DOCUMENT RESUME

ED 066 147 JC 720 132

AUTHOR Roberts, Frank C.
TITLE Analysis of Effects of Placement Exam Scores on Grading Practices: A Thwarted Esoteric Study.
INSTITUTION Antelope Valley Coll., Lancaster, Calif.
PUB DATE 2 May 72
NOTE 18p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *College Entrance Examinations; Comparative Analysis; *Grade Point Average; *Grading; *Junior Colleges; Placement; *Statistical Analysis

ABSTRACT

The purpose of this study was to analyze the effects of entrance examination scores, used for placement, on grading practices of teachers at Antelope Valley College (California). Correlations between students' GPA and their entrance exam scores were calculated. These were compared with correlations between grades given by 10 instructors and the students' entrance exam scores to determine whether prior knowledge of a student's score on the entrance exam had any effect on the instructor's grading practices. It was concluded that among the sample teachers studied, there was little, if any, effect. (RN)

ANALYSIS OF EFFECTS OF PLACEMENT EXAM SCORES ON GRADING PRACTICES: A THWARTED ESOTERIC STUDY

ANTELOPE VALLEY COLLEGE

Research Report

Frank C. Roberts, Research

May 2, 1972

ANALYSIS OF EFFECTS OF PLACEMENT EXAM SCORES ON GRADING PRACTICES: A THWARTED ESOTERIC STUDY

Some junior college personnel are becoming increasingly concerned with the use of "cut-off" scores from entrance exams for placement of students in college level classes. Teachers claim that better homogenity and a greater learning experience results when test scores are used in the counseling and placement process. They further claim greater success percentages in classes where careful screening has taken place. On the other hand, a recent study at a southland college (Golden West) indicates that decreased attrition and greater success resulted when students were allowed to enroll in English IA without regard to outcomes from the verbal ability portion of entrance exams. A 5.1% increase in number of students completing the course successfully is reported (significance at the .01 level of confidence is claimed for the 5.1% increase). If the claim is true, a point of divergence arises and investigation seems in order.

Another interesting aspect of using entrance exam scores is the well-known thesis relating knowledge of student records by teachers to degree of success of students. This thesis claims that if teachers are led to believe that their students are superior, the students will excel. The Golden West English IA study and a recent monograph on placement testing written by Dr. Steve Sheldon and published by ERIC caused some critical thinking regarding the use of entrance exam scores for placement purposes. Since some researchers suspect that the use of "placement" exam scores as predictors in college classwork is difficult to justify, a close examination of the placement exam situation seemed in order.

In an attempt to shed some light on the divergent opinions discussed above and because of an interest in the validity of the classical entrance/placement exam program, the following study was proposed. The main point of the study is to assess the value and effect of the placement test program as it is presently used by teachers and counselors at Antelope Valley College. Since the administration and use of placement exams has been in effect for over a decade, statistics for use in a "before and after" comparison would be age-dated and their validity would be suspect. Therefore, the "before" conditions are difficult to analyze. To make matters worse, if the thesis that knowledge of entrance exam results influence teachers grading is true, the problem becomes impossible to assess. What is more serious, is the effect that the exam program might have on a student with "poor ability" as evidenced by low scores. This type of student could be penalized two ways: 1) He would be relegated to remedial class level studies, and 2) The grades awarded, even after completing the pre-college work, could be influenced on the basis of knowledge of these original test scores by teachers. From this line of reasoning came the realization that the worst thing that could come from an improperly used placement system would be the penalty factor mentioned--a student receiving low test scores could be doubly penalized. It is known that the low achiever on entrance exams and placement tests is among the high attrition population in most instances. The above average student would be effected very little by the testing program and the gifted student, (that would naturally score high), would actually be helped to succeed by the system.

Since the most serious consequence of the program could accrue to the student that the very program was set up to assist <u>if</u> test grades were used

in any way to influence grading practices, this problem was given top priority and was designated as the item to be studied in this research paper. It was decided to Analyze the Effects of Placement Exam Scores on Grading Practices and only superficiously examine the effectiveness of the placement process.

STUDY DESIGN

At Antelope Valley College, SCAT, ACT, and SAT test scores are furnished to counselors and available upon request by teachers. In past years the grades were printed and published--all one had to do was ask for a copy of Freshman exam grades. At the present time, an instructor must ask the Dean of Student Personnel Services for such a list. Some instructors consistently request such information. Other instructors individually go to the transcript files and look up the scores made by students. (It should be pointed out that the entrance exam scores are no longer published, but are still easily obtainable.) To analyze knowledge of exam scores vs. grades awarded, a comparison type study was decided upon. The comparison of G. P. A. vs. entrance exam scores on a Pearson r basis (that is, a comparison of r to ρ) between treated and untreated samples was proposed. The scheme was as follows:

1. After enrollment of all freshman students is complete, about one-eighth of the enrollees scores will be treated as follows (N=150): Verbal ability percentile scores ranging between the 60th and 80th percentile will be lowered 40 percentile points. Scores between the 20th and 40th percentile will be elevated 40 percentile points. This phase is to be carried on "esoterically"--only the researcher and essential personnel should know of these changes.
2. The test scores will be handled normally at this point. Those instructors normally requesting scores will be encouraged to use them and if

any discrepancy on the part of a student or instructor is encountered, the incorrect score, it will be explained, could be as a result of clerical error. All "doctored" scores will be carefully recorded and changed back to the correct value the semester following the study.

3. After grades are awarded, a comparison of expected vs. actual outcome on the basis of exam grade will be made. Statistically, the correlation factor ρ (rho) of the population's G. P. A. vs. V and Q will be calculated. A similar correlation factor of GPA received vs. V and Q will be calculated for the treatment group. These values would be converted to z scores and tested for statistical significance.

4. It is hypothesized that no significant difference will exist between ρ and r when ρ is the population correlation and r is the correlation of GPA vs. V and Q of the doctored sample. The outcome from this test could provide insight on influence of knowledge of entrance exam scores on grading practice.

MURPHY'S LAW AND THE THWARTING OF A STUDY

While arrangements for conducting the esoteric study were being finalized and random procedures for selection of the 150 samples in the 40 to 80 percentile and 20 to 40 percentile bands were being devised, Murphy's law was operating. In the initial design stages of the study, the high risk of legal recourse by students claiming damage if results showed an effect on GPA were known. This problem had been considered even earlier by a group during the concept stage and the majority opinion among nine consulting researchers was that the information to be gained would be well worth the risk involved. Knowledge of the effect of entrance exams on student success was deemed to warrant risking any consequences.

Students were enrolled and the sampling was about to take place when County Council, having been consulted, announced that student records were sacred and could not be tampered with. It was even speculated that as a result of the proposal and the possibility of doing such a thing, maybe an examination of student record handling in the state be incited. Conceivably, legislation on student records and their easy accessibility by teachers and administrators could be a result.

The researcher involved realized his liability from the beginning and had decided to accept full responsibility. The district, however, had valid concern and Murphy's law took effect. An alternate plan based on unobtrusive techniques would have to be used.

ALTERNATE PLAN

As a result of the County Council ruling, an alternate plan to study the problem was devised. The new scheme made use of existing student records. It was decided to calculate the ρ's of the population (V vs. GPA and Q vs. GPA), using random sampling techniques, then calculate r's of various randomly selected instructors looking for any statistically significant discrepancy between ρ's and r's. If a difference was detected, further study of the individual instructor and his use of entrance exam scores could be undertaken. Unless a majority of instructors use scores to influence their grading practices, the result $r \approx \rho$ would enable one to infer that knowledge of placement tests plays very little part in final class grades. If $r > \rho$ in any individual case, further investigation of the degree of usage of test scores by the individual should be undertaken. Any extreme variation from the value of ρ on the part of a sample instructor would warrant further investigation. From the continued investigation, inferences as to the

effect of knowledge of test grades on grading practices of instructors could be drawn.

SAMPLING METHOD

Samples were drawn on a random basis from IBM enrollment sheets using a random number table. Each sample of 100 was selected from the sheets using the following procedure: The first two digits selected from the table were used as page indicators. The second set of two digits were used to select a sample by line on the page. If no test scores were recorded for the student thus selected, the next entry on the page would be the accepted sample. ACT, SCAT, SAT scores were all converted using appropriate conversion scales (as furnished by the individual companies) to a common percentile base. Because Antelope Valley College administers the SCAT testing program, scores are predominently SCAT scores. It is to be noted that the samples involve less than 100 individuals. After careful consideration of how to treat withdrawals, it was decided to eliminate them and not re-draw. This reduced the sample size, but the smallest sample is still over 70 in number. Table IV presents average V and Q scores of withdrawals (purged withdrawals were total-withdrawals and therefore could not be used in the bivariate sense, i.e., GPA ranges from A=4 to F=0 and there isn't an established value assigned to W's).

Samples I through III were drawn from the most recent enrollment sheets. Samples IV and V were drawn from the first IBM enrollment printout. IBM equipment was first used at AVC for extensive student data handling in 1964-65. (The 1964-65 samples were drawn to investigate a notion that students motivation to obtain high grades had changed and this might cause the present bright student to receive only fair grades while the student of a few years back was in competition for top grades.)

CACULATIONS

--- The well-known formula for Pearson product-moment correlation coefficient $r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}}$ was used with a desk top programmable computor for all calculations. Each calculation was repeated switching variables to check accuracy and allow the calculation of a_{xy}, a_{yx}, b_{xy}, and b_{yx}. The computor program used provided regression line slope and intercept constants. By changing variables both lines of regression could be examined and plotted. Figure 1 is a sample scatter plot of the data from Sample I with regression lines shown.

--- Mean r's, when needed, were obtained using Fisher's Z transformation technique.

--- Significance levels, as presented on Tables I, II, and III, are taken from Table VI of Fisher and Yates: Statistical Tables for Biological, Agricultural, and Medical Research, published by Oliver and Boyd LTD., Edinburgh.

--- Since the computor was used for all calculations, sample calculations are not shown in this paper.

INTERPRETATION OF RESULTS

The hypothesis: It is hypothesized that there is no significant difference between ρ of the population and r_i, i=1...10, of any individual teacher in the ten teacher sample. If a difference is found at the .05 level of significance, one could state that knowledge of exam scores influence teacher grading. This test will be discussed in the conclusion.

Some inferences that can be made from the data:

1. Correlations of verbal ability and quantitative ability vs. grade point averages on a population basis are, at best, only about 11% better than what might occur by chance alone (using r from II_T of .35).
2. Correlations improve as one becomes selective: If students with V > 50 are correlated against their earned GPA's, an increase of 9% is effected, that is, an r of .62 seems to result which means this correlation is 20% in excess of chance. See Figure 2.
3. Current correlations are lower than those of seven years ago. Possibly because of more effort on the part of educators to provide remedial classes in which all succeed (and succeed well). This inference can be substanciated somewhat by examination of the scatter plot of Figure 1. The group that disturbs the correlation by not performing as predicted is encircled in dashed lines. This group is, for the most part, enrolled in remedial classes and receive good grades for work that is below transfer level rigor and/or technical-vocational.

A corollary to 3: It is possible that the lack of sizable correlation is indication that entrance exam scores are successfully being used for placement purposes. With the exception of the 0 GPA scores, and a small cluster of students (7) in the lower quarter of Figure 1, most all students fall into the success region. Figure 3 shows a plot of those students ineligible to enter most college transfer classes without remedial work. Sixty-five percent of these students in the lower 40 percentile verbal ability group have a GPA $\geq$ 2.0 with 56% having a GPA > 2.5.

REPORTS FROM OTHER RESEARCHERS

--- Paul Games and George Klare in their book, Elementary Statistics, Data Analysis for the Behavioral Science: McGraw Hill, 1967 discuss theoretical reasons for a lowering of correlation as the range of talent decreases. They report scholastic aptitude vs. academic achievement correlations around .80 in K through sixth grades. "By the time the group finishes high school, however, some of the students--chiefly but not entirely those at the low end of the range of aptitude--have dropped out. With this more restricted range of talent, the correlations between scholastic aptitude and achievement scores typically drop to about .70 or so. As the group goes on to college, the range of aptitude becomes still more restricted and again correlation drops, usually to around .60." The authors go on to say that graduate school correlations are in the .50 range. (pages 369-370).

--- From this information, it seems that figures from the study are "in the ballpark" if we include only college transfer data.

CALCULATION OF ρ FROM r

$r_{I_T} = .24 \longrightarrow Z = .245$

$r_{II_T} = .23 \longrightarrow Z = .234$

$r_{III_T} = .08 \longrightarrow Z = .080$

$\rho_V = \bar{r} \approx .19 \longleftarrow \bar{Z} = .186$

$r_{I_T} = .33 \longrightarrow Z = .343$

$r_{II_T} = .20 \longrightarrow Z = .203$

$r_{III_T} = .19 \longrightarrow Z = .192$

$\rho_Q = \bar{r} \approx .24 \longleftarrow \bar{Z} = .246$

Note: r calculated for V using all samples in group I, II, and III combined is .20

Note: r calculated for Q using all samples in I, II, and III combined is .26

COMPARISON OF TEACHERS r's to ρ

A cursory examination of Table III discloses that GP vs. V and GP vs. Q need be examined for six instructors and three instructors respectively. More careful thought however, leads one to a realization that the r values calculated are unrealistic because of the many factors operating in the college placement programs. Maybe a more realistic value for r would be one calculated from Table 1 omitting the enclosed variates. An r of .56 results from this calculation. Much hypothetical testing could go on at this point with one-tail tests being most appropriate because we are interested only in r's of instructors significantly higher than ρ's. But even the use of sophisticated statistics could not hide the fact that the greatest correlation exhibited by a teacher sample is below an amount ($r=.45$) that carries with it only a 15% improvement over chance alone.

We will accept our hypothesis on the basis of the considerations mentioned above. It is concluded, therefore, that little, if any, effect is caused by knowledge of entrance exam scores on grading practices--at least among the sample teachers tested.

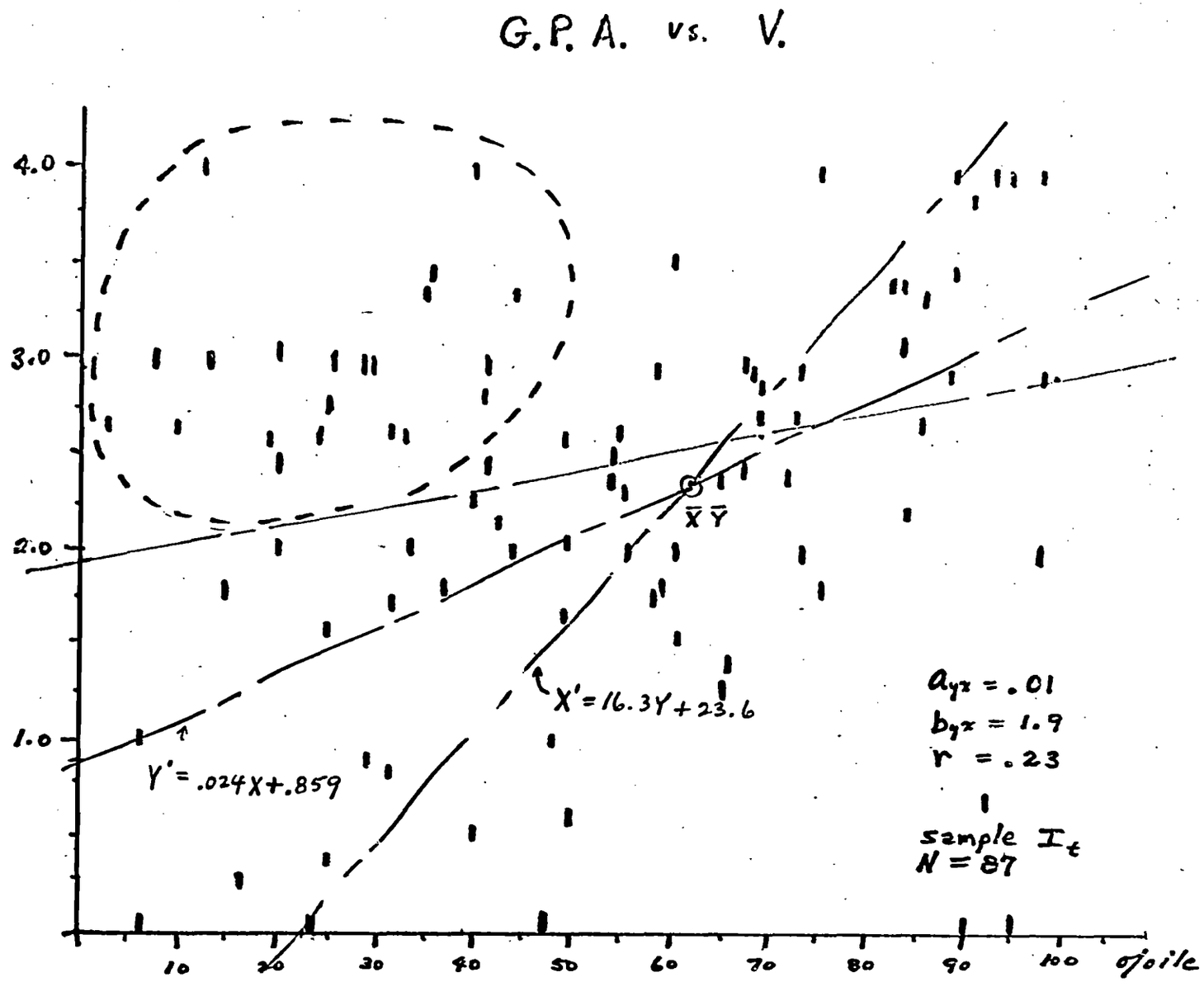

Figure 1 Scatter Plot of Data from Sample I

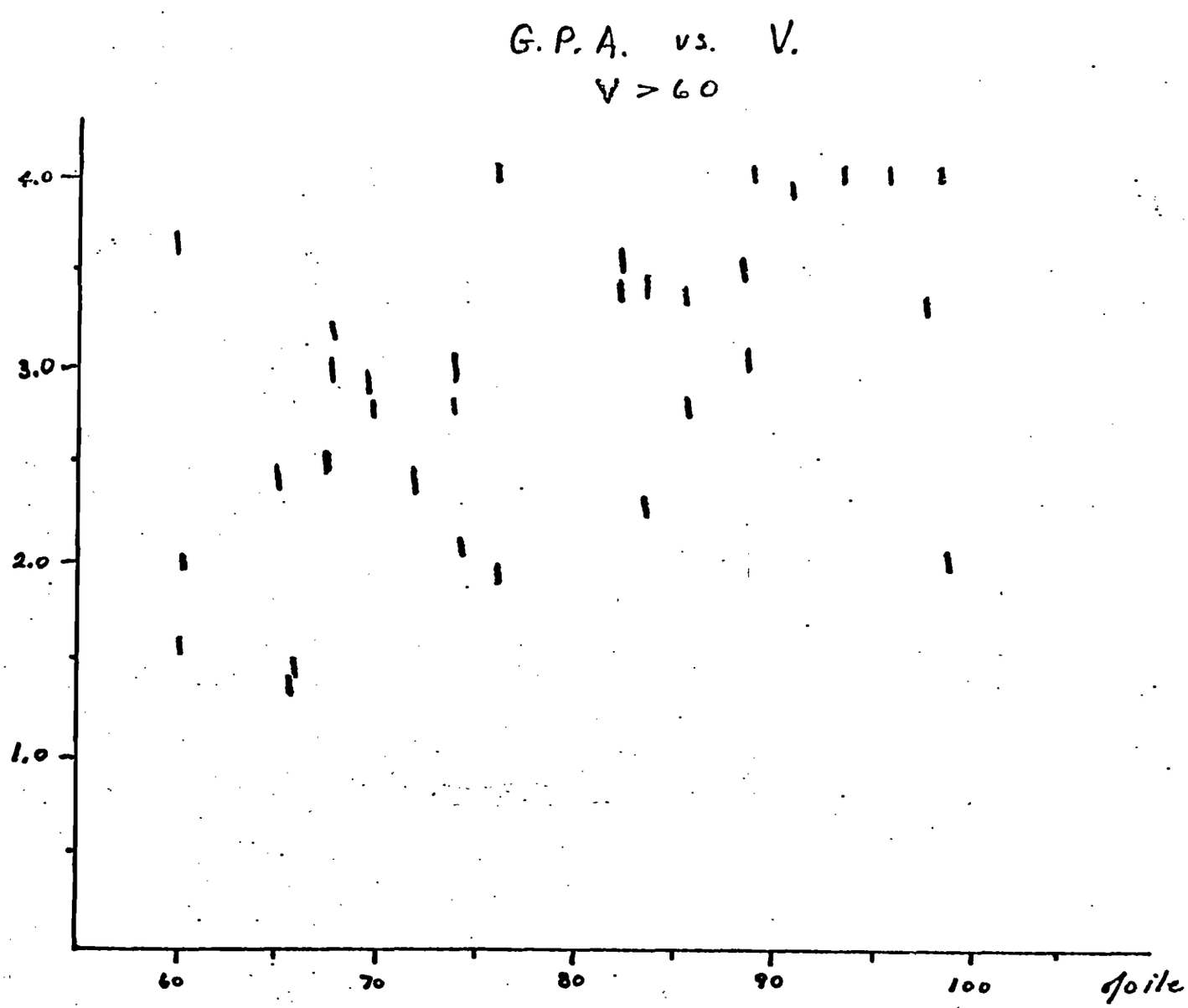

Figure 2 Scatter Plot of Sample I V's > 60

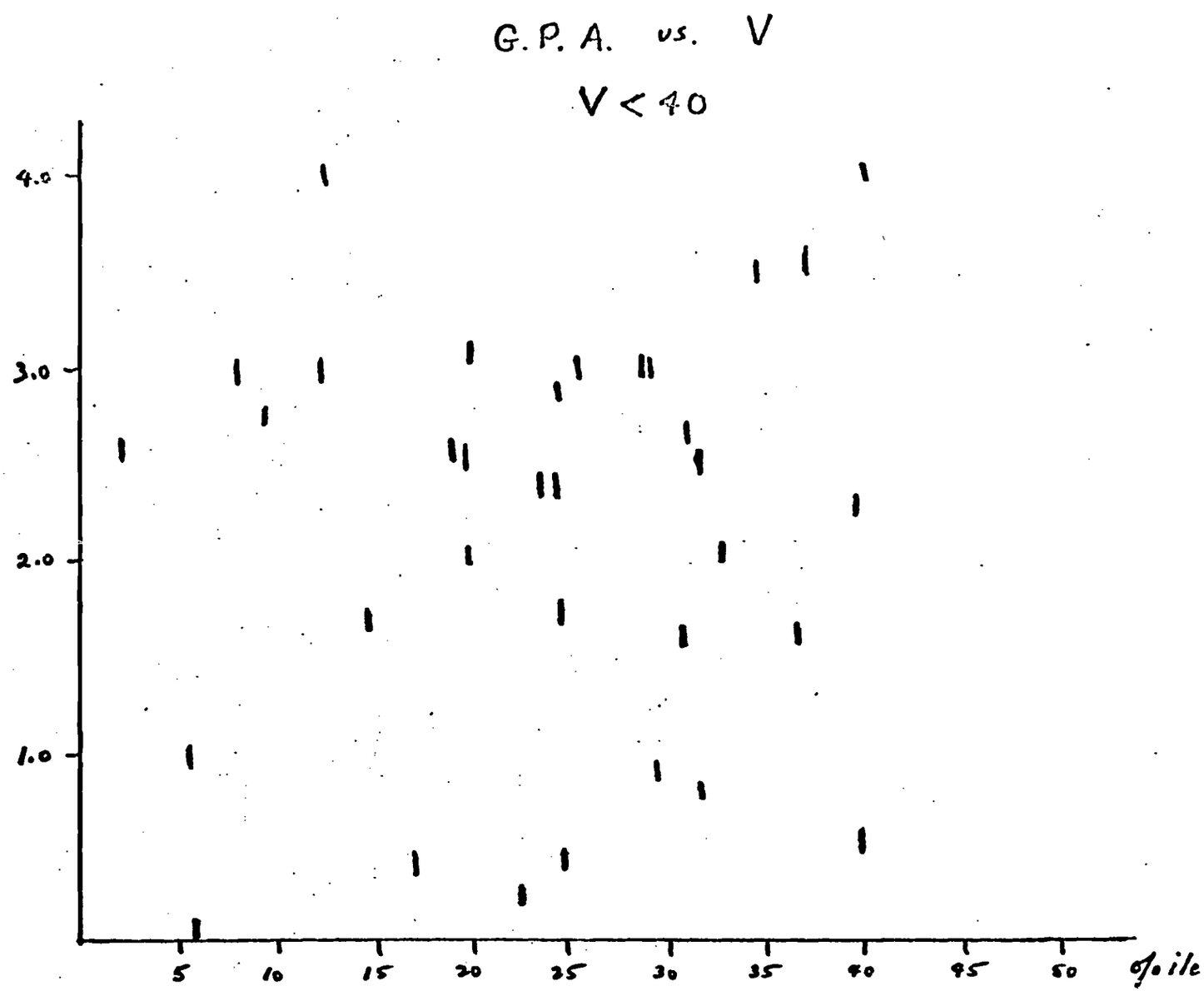

Figure 3 Scatter Plot of Sample I_TV's < 40 (Remedial & Tech.-Voc. Types)

TABLE I GENERAL CORRELATIONS

INDEX	SAMPLE	N	r GPA vs. V	Level of Sign. GPA vs. V	r GPA vs. Q	Level of Sign. GPA vs. Q	Notes
1	I_T	87	.24	.05	.33	.01	Total sample 1971-72
2	I_{FL}	28	.35	.10	.34	.10	Sample I, full load only units ≥ 12
3	II_T	80	.23	.05	.20	.10	
4	II_{FL}	18	.28	ø	.56	.02	units ≥ 12
5	III_T	74	.08	ø	.19	ø	
6	III_{FL}	23	.20	ø	.45	.05	units ≥ 12
7	$I_T + II_T + III_T$	241	.20	.05	.26	.01	
8	I_{M1}	33	.10	ø	.11	ø	40 > V ≥ 0 all included
9	I_{M2}	32 (V) 27 (Q)	.56	.001	.42	.05	V: 100 > V > 60 Three extreme cases removed Q: 100 > Q > 60 One extreme case removed
10	I_{M3}	44 (V) 33 (Q)	.62	.001	.10	ø	V ≥ 50 Three extreme cases removed Q ≥ 50 (no apparent extreme cases)
11	IV_T	94	.41	.001	.23	.05	Total first sample from 1964-65
12	V_T	89	.39	.001	.30	.02	Total second sample from 1964-65
13	$IV_T + V_T$	183	.41	.001	.27	.01	GPA vs. Q — mean using z's

See Table II for Subscript Code

Subscript code: T = total sample
FL = full load
M = modified sample

TABLE II CORRELATIONS OF SPECIAL GROUPS

INDEX	SAMPLE	N	r GPA vs. V	Level of Sign. GPA vs. V	r GPA vs. Q	Level of Sign. GPA vs. Q	Notes
1	VI_T	186	.22	.05	.12	ø	Honor students only 12 units/semester with 41 G.P./semester having GPA ≥ 3.0, no D's or F's 1971-72 sample
2	$I_{S1} + II_{S1} + III_{S1}$	89	-.03	ø	.15	ø	1.5 < GPA < 2.49 (1971-72)
3	$I_{S2} + II_{S2} + III_{S2}$	25	.03	ø	.31	ø	.5 < GPA < 1.49 (1971-72)
4	$I_{S3} + II_{S3} + III_{S3}$	223	.29	.01	.27	.01	"0" GPA's purged
5	VII_T	45	.24	ø	-.07	ø	Students with 12 units/ semester GPA ≥ 3.0 NOT ON HONOR ROLL
6	VI_{M1}	29	-.23	ø	--	--	Honor students V < 40
7	VI_{M2}	33	--	--	.14	ø	Honor students Q < 40
8	VI_{M3}	126	.24	.02	--	--	Honor students V > 60
9	VI_{M4}	124	--	--	.15	ø	Honor students Q > 60

Subscript Code: s = "strata" sample
T = total sample
FL = full load
M = modified sample
ø = significance less than .10

TABLE III CORRELATIONS BY INSTRUCTOR

INDEX	CODED TEACHER NO.	N	GP vs. V	LEVEL OF SIGN. GP vs. V	GP vs. Q	LEVEL OF SIGN. GP vs. Q	CONDITIONS
1	151110	92	.23	.05	.10	ø	Science (majors & non majors)
2	151210	91	.21	.05	--	--	Humanity
3	271210	69	--	--	.29	.02	Science
4	271810	66	.19	ø	.17	ø	Science
5	381310	105	.28	.01	.42	.001	Science
6	381410	85	.35	.01	--	--	Humanity
7	481510	107	.32	.01	--	--	Humanity
8	491410	85	.35	.01	--	--	Humanity
9	551110	102	.44	.001	.38	.001	Science (majors only)
10	591610	42	.45	.01	--	--	Remedial

A slight lowering of r has been suggested because of the analog nature of the V and Q scores vs. the digital nature of the GP scores. Plus and minus factors were not available.

TABLE IV WITHDRAWAL INFORMATION

INDEX	SAMPLE	N	$\bar{V}$	$\bar{Q}$
1	I	14	38.6	24.9
2	II	15	45.1	41.1
3	III	19	42.6	32.8
4	IV V	8	65.2	34.5

Explanation: N is the number of total withdrawals purged from the original samples. $\bar{V}$ is the arithmetic mean of the verbal ability scores of withdrawals. $\bar{Q}$ is the arithmetic mean of the quantitative ability score of withdrawals.

(Withdrawals excluded from ALL samples)