

“Wise Schooling”: An Alternative Explanation for Steele’s (1997) “Existence Proof” for the Effect of
Stereotype Threat on the Academic Achievement of Black College Students

William Collins

University of Michigan

January, 2000

Abstract

The data for Steele's (1997) "existence proof" for his model of "wise schooling" for black college students was examined and found to rely upon a methodological flaw that may not have been apparent to researchers as the July 1997 American Psychologist article was prepared. The result was that Steele's analysis created a misimpression of what the data actually mean. Correcting the flaw in Steele's original database and re-analyzing the data using Analysis of Covariance led to an alternative explanation for his findings.

“Wise Schooling”: An Alternative Explanation for Steele’s (1997) “Existence Proof” for the Effect of Stereotype Threat on the Academic Achievement of Black College Students.

A previous examination of Steele’s (1997) “existence proof” for the effect of “wise schooling” on black student academic achievement in college failed to replicate his findings (Collins, 1998). In essence, Steele’s findings relied upon a faulty analysis of the data leading to results that are easily misinterpreted and therefore misleading. The current report provides a closer look at the data and how they came to be misinterpreted. Steele has maintained that a “wise schooling” approach based on his theory of stereotype threat 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.

To more fully understand the data, 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 adjustment and success, some are residential programs such as Steele’s, while others are not. 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 threat and which attempts to lessen or eliminate threat 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 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. 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 selected for CSP and for Summer Bridge by the admissions office or by their request to be affiliated with the program. 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. Given this overall local picture, at least five groupings of black students are possible for the two years on which Steele reported:

- 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)

The data for this analysis were provided by Claude Steele and are the same as those used for his June 1997 article in the American Psychologist, but with the Summer Bridge students identified as a group. I am grateful to Steele and his colleagues for their cooperation in making their data available.

Two features of Steele's analysis deserve attention. The first is his use of 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 threat] theory could stop or reverse a tenacious negative trajectory in the school performance of stereotype-threatened students” (Steele, 1997). 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 the groups under comparison. In fact, it suggests a distribution that does not exist. That is, students from different programs tend to cluster at different points along the *x-axis* (i.e., test score) of his Figure 5 and there is little or no overlap between the groups along the *x-axis*. Steele's Figure 5 shows 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). In fact, there are not. Steele's 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.

The second feature of Steele's analysis that deserves attention is perhaps more compelling in terms of understanding just what the data suggest. The criterion variable used for his analysis is First-semester grade-point average (FGPA) as recorded in institutional records. This is probably the most reliable source one could have, but it remains important that the data in those records are valid for the purposes of statistical analysis.

 Insert Figure 1 about here.

Figure 1 is a scatterplot of FGPA by Test Score for all black students used by Steele in his analysis and with distinct markers for the different groups of students. 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 represent each group except for Steele's 21st Century Program. What do the FGPA's of "0" mean? In Steele's analysis they are treated as failures and are averaged in with other FGPA's when computing means and other subsequent analyses. Steele may not have realized that, in fact, almost all of the data points with an FGPA of "0" are of students in the Residential College at Michigan. The Residential College was created by psychologist Ted Newcomb and Don Brown in the 1967 as a "college within a college" devoted solely to undergraduate education. Residential College students are graded by written evaluation instead of letter grades and grade point averages are not computed 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 in letter graded courses. 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. Other unusual circumstances could result in the student's record reflecting an FGPA of zero as well, such as a student electing extension courses or withdrawing from the University late in the term, typically for personal reasons such as illness or trauma. Of course, it is also possible for a student to simply fail all courses, but this is actually quite rare and is not an issue for the current analysis.

Figure 2 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 different groups differ only at the extremes, but relatively few students are to be found there and even so, there is little overlap between groups. 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 3 shows the same regression lines (with data points omitted to aid visualization) using loess smoothing and shows not only how the data are distributed along the *x-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. 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. Yet, the weaker students, and those in a program which according to Steele may heighten racial awareness and consequently threat, performed better at almost every point.

Figure 3 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 FGPA

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 the five groups identified above. The homogeneity of 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, $F(6, 769) = 14.2$; $p < .001$, and adjusted FGPA means of 2.86, 2.61, 2.46, 2.66, and 2.69 for the Bridge, CSP, 21CSP, 21CP, 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 six years during which Steele's program operated, the results consistently show that students in the Summer Bridge Program demonstrate the highest gain in predicted FGPA. 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.

This examination suggests that the existence proof for Steele's intervention model is lacking due to a methodological fault. There appear to be two components of a more parsimonious explanation for the differences that do exist: 1) students who are better prepared tend to perform better academically and 2) programs that help students improve their preparation for academic work or which pointedly seek to develop their academic abilities lead to improved performance. A theory of stereotype threat is appealing in this context as an explanation of student achievement, but at this time must be viewed as suggestive only. A notion like stereotype threat is certainly worthy of consideration as the basis for one among many tools programs have shown are required for meeting the challenges students face in college. 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.

Collins, W. (1998) "Wise Schooling": An Examination of Steele's (1997) "Existence Proof" for the Effect of Stereotype Vulnerability on the Academic Achievement of Black College Students – and an Alternative. Ann Arbor, MI: University of Michigan Comprehensive Studies Program Report.

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.

List of figure captions.

Figure 1. Scatterplot of FGPA by Test Score.

Figure 2. Regression of FGPA on standardized test score for five groups of black college students (Linear fitting).

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

Figure 1. Scatterplot of FGPA by Test Score.

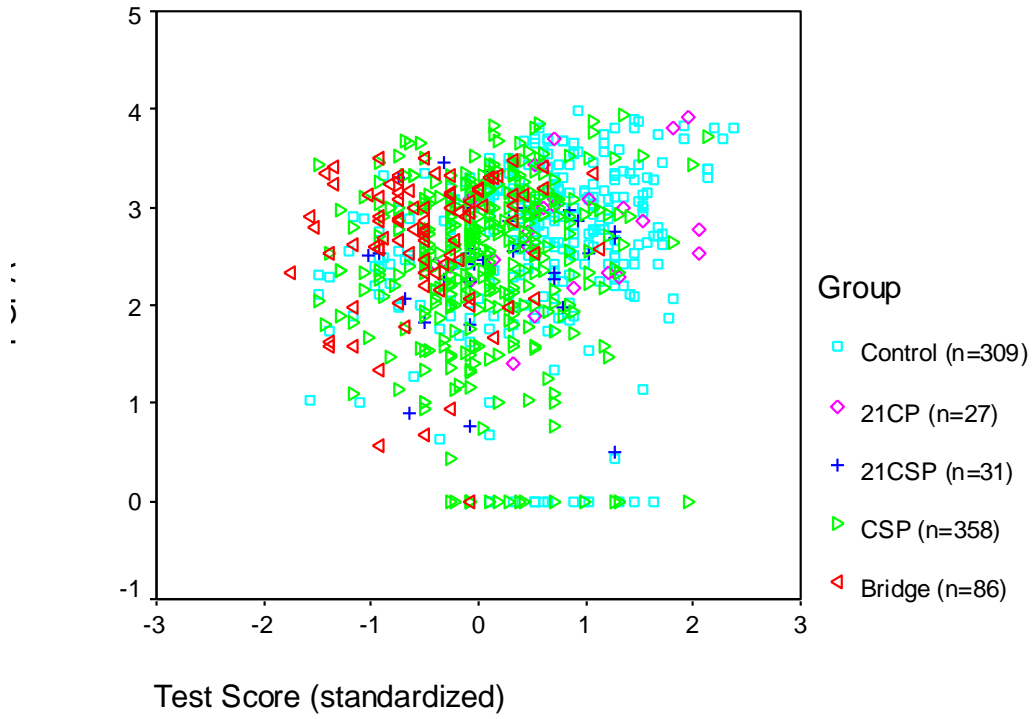


Figure 2. Regression of FGPA on Test Score
for five groups of black college students.

(Linear fitting)

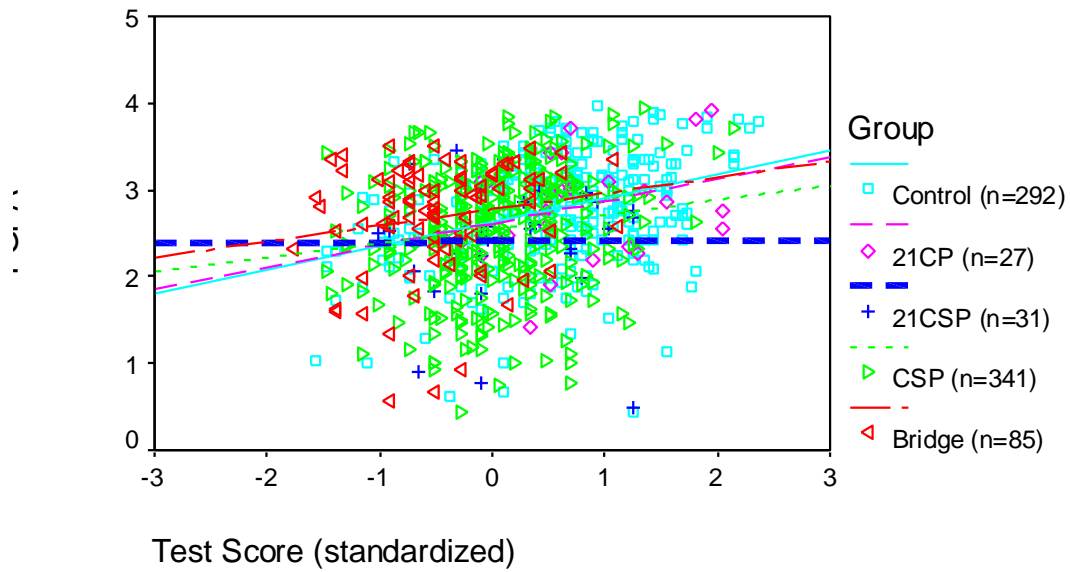


Figure. 3 Regression of FGPA on Test Score for five groups of Black college students.

(Loess fitting)

