES IN AN ACADEMIC-PREDICTION REGRESSION EQUATION. DATA WERE OBTAINED FROM THE THREE PUBLIC, PREDOMINANTLY NEGRO COLLEGES IN GEORGIA. THE STUDENTS WERE THOSE WHO ENTERED IN THE FALL QUARTER OF 1964 AND COMPLETED THE ACADEMIC YEAR. THE STUDENTS WERE DIVIDED INTO SEVERAL GROUPINGS ACCORDING TO THEIR SCHOLASTIC APTITUDE TEST (SAT) SCORES. CORRELATIONS WERE COMPUTED BETWEEN SCORES AND 1ST-YEAR GRADE AVERAGE. THESE DATA SEEM TO INDICATE THAT BELOW-CHANCE TEST SCORES ARE AS PREDICTIVE OF PRACTICAL CRITERION (COLLEGE GRADES) AS ARE ABOVE-CHANCE TEST SCORES. THE STUDY ALSO EXAMINED THE USEFULNESS OF RANGE-RESTRICTION-ADJUSTMENT PROCEDURES IN SUCH APPLICATIONS. REGRESSION TESTS FOR SEVERAL SAMPLES WITH BELOW-CHANCE SCORES WERE NOT DIFFERENT FROM THE REGRESSION LINES IN THE ABOVE-CHANCE SAMPLES. THE RANGE-RESTRICTION-ADJUSTMENT PROCEDURES GAVE ERRATIC RESULTS SUGGESTING THAT THEY SHOULD NOT BE RELIED ON WHEN VARIABILITY IS AS SEVERELY RESTRICTED AS IS THE CASE IN STUDYING BELOW-CHANCE SCORES. (AO)

Research Bulletin 3-66

PREDICTING GRADES FROM BELOW CHANCE TEST SCORES

John R. Hills

and

Marilyn B. Gladney

Office of Testing and Guidance
Board of Regents
University System of Georgia
244 Washington Street, S.W.
Atlanta, Georgia 30334

May 27, 1966

Research Bulletin 3-66

Predicting Grades from Below Chance Test Scores

by

John R. Hills
and
Marilyn B. Gladney
Board of Regents, University System of Georgia

As college-admissions pressures increase, some institutions will turn toward selectivity in admission as an alternate to expansion as a means of alleviating some of these pressures. For some colleges, selectivity will be a new procedure. At first, very few students will be turned away, and these will be of the lowest academic potential. In fact, they may very well be students who have scored so low on admissions tests that theoretically their scores could have been obtained by marking the answer sheets at random. Some students with such low scores may be admitted while others are turned away. This suggests that the choice among such students is based on below-chance scores, but scores below the chance score may be thought of as implying no measurable aptitude. In fact, some writers (Gulliksen, 1950, page 263) state that a score that is even within one or two standard deviations of a chance score should not be interpreted as signifying any knowledge of the subject matter of the examination. One might be concerned, then, that choices among these low-scoring applicants were being made on essentially random numbers, or at least numbers that are not related to their academic aptitude. This study is an attempt to determine whether the use of below-chance scores can be expected to give different results in prediction of grades than the use of above-chance test scores, i.e., whether it is sound to use below-chance scores in an academic-prediction regression equation.

Previous Studies

In 1956 Cliff studied the value of chance-level scores for predicting other test scores. She used data from the School and College Ability Tests (SCAT) at the high school and college levels, comparing the regression coefficients and the standard errors of estimate for predicting scores on an equating test from scores below and above chance on the V and M sections of SCAT. She concluded that there were no differences between the regression weights or the standard errors of estimate for the below-chance groups and the above-chance groups in three of her four comparisons. For the college-level SCAT scores, the relationships between scores of the below-chance groups and scores on the equating test were significantly different from zero, revealing that the below-chance scores were predictive at the college-level. At the high-school level, the below-chance scores produced regression weights that were not significantly different from zero. She felt that this indicated that below-chance scores might be predictive of other variables under certain circumstances. This study does not examine the prediction from below-chance scores of a criterion of practical importance.

The problem of below-chance scores in selection may be most acute in colleges with many low-scoring students such as the predominantly Negro colleges in the south. According to published data, these institutions have more than the usual numbers of students who score at the low end of the SAT score scale (Hills, et al., 1965). Some of these institutions, such as those in the University System of Georgia, do reject some applicants. In Georgia the percent rejected by individual public predominantly-Negro colleges over a recent calendar year ranged from 1% to 8% (Bush; 1964). For the Fall Quarter of 1964, the percent rejected for these institutions ranged from 3% to 14%, the latter being a higher percentage of rejections than was reported for the major state university for that same quarter (Klock; 1965).

The SAT scores in these institutions not only are, on the average, low, but they also are unusually homogeneous. The standard deviation will more often be in the 45 to 55 range than in the 95 to 105 range. (See, for example, Hills, et al. 1965). Similar narrow ranges occur on other admissions tests (Munday). This has caused some investigators to consider how high the correlations between these SAT scores and grades might have been if the spread of scores had been greater. Usually the investigators have approached the problem through applying corrections for restriction of range (Biaggio and Stanley; Munday; Stanley, Biaggio, and Porter). It appears, however, that the usual procedure for correcting for restriction in range may be inapplicable in such situations as these. The restriction in these distributions is not so much a matter of selection as a matter of such things as inadequate floor on the test or the reporting of an arbitrary low score for all raw scores at or below a given level. As Gulliksen states (1950, page 112), "As we approach the floor or the ceiling of a test, the error variance is clearly affected, but the theory presented in this chapter (Effect of Group Heterogeneity on Test Reliability) has nothing to do with such effects." Stanley, Biaggio, and Porter attempted to take into account the floor effect by improving the estimate of the standard deviation to be used for the restricted group in the adjustment for restriction in range. However, they recognize that further studies are needed to examine the validity of their assumptions.

Our study was not concerned with the problem of whether the validities of admissions tests for these low-scoring students were as high as, or higher than, the validities for higher-scoring predominantly Caucasian students. It was concerned with the question of whether for the lowest-scoring of these low-scoring students the validities were similar to the validities for the higher-scoring of the group. Specifically, if the scores are so low that when the usual correction for guessing in multiple-choice items is applied the raw scores are zero or less, is the regression of grades on scores different from the regression of grades on scores above that level?

The Data

Data were obtained from the three public predominantly-Negro colleges in Georgia. The students were those who entered in the Fall Quarter of 1964 and completed an academic year without dropping out. The distributions of College

Board SAT scores for these students appear in Table 1 For each of these students the date on which he took the SAT was obtained from the institution Dr. Robert Boldt of the Educational Testing Service kindly provided us with the chance scores for the relevant testing dates. These were the SAT scaled scores which corresponded with raw scores of zero when corrected for guessing by the usual formula (Gulliksen, page 249). For 66 of the students in Table 1 the chance score could not easily be determined since they took the SAT on other than the national testing dates. These cases were eliminated from further consideration in the study.

Table 1

SAT Score Distributions on Three Southern Predominantly-Negro Colleges

	SAT V	SAT M
600-619		1
580-599		2
560-579		1
540-559	1	
520-539	3	1
500-519		1
480-499	2	3
460-479	3	2
440-459	1	5
420-439	8	6
400-419	6	9
380-399	12	12
360-379	10	32
340-359	32	55
320-339	45	75
300-319	60	114
280-299	69	118
260-279	106	116
240-259	113	77
220-239	95	25
200-219	101	12

The students were divided into several groupings. One group consisted of those who scored at chance level or below. Another group was composed of those in each institution who scored immediately above the chance level. The size of this group in each institution was equal to the size of the below-chance group in that institution. A third group in each institution was chosen from those at the high end of the test-score distribution, but again this group was equal in size to the below-chance group. A fourth group comprised all those who scored above the chance score. Finally, all the students in each college were considered together as a single group. The chance scores on V and on M were not identical, of course. The chance score in V was roughly on the level of 225 to 230, depending on the test form. The chance score on M was roughly on the level of 275 to 280, again depending on the test form. The division into the various groups was done on SAT V, and it was also done on SAT M. The numbers of people in each of the groups appear in Table 2. The colleges are labeled A, B, and C.

Table 2

Sample Sizes

Colleges	Below Chance On V	M	Equal Just Above Chance On V	M	Equal Extreme Above Chance On V	M	Total Above Chance On V	M	Total N
A	18	44	18	44	18	44	126	100	144
B	36	61	36	61	36	61	162	137	198
C	28	61	28	61	28	61	231	198	259

Analyses

In each of the subgroups and in the total group within each college the correlations were computed between V and first-year average grade (FAG), between M and FAG, and between V and M. These raw correlations appear in Table 3.

Table 3

Raw Correlations

		For Group Below-Chance On V	M	For Group Of Equal Number Just Above Chance On V	M	For Highest-Scoring Group On V	M	For Total Group Above Chance On V	M	For Total Group V	M
A	F vs. V or M	-.28	.13	.17	.14	.36	.32	.37	.42	.40	.45
	V vs. M	.70	.02	.35	.04	.50	.32	.51	.48	.54	
B	F vs. V or M	.22	.03	.10	.08	.36	.37	.50	.31	.50	.38
	V vs. M	.36	.10	-.05	-.02	.44	.31	.36	.32	.34	
C	F vs. V or M	-.16	-.06	-.08	.18	-.09	.06	.38	.27	.41	.31
	V vs. M	.16	.06	.23	.22	.39	.36	.45	.43	.48	

Corrections for Range Restriction

Adjustments for restriction of range were applied to the correlations for each of the subgroups to estimate from them what the correlation would be in the total group. The adjustments were based on the usual formula (Gulliksen, page 137, Equation 17). The adjusted correlations appear in Table 4. The values for the total group in each college are to be compared with the adjusted values from each subgroup to evaluate the soundness of the estimates obtained from corrections for range restriction.

Table 4

Correlations Corrected for Range Restriction

		Based On Below Chance On		Based On Group Of Equal Number Just Above Chance On		Based On Highest-Scoring Group On		Based On Total Group Above Chance On		Raw Total Group	
		V	M	V	M	V	M	V	M	V	M
A	F vs. V or M	-.81	.35	.78	.52	.37	.37	.39	.46	.40	.45
	V vs. M	.98	.05	.94	.17	.52	.37	.53	.52	.54	
B	F vs. V or M	.74	.07	.43	.27	.47	.44	.53	.36	.50	.38
	V vs. M	.88	.23	-.23	-.07	.56	.37	.39	.37	.34	
C	F vs. V or M	-.68	-.20	-.52	.70	-.10	.06	.39	.28	.41	.31
	V vs. M	.68	.20	.87	.77	.43	.35	.46	.45	.48	

Certain features in Table 4 stand out rather clearly. First, the estimates of the correlations in the total groups which are derived from the below-chance groups and from the group of equal size immediately above chance are quite inaccurate. Some of them are large and negative while the total population values are substantial and positive. Second, the estimates from the groups at the upper extreme on aptitude scores are fairly accurate, with a few exceptions at one college. Third, the estimates from the entire groups who scored above chance are quite sound, differing from the values for the total group by no more than .05.

Several hypotheses can be evaluated from these data. One might have thought that the correction formula would have been inappropriate because the total score distributions deviated markedly from normality or even symmetry. In Table 1 it can be seen that both distributions have appreciable positive skew. However, if departure from normality made the correction procedure inappropriate we should not have gotten the sound estimates we obtained from the total above-chance group and the extreme above-chance group.

One might think that the problem with the adjustments from the below-chance and near-chance groups were inferior because those groups were considerably smaller than the total above-chance groups. However, good estimates were obtained from the extreme above-chance group which was of the same size as the below-chance group.

One might think that the estimates from the below-chance group were faulty due to the College Board's policy of reporting as 200 all scores that would convert to below 200 on the Board's standard score scale. That probably accounts for the increase of frequency of scores between 200 and 219 for SAT V in Table 1. However, this effect does not appear for SAT M. There may have been few or no below-200 scaled scores for SAT M. If this particular kind of floor effect were the culprit in our estimates from below-chance scores, it should not have been involved in the corrections of correlations between SAT M and FAG, and it should not have been of particular significance in the corrections from scores just above chance as opposed to corrections from scores at the extreme upper end of the distribution.

Gulliksen admonished that floor effects would influence the error variance of scores, and that the corrections would not apply if such effects existed. It seems likely that floor effects are occurring in these data since we by design are operating near the lower end of the distribution of SAT scores. However, it is not clear from Gulliksen's comments that the floor effects would operate only to the detriment of adjustments made from the lower scoring groups, as was found in our data. Table 5 contains the standard deviations of SAT V and SAT M in the below-chance, the immediately above-chance, the extreme above-chance, and the total groups. Clearly the standard deviations for the below-chance and the immediately-above-chance groups are distinctly smaller than the standard deviations for the other groups. This suggests that the problem in corrections from the below-chance group may be not so much a floor effect, which

Table 5

Standard Deviations for SAT V and SAT M

College	SAT	Below Chance	Equal Just Above Chance	Equal Extreme Above Chance	Total Group
A	V	11.1	7.2	50.8	52.9
	M	18.7	12.4	45.3	53.1
B	V	10.7	10.9	37.6	52.1
	M	20.1	13.2	38.0	46.8
C	V	10.8	8.2	55.4	62.6
	M	17.6	11.1	60.9	59.6

should have been but was not more distinct in the below-chance group than in the immediately-above-chance group, as an effect of the severity of restriction. The groups which yield poor adjustments have in common exceedingly small standard deviations which produce very large adjustments unless the raw correlations are close to zero. It is interesting that the larger standard deviations obtained in the extremely high-scoring groups are a function of the positive skew of the distributions, a factor which might have been expected severely to distort adjustments.

One other hypothesis about these data can be examined through the correlations. Presumably if below-chance scores represent random variations, then in a scatterplot which shows a moderate degree of correlation there should be a distinct curve at the lower end, with the arrays at the lower end being nearly level, but with the arrays progressively rising in the above-chance region. Scatterplots of these data appear in Tables 6 through 11. The x's in these plots indicate the cells which include the median of each vertical array. Only in the plot of SAT V vs. FAG for College C does there appear to be this kind of curvature. Generally the central tendencies of these arrays depart no more from linearity than do those in, say, Snedecor's illustration (Snedecor, 1956, page 396) in which he states that the values give an impression of linearity. Thus a lack of linearity, especially a drop of correlation to near zero level below the chance-score level, cannot readily be blamed for the poor results from adjusting correlations from below-chance level for restriction in range. It is interesting that in most of these plots there is not even a clear departure from homoscedasticity, judging from the range of values represented in each of the vertical arrays.

To summarize, then, the usual procedures for adjusting for restriction in range do not give very satisfactory results when one is attempting to determine whether below-chance test scores provide predictive information comparable to that obtained from above-chance scores. The erratic results do not seem to be clearly attributable to any of the following factors: sample size, skew, curvilinearity, floor effect, or homoscedasticity They may be brought about by the degree of severity of restriction in range which obtains with below-chance scores, but the theory for corrections for restriction does not take such a factor into account.

Regression Tests for Several Samples

Another procedure is available for studying the question of whether the regression in the below-chance group is similar to the regression in the rest of the data. This is provided by Gulliksen and Wilks' regression tests for several samples (Gulliksen & Wilks, 1950). Their procedure tests the hypothesis that sampling has been done from different portions of the same universe where the division of the universe has been with respect to a predictor variable. It assumes that the criterion distribution is normal. This would seem to be quite appropriate to our problem since we have done exactly what their procedure requires, i.e., selected from one portion of the universe (the below-chance portion), and the FAG distributions are quite normal in appearance.

The Gulliksen-Wilks tests evaluate the similarity in the different samples of standard errors of estimate, of slopes of the regression lines, and of the intercepts of the regression lines. In our case, the samples were the below-chance and the above-chance groups of subjects for V and for M in each college. The explicit selection on the predictor came at the chance score. For each of the six sets of data the three significance tests proved to be nonsignificant. The conclusion is that the below-chance groups can be considered to have the same regression lines as the above-chance groups. If it were true that the regression line changed to zero slope at the chance score, we should have found that the regression slopes were different for the below-chance and the above-chance groups, but this was not the case.

Table 6

Scatterplot of SAT V vs. FAG
College A, Males and Females
N= 165 r= .41

FAG	200	230	260	290	320	350	380	410	440	470	500
34				2	1		1				
31		1	3	2		3	1				
28	2	2	1	2	1					1	1
25	4	4	4	7	8x		1	1			
22	6	12x	6	5x	2	2x					
19	6x	8	11x	7	2	2					
16	9	5	6	3	1		1				
13	7	2	4		1						
10	1	1	1								
		1									

SAT V

Table 7
Scatterplot of SAT V vs. FAG
College B, Males and Females
N= 204 r= .51

FAG \ SAT V	200	230	260	290	320	350	380	410	440	470	500	530
37					1	1						
34			1	1		1	1	1				
31			2	1	2	2		1	1			
28	2	4	4	1		1		1	1			
25	3	2	5	5	4x	5x	1					
22	3	11	8	8x	3	4	1					
19	10	9x	17x	7	1							
16	12x	10	10	3								
13	7	6	6				1					
10	2	5	3		1							

SAT V

Table 8
Scatterplot of SAT V vs. FAG
College C, Males and Females
N= 298 r= .43

FAG \ SAT V		200	230	260	290	320	350	380	410	440	470	500	530	560
37			1											
34				1		1	2	1		1		1		
31		2	2		2	5	2	3	1		1			
28		1	5	9	7	5	2x	1x	3x	1				
25		6	10	10	10	7x	1	2			1			
22		6	14	8	6x	7	2		1				1	
19		12x	17x	23x	8	4	1	2						
16		12	15	11	7	2	1				1		1	
13		9	6	4	2	1								
10		1	2	1										
		1	2											

SAT V

Table 9

Scatterplot of SAT M vs FAG
College A, Males and Females
N = 165 r = .42

FAG \ SAT M	200	230	260	290	320	350	380	410	440	470	500	530	560
37													
34					2		1			1			
31		1	1	2	1	3		1					1
28		2	2		4			2					
25		4	5	10	4	5x	1						
22	1	3	13	4	7x	5							
19	2	6x	9x	11x	5	2	1						
16	3x	5	8	5	3	1							
13		3	5	5			1						
10		1	2										
	1												

Table 10
Scatterplot of SAT M vs. FAG
College B, Males and Females
N= 204 r= 34

FAG \ SAT M	200	230	260	290	320	350	380	410	440	470	500	530
37					1					1		
34				1	3	1						
31		2	1	2	2			1			1	
28		1	4	4	3	1		1				
25	1		8	9	4			1				
22	1	11	7	10x	7x	1	2					
19	3x	11x	9x	10	7	4x						
16	2	6	14	7	6	3						
13	1	5	9	3	2							
10	1	2	4	2	2							

SAT M

Table 11
Scatterplot of SAT M vs. FAG
College C, Males and Females
N= 298 r= .33

FAG	200	230	260	290	320	350	380	410	440	470	500	530	560	590
37			1											
34					2	2			1	1				1
31		1	3	3	3	2	2	2	1			1		
28		2	5	12	5	3	4		3x					
25	1	3	6	14	10	7x	5x		1					
22	1	5	9	15	5x	6	1	2	1					
19	1x	8x	18x	20x	14	3	3							
16	3	6	16	16	5	3								1
13		3	8	7	3									1
10		1	1	1			1							
			2	1										

SAT M

Conclusions

These data seem to indicate quite clearly that below-chance test scores are as predictive of, and are predictive in the same way of, a practical criterion (college grades) as are above-chance test scores. There would seem to be no need to be concerned about the validity of making selection decisions on the basis of these low scores if the selection instrument is generally valid and if the regression is as rectilinear as was the case in these data.

The analyses also suggest that correction for restriction in range may be very deceptive in situations such as these. It is not clear what characteristics of these situations distort the restriction-correction procedures, but it may be that those procedures should not be applied when the variability in the predictor has been restricted to a very narrow amount such as 20% to 50% of the variability of the total group as was the case in this situation.

Summary

This study attempted to determine whether below-chance scores on the College Entrance Examination Board's Scholastic Aptitude Test were as useful for selection, and useful in the same regression equation, as above-chance scores on that instrument. The study also examined the usefulness of range-restriction-adjustment procedures in applications such as this. The Gulliksen-Wilks regression tests for several samples indicated clearly that the regression lines in the samples with below-chance scores were not different from the regression lines in the above-chance samples. The range-restriction adjustment procedures gave erratic results suggesting that they should not be relied on when variability is as severely restricted as is the case in studying below-chance scores.

References

Biaggio, Angela B. & Stanley, J.C. Prediction of Freshman Grades at Southern State Colleges. Paper read at Ninth Interamerican Congress of Psychology, Miami, 1964.

Bush, Marilyn L. Rejection Rates in the University System of Georgia. Research Bulletin 64-7. Atlanta: Regents of the University System of Georgia, 1964, (mimeo).

Gulliksen, H.O. Theory of Mental Tests, New York: Wiley, 1950.

Gulliksen, H.O., & Wilks, S.S. Regression tests for several samples. Psychometrika, 1950, 15, 91-114.

Hills, J.R., Klock, J.A. & Bush, Marilyn L. Freshman Norms for the University System of Georgia, 1964-65. Atlanta: Regents of the University System of Georgia. Pages 5, 7, 12, 27, 48.

Klock, J.A. Rejected Students in the University System of Georgia: Fall Quarter, 1964. Research Bulletin 65-8. Atlanta: Regents of the University System of Georgia, 1965, (mimeo).

Munday, L. Preciting college grades in predominantly Negro colleges. Journal of Educational Measurement, 1965, 2, 157-160.

Snedecor, G.W. Statistical Methods, Fifth Edition. Ames, Iowa: Iowa State College Press, 1956.

Stanley, J.C., Biaggio, Angela B., & Porter, A.C. Relative predictability of freshman grade-point averages from SAT scores in Negro and white southern colleges. Paper read at annual convention of American Educational Research Association and National Council on Measurement in Education, Chicago, 1966.