

Misinterpreting Data

Misinterpreting Data Produces Misleading Results: The Example of
Steele's (1997) "Existence Proof" for Wise Schooling.

William Collins

University of Michigan

March, 1999

Abstract

Steele's (1997) "existence proof" for his model of "wise schooling" for black college students is examined. Black students who did not participate in Steele's model were found to belong to distinct programs designed to promote academic achievement and were differently prepared for college than students in Steele's model. Difference in preparation was actually the basis used by admissions officers to select students and to assign them to a given program. Although the groups of students differed in terms of their characteristics, the groups did not differ significantly on the criterion measure of First Semester Grade Point Average (FGPA). Analysis of covariance and computation of least squares means estimates resulted in findings which fail to replicate those reported by Steele. Steele's "existence proof" for a model of wise schooling for black college students based on a theory of stereotype vulnerability is found to be lacking and susceptible to misinterpretation. A more parsimonious explanation and an alternative model for promoting academic achievement in black college students are offered.

Misinterpreting Data

Misinterpreting Data Produces Misleading Results: The Example of Steele's (1997) "Existence Proof" for Wise Schooling

Black Student academic achievement in college was the subject of an influential report by Steele (1997) in which he described an intervention designed to promote academic achievement. But an examination of his "existence proof" for the effect of his version of "wise schooling" on academic achievement finds it to be not only lacking, but easily misinterpreted as well. I must point out that I am no disinterested reader of his article; rather I am the director of the University of Michigan program which Steele referred to as remedial in his article. So this commentary runs the risk of being dismissed as the mere "sour-grapes" rantings of a disgruntled party. However, I trust that the substance itself of this commentary will prove otherwise. Moreover, the issue of black student academic progress at the college level is much too important for the nation as a whole to be cast as a disagreement between individuals or theories, and instead requires the kind of considered analysis Steele attempts to provide.

Steele has maintained that a "wise schooling" approach based on his theory of stereotype vulnerability can be effective in promoting the academic achievement of college students in comparison to other approaches. Specifically, the model he and his colleagues developed at Michigan and which is called the 21st Century Program (21CP) was compared to another program (not mentioned by name in the article), the Comprehensive Studies Program (CSP), and to a control group of students who were not subject to intervention. Steele offered a graph in his article (Figure 5 in the June 1997 American Psychologist article) which he argues provided an existence proof "that an intervention derived from the [stereotype vulnerability] theory could stop or reverse a tenacious negative trajectory in the school performance of stereotype-threatened

Misinterpreting Data

students" (Steele, 1997). I suggest that the effect itself remains to be demonstrated and that there is an alternative and more plausible explanation for the effect shown in Steele's Figure 5 than stereotype vulnerability and the intervention derived from it; that explanation is preparation for college work as indexed by standardized test score. An unfortunate feature of Steele's Figure 5 is that it provides the reader with no sense of the distribution of standardized test scores within groups. In fact, it suggests a distribution that does not exist. I provide here a similar analysis as it applies to Steele's data and in a fashion that allows the reader to understand the character of the distributions of standardized test scores for the different groups.

To begin, some consideration must be given to the local picture at the University of Michigan within which Steele's model was tested. Michigan is a large university with over 36,000 students; in reality a number of intervention strategies exist to promote student success at Michigan, but three distinct programs include minority student retention among other objectives and form the comparison groups for our analyses. The 21st Century Program is a retention program that is based on Steele's theory of stereotype vulnerability and which attempts to lessen or eliminate vulnerability among participants. The Comprehensive Studies Program (CSP) is a student retention program that emphasizes an intensive instructional and advising model; that is it stresses the development of a proper work ethic as well as academic skill building among students and provides the opportunity for more contact with teachers and advisors than is typically the case. The Summer Bridge Program (SB) is a conditional admission program that allows a select group of students to begin their university studies in the summer preceding the freshman year and to develop skills in such areas as mathematics or writing prior to fall semester enrollment. It is important to note that students selected for the Summer Bridge Program

Misinterpreting Data

typically are chosen precisely because they have relatively low standardized test scores, yet exhibit outstanding potential for college success in other ways, for example through good grades or leadership activities in high school; it also should be noted that, except for the conditional admission program, these students would not otherwise have the opportunity to enroll at Michigan. The Summer Bridge Program is a subset of the Comprehensive Studies Program and represents about ten percent of all CSP students. Students may elect to participate in any combination of the three programs described. Students are normally selected for CSP and for Summer Bridge by the admissions office. Prospective students in the 21st Century Program are identified by its staff through a separate application process for admission to a "Residential Learning Community" and includes assignment to a specific residence hall; students may also be encouraged to apply by staff via telephone.

Steele refers to CSP as a "remedial" program, which probably is not an appropriate description; rather CSP embodies a comprehensive model for facilitating academic achievement, which emphasizes intensive instruction, regularly scheduled active advising opportunities, and student development through such efforts as collaborative learning, peer advising, and freshman interest groups. It is unfortunate that the term "remedial" has developed a pejorative cachet because, whether used appropriately or inappropriately, it serves to deflect attention from any true benefits that may result from special efforts to promote student success, remedial or otherwise. Nevertheless, to the extent that the term "remedial" encompasses a focus on the development of good study habits and concern for improved academic competence, then clearly all three programs qualify. To the extent that "remedial" means correcting deficiencies, then none of the programs qualifies, although I would not quibble with one who insisted on such a label for the Summer Bridge

Misinterpreting Data

Program due to the large differences in standardized test scores its students exhibit in comparison to others in the competitive Michigan context.

Given this overall local picture, at least five groupings of black students are possible. Those who participated in the Summer Bridge Program, those who participated in CSP but not Summer Bridge, those who were in both CSP and the 21st Century Program, those who were in the 21st Century Program but not in any CSP Program, and a Control Group of black students who participated in neither CSP nor 21st Century Programs. The subjects and data for this examination and analysis were provided by Claude Steele and are the same as those used for his June 1997 article in the American Psychologist. Thus, an initial concern is that he makes no distinction between students in Summer Bridge and CSP and there should be. The assignment of subjects to groups is more appropriately indicated by the following:

- Summer Bridge (a subset of CSP; but not in 21CP) (n=101)
- CSP only (i.e., not Summer Bridge) (n=359)
- 21st Century and CSP (n=35)
- 21st Century only (n=27)
- Control Group (Blacks not in CSP or 21CP) (n=313)

This examination will be concerned with the black students in Steele's study as they clearly are the focus of his intervention model. Steele's existence proof argues that students in the 21CP perform better academically than others and that the slope of their regression line for academic achievement is steeper than for blacks in the so-called remedial program. Yet, only the graph in his Figure 5 is offered to demonstrate this effect. Other statistics that might better give the reader a fuller picture of the nature of the variables used in Steele's analysis simply are not provided. An initial point might be to look at

Misinterpreting Data

student characteristics before they entered college and then to establish whether there are differences in academic achievement before proceeding to a comparison of regression lines. In other words, for the five groups of students who comprise Steele's subjects, what is the basic structure of the variables used and how do the different groups compare?

Insert Figure 1 about here.

Figure 1 shows mean levels of academic achievement for the five groups of students. The graph shows High School Grade Point Average (HSGPA), Standardized Test Score (ACT or SAT where Test Score is standardized based on the national population of test takers), and First-Semester Grade Point Average (FGPA). One sees immediately that there are substantial differences between the groups in terms of mean standardized Standard Test Scores and in terms of HSGPA; that is, there are wide discrepancies among them in terms of preparation for college work. But one also sees that the first-semester GPAs for the different groups are rather comparable. All the groups occupy a narrow band of FGPA achievement between about 2.5 and 3.0. Observe as well that 21CP students have higher scores than the other groups on each of the achievement variables. Thus, a more detailed examination is in order for these data.

Table 1 shows the results of analyses of variance comparing 21CP students to the other groups. 21CP students achieved a significantly higher standardized Test Score than students in Summer Bridge, in CSP only, and those who were in both CSP and 21CP (in each case $p < .05$). 21CP students did not have a significantly higher Test Score in comparison to the Control Group. Similarly, 21CP students had significantly higher HSGPAs than those in Summer Bridge and those in both CSP and 21CP

Misinterpreting Data

($p < .05$), but did not have significantly higher HSGPA in comparison to students in CSP- only or those in the Control Group.

Insert Table 1 about here

Although these data suggest that the academic achievement of students in the 21st Century Program and the comprehensive program is mediated by levels of pre-college preparation, Steele emphasizes the difference in slopes of the regression lines for GPA vs. standardized test score as the really important issue and this question requires closer attention. At base, Steele asserts that stereotype vulnerability depresses the academic performance of black students and also that programs designed to address specific academic needs, such as the Summer Bridge or comprehensive program described here, can have the effect of accentuating both stereotype vulnerability and its depressive effects on achievement. As proof he offered a graph, his Figure 5, depicting first-semester grade-point average (FGPA) as a function of program and race controlling for high school GPA (HSGPA). The graph depicts a linear relationship between variables, reflecting the assumption of the ordinary least squares regression analysis; the graph also suggests a wide distribution of subjects along the entire regression line, which would mean that there were large numbers of subjects from each group at the extremes (that is, two standard deviations beyond the mean in Steele's Figure 5).

However, it should be pointed out that the University of Michigan is a highly selective institution and standardized test scores for all groups of students are higher than national averages. But when the standardization is based on the local population the well-known difference of one standard deviation in mean test score between blacks

Misinterpreting Data

and whites is apparent. Thus, all of the black groups have a mean test score that is lower than the local mean, but the 21CP group is highest on every dimension. The mean standardized test score for black students who participated in the 21st Century-only was well above the national mean (indeed no students were below it), while in contrast the mean standardized test score for participants in the Summer Bridge Program was below the national mean. Therefore, for whatever reason, the students who elected to join the 21st Century Program tended to be exceptionally well prepared before entering college in comparison to other black students in the study, while the Summer Bridge participants, in contrast, were chosen for that program precisely because they were not so well prepared. The students in the 21st Century Program were concentrated at the high end of academic preparation, while Summer Bridge students were concentrated at the lower end. Steele's analysis, illustrated by the graph in his Figure 5, obscures any group differences that may exist in the distributions of students along the dimension of standardized test score and creates an inaccurate impression of the relationship between FGPA and test score by program and race.

Insert Figure 2 about here.

Figure 2 is a scatterplot of FGPA by Test Score for all black students in the study with distinct markers for the five groups. Although this graph is densely populated with data points, it is important for the purpose of drawing attention to those points which appear at the bottom of the x-axis and which show a FGPA of "0." Michigan is a difficult school, but the admissions office does a good job of selecting students who are expected to succeed, so these data points, which number about 30, are of interest. Closer examination shows that these data points appear to be roughly equally distributed between

Misinterpreting Data

the CSP and Control groups, but none of these data points are of students in 21CP. In fact, 26 of these data points are of students in the Residential College at Michigan, a college that does not compute grade point averages for its students; institutionally this fact is recorded as a grade point average of zero in the database even if such students earned all A's. Because both 21CP and Residential College are residential programs, students enrolled in one cannot be enrolled in the other and this explains why none of these data points are associated with 21CP. The grade point averages of zero for Residential College students are meaningless and should be treated as missing rather than included in any analyses. Of the remaining four cases with a FGPA of zero, two withdrew for personal reasons and thus had not FGPA, and the other two appear to be cases in which course selection was inconsistent with the normal advice given to first-year students. However, these last four cases, though clearly outliers, are included in the analyses which follow.

Figure 3 shows a linear fitting of FGPA by Test Score for the five groups (with Residential College students omitted from the analysis). The lines for the Summer Bridge, 21CP, and Control Groups are essentially identical. The line for students in both CSP and 21CP is rather flat and the line for CSP-only students is slightly elevated. But because these groups are not equally distributed along the axis for Test Score, a linear fit may not provide the best picture of the true relationship between the variables for the five groups. Cleveland (1979) has identified locally weighted regression analysis (Lowess) as a means of aiding data visualization when underlying patterns may not be so apparent. As in our current case, the underlying structure of the data is not readily apparent from the linear fit because it obscures the distribution of the Test Score variable. Figure 4 shows the same data plotted with loess smoothing and shows not only how the data are

Misinterpreting Data

distributed along the y-axis, but also shows that for the hypothetical upper range of scores it is not the 21CP students who are at the top, rather it is the CSP and the Control Group, although the 21CP students are not far behind. Moreover, even the line for the Summer Bridge group exceeds that for 21CP at every point except for a small area where the lines for Summer Bridge and 21CP students are seen to converge. Such a finding is all the more remarkable given the huge discrepancy in preparation characteristics evidenced by Summer Bridge students. Figure 4 indicates a slightly positive relationship between Test Score and FGPA for students in each program. But more telling is the ability to visualize the locations of the test score distributions for the various groups as each line in the loess smoothing is limited to its particular range of scores. The distributions also indicate that the different groups do not start their college careers at the same point as measured by standardized test score. If standardized test score is a measure of preparation, then some groups are decidedly better prepared than others as they enter the first year of college study. Such differences in preparation undoubtedly contribute to differences in achievement. Yet, the weaker students, and those in a program which according to Steele may heighten racial awareness and consequently vulnerability, performed better at almost every point.

Figure 4 illustrates that there is no obvious evidence that the regressions within the black groups differ from one another. Thus, what we really see are groups that differ in their levels on the predictor variable, from which differences in FGPA should follow. The calculation of adjusted mean GPA estimates provides a good way of appreciating what this implies. To illustrate the point, I constructed an FGPA-test score-HSGPA regression model for black students in three groups: all those who participated in the 21CP; those who participated in CSP, but not 21CP; and those who participated in neither 21CP nor CSP. The homogeneity of

Misinterpreting Data

slopes assumption for this model was met as there were no significant interactions between the treatment (i.e., Group) and covariates. This model yielded a significant treatment effect, $F(2, 207) = 5.99; p < .01$, and adjusted mean FGPA estimates of 2.58, 2.94, and 2.47 for the 21CP, CSP, and control groups, respectively. Effectively, these are estimates of what mean first-semester GPAs for the groups would have been if each had had a common standardized test score and common HSGPA identical to the actual means across all groups. In fact, when a comparable analysis is applied to students in these different programs for each of the six years during which Steele's program has operated, the results consistently show that students in the Summer Bridge Program demonstrate the highest gain in predicted FGPA and not students in the 21CP. This is significant not only because it fails to replicate Steele's findings, but also because students in the Summer Bridge Program are the most academically at-risk, are required to participate in their program as a condition of admission, and therefore should be the most susceptible to heightened stereotype threat such as posited by Steele. Figure 5 is a graph showing mean FGPA adjusted for test score and HSGPA for the different groups in each of six years examined.

This examination of students in the 21CP in comparison to other programs suggests that the existence proof for Steele's intervention model is lacking. What the analysis actually suggests is that, although the concept of stereotype vulnerability is intellectually appealing, its impact on black student achievement in a real school context is questionable. More importantly, there appears to be a more parsimonious explanation for the differences that do exist: students who are better prepared tend to perform better academically; and programs that help students improve their preparation for academic work or which pointedly seek to develop their academic abilities lead to improved performance. A complex theory of stereotype vulnerability simply is not needed to

Misinterpreting Data

account for the differences in academic achievement that have been observed.

An alternative to a student success model based on overcoming stereotype vulnerability is the comprehensive model described here and for which the present data ironically provide an effectiveness existence proof. The comprehensive approach acknowledges the different circumstances from which students may emerge as they seek to realize their potential through higher education. Steele (1997) is almost certainly correct in his assumption that sustaining success in school requires identification with school achievement and that one must perceive good prospects for achievement in the schooling domain as well. Likewise, those who pursue higher education clearly identify with schooling. But realizing one's potential in the face of substantial disparities in preparation is a daunting task; it is rather like running a footrace but starting many meters behind the other runners. To win the race, you must first close the gap. The comprehensive model emphasizes doing so early and places a positive focus on such effort while being honest with students about what is required of them in terms of commitment to their goals. The comprehensive model includes intensive instruction, both academic and personal advising, the development of sound study habits, and active involvement in the total university community. Many programs adhering to similar models exist nationwide and they are unabashedly eclectic, welcoming --indeed, even seeking out-- effective concepts and approaches wherever they may arise. A notion like stereotype vulnerability is certainly worthy of consideration as the basis for one among many tools these programs have shown are required for meeting the challenges they face. But the true practical significance of the concept remains to be demonstrated.

References

Cleveland, W. S. (1979) "Robust locally weighted regression and smoothing scatterplots." *Journal of the American Statistical Association* 74:829-836.

Steele, C. M. (1992, April). Race and the schooling of black Americans. *The Atlantic Monthly*, 69-79.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, 52 (6), 613-629.

Misinterpreting Data

List of table captions.

Table 1. Means of selected variables; significant differences between 21CP and other groups indicated by asterisks (Tukey HSD, $p < .05$).

Table 2. Summary of regression analyses employing grades and test scores as predictors of college academic achievement for selected groups.

List of figure captions.

Figure 1. Mean achievement for selected groups.

Figure 2. Scatterplot of FGPA for select groups.

Figure 3. Regression of FGPA on standardized test score for each of five groups of black college students (Linear fitting).

Figure 4. Regression of FGPA on standardized test score for each of five groups of black college students (Loess fitting).

Figure 5. Mean FGPA for select groups adjusted for test score and HSGPA for each of four years examined.

Misinterpreting Data

Table 1. Means on academic achievement variables for selected groups.

	HSGPA	Test Score (standardized)	FGPA
SB	2.98	-.49	2.52
CSP	3.18	-.002	2.54
21CSP	3.15	.08	2.46
21CP	3.43	.64	2.80
Ctrl.	3.38	.62	2.77

Misinterpreting Data

Table 2. Years in which there were statistically significant differences between 21CP and other groups on variables of interest over a six year period.

<u>Group</u>	<u>Variable</u>		
	HSGPA	Test Score	FGPA
21CP vs.			
Summer Bridge	'91, '93, '94 '95 & '96	'91, '92, '93 '94 & '96	
CSP	'94, '95, '96	'91, '92, '93, '94 & '96	
21CSP	'91	'93 & '94	
Control			

(No significant differences were found between 21CP and other groups on FGPA for any years between 1991 and 1996; nor were there any significant differences between 21CP and the Control Group on any variable).

Misinterpreting Data

Table 3. Summary of regression analyses employing grades and test scores as predictors of college academic achievement for selected groups.

Group	B hsgpa	B test score	R	R2
SB (n=101)	.121	.145	.15	.02
CSP (n=359)	.389**	.069	.18	.03
21CSP (n=35)	.392	.043	.22	.05
21CP (n=27)	.293	.214	.38	.14
Ctrl. (n=313)	.275**	.067	.21	.04
All Black Students (n=819)	.288**	.074**	.202	.04
All White Students (n=6,705)	.384**	.049**	.16	.02

* p<.05

** p<.01

Fig. 1 - Mean Achievement for Selected Groups

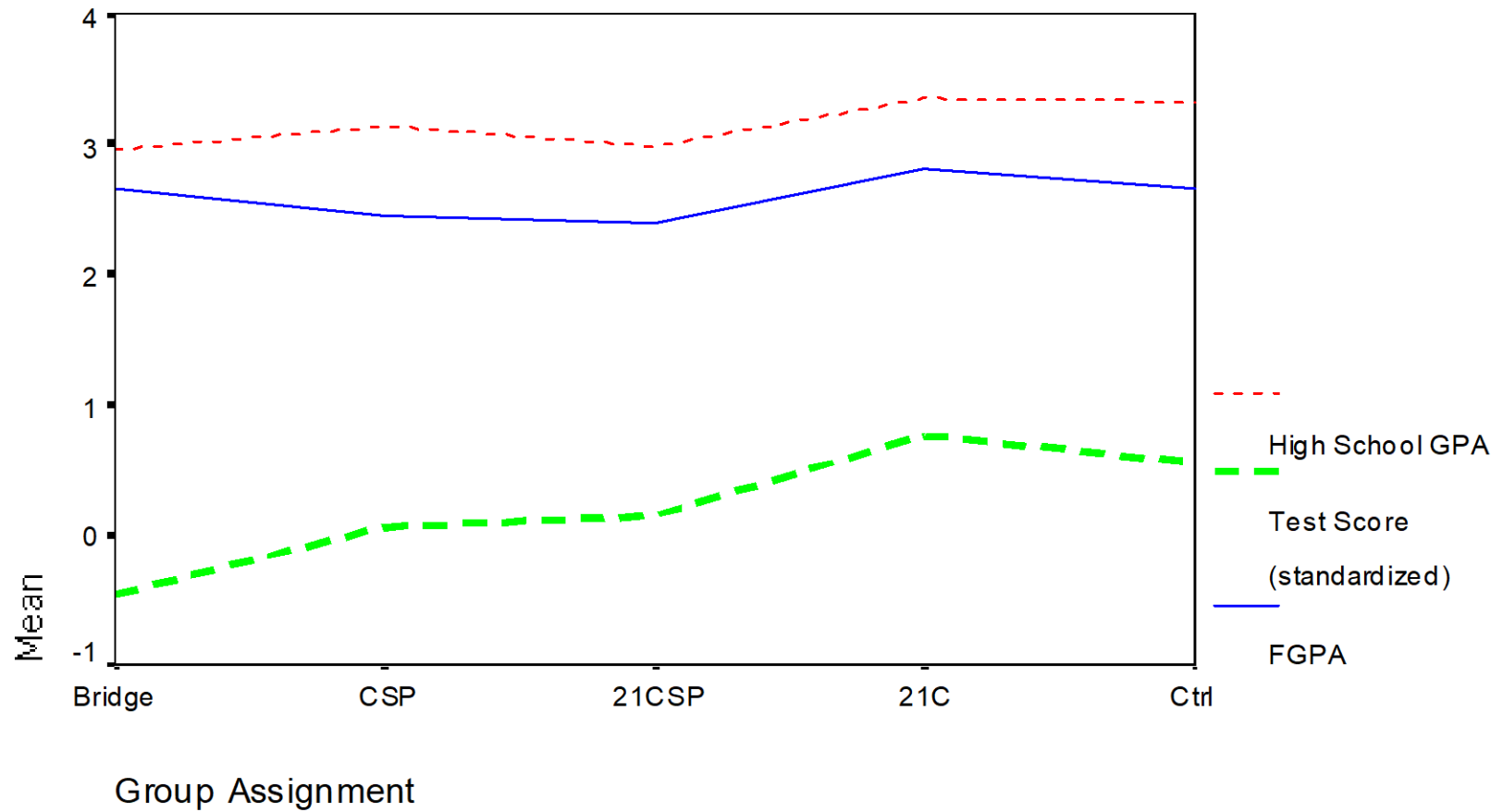


Fig. 2 - Scatterplot of FGPA by Test Score

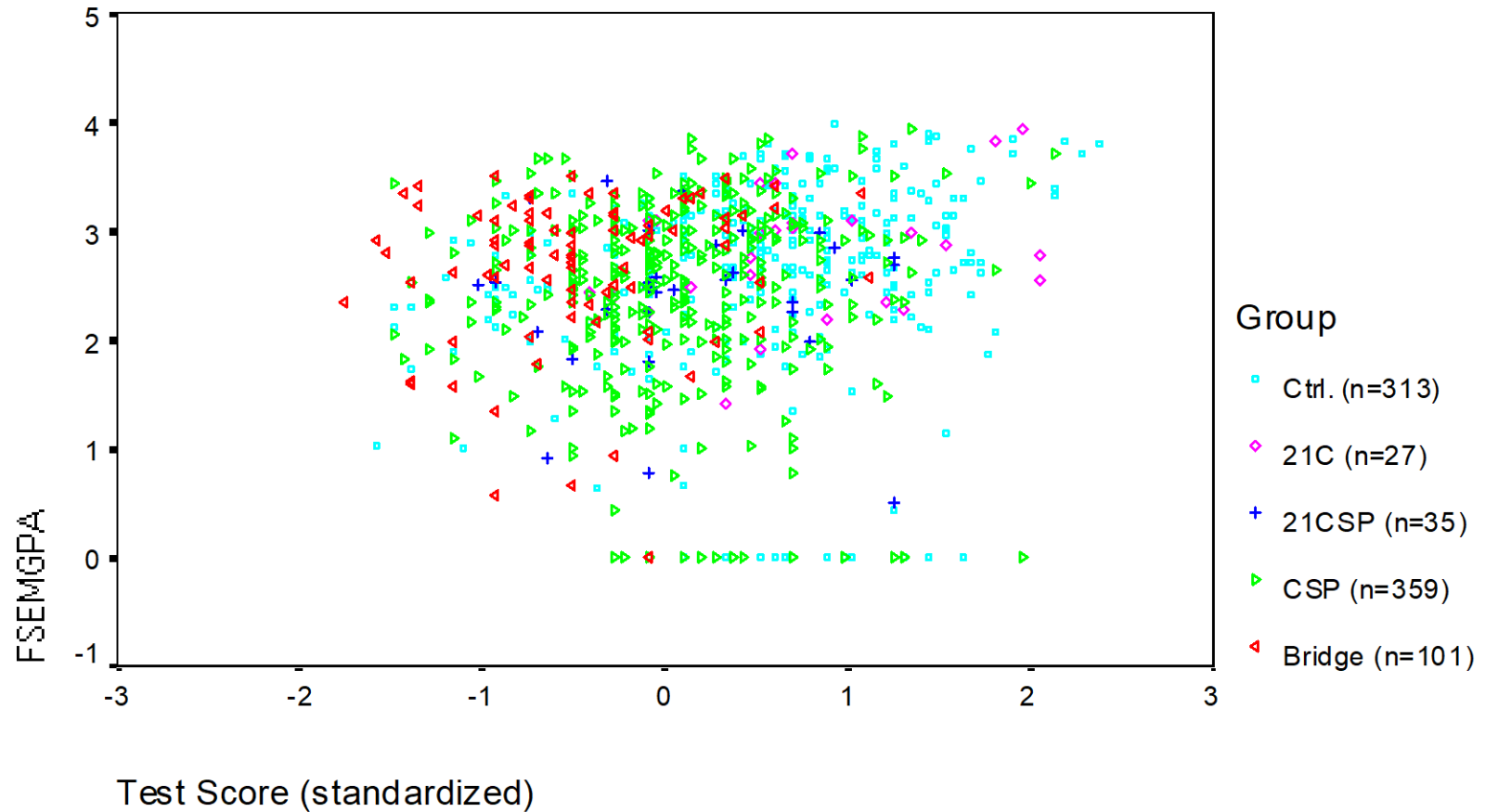


Fig. 3 - Regression of FGPA on Test Score for five groups of black college students.

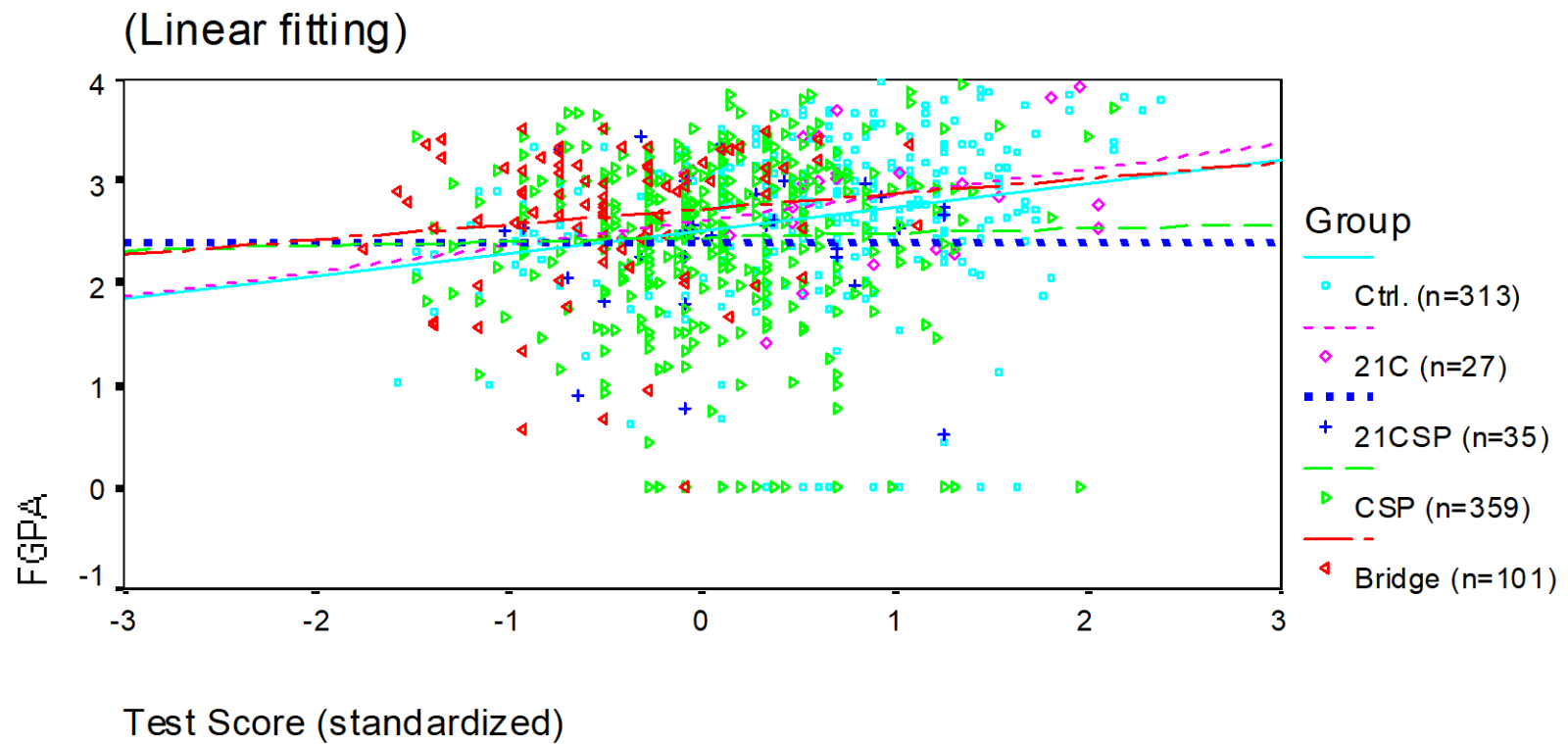


Fig. 4 - Regression of FGPA on Test Score for five groups of black college students.

