College Board Report No. 98-2
ETS RR No. 98-9

Inquiring About Examinees' Ethnicity and Sex: Effects on Computerized Placement Tests™ Performance

LAWRENCE J. STRICKER
and WILLIAM C. WARD

College Entrance Examination Board, New York, 1998

Lawrence J. Stricker is a Principal Research Scientist and William C. Ward is a Senior Development Leader at ETS.

Additional copies of this report may be obtained from College Board Publications, Box 886, New York, New York 10101-0886. The price is $15. Please include $4 for postage and handling.

Printed in the United States of America.

Acknowledgments

Thanks are due to Central Piedmont Community College for cooperating in the study, David A. Rhoden for coordinating the data collection, Margaret L. Redman for preparing the data for analysis, Laura M. Jenkins and Xuefei Hui for doing the computer analysis, and Rick Morgan and Gita Z. Wilder for reviewing a draft of this report.

Contents

Abstract.....*1*

Introduction *1*

Method.....*2*
Sample *2*
Procedure.....*2*
Measures.....*3*
Analysis.....*3*

Results and Discussion *4*
Intercorrelations.....*4*
Analyses of Covariance.....*4*

Conclusions *4*

References.....*7*

Appendix.....*8*

Tables
1. Summary of Sample Characteristics *2*
2. Intercorrelations of Scores and Times on Tests...*4*
3. Summary of Analyses of Covariance of Scores and Times on Tests.....*5*
4. Mean Scores and Times on Tests for Ethnic Groups and Men and Women *5*

This page is intentionally blank. Please continue on to the next page.

Abstract

Laboratory experiments by Steele and Aronson (1995) found that African-American subjects' performance on difficult verbal items, described as a verbal problem-solving task, was adversely affected when they were asked about their ethnicity just before working on the items. These results were attributed to stereotype threat: asking about ethnicity primes African-American subjects' concerns about fulfilling the negative ethnic stereotype about their intellectual ability, thereby disrupting test performance. The present field experiment assessed the effects of asking community college students taking the Computerized Placement Tests™ (CPTs™), in an actual operational setting, about their ethnicity and sex. This inquiry had no statistically and practically significant effects on how well the examinees did on the tests or how long they worked on the tests.

Introduction

Two experiments by Steele and Aronson (1995), done with Stanford undergraduates, found that African-American subjects' performance on difficult GRE General Test (Briel, O'Neill, and Scheuneman, 1993) verbal items was adversely affected when they were asked about their ethnicity just before they began working on the items, while white subjects' performance was unaffected. The African-American subjects who were asked about their ethnicity answered fewer items correctly, answered correctly a smaller percentage of attempted items, attempted fewer items, and spent more time working on the items. The purpose of the experiments was described to the subjects as "nondiagnostic"—to understand the psychological factors involved in solving verbal problems; individuals' ability was not being evaluated, though they could receive feedback about their performance. Steele and Aronson explain their results as coming about because asking about ethnicity primes African-American subjects' concerns about fulfilling the negative ethnic stereotype about their intellectual ability, thereby disrupting the subjects' test performance. (See also Steele, 1997.) Other research by Spencer, Steele, and Quinn (1997) suggests that this same kind of "stereotype threat" affects the performance of women on quantitative test items, given the negative stereotype about women's ability in this sphere.

The Steele and Aronson studies have obvious parallels with the test administration procedures for the Computerized Placement Tests (College Board, 1995; Ward, 1988) and Advanced Placement (AP®) Examinations (College Board and Educational Testing Service, 1995), as well as for other standardized tests that require examinees to answer questions about their ethnicity and sex just before they take the tests. But the Steele and Aronson studies and the CPTs and AP procedures may also differ in important respects. First, the Steele and Aronson subjects were taking the items for research purposes whereas CPTs and AP examinees take the tests for important personal reasons—to guide their course placement or to get advanced credit in college—and hence may be more motivated to do well on the test material. Second, the experimental task in the Steele and Aronson studies was portrayed as innocuous problem solving whereas CPTs and AP examinees are aware that they are taking tests that reflect their mastery of important academic skills or specific course content. Steele and Aronson have also found that stereotype threat is heightened when the experimental task is described as diagnostic of the subjects' intellectual ability. Thus, inquiring about ethnicity and sex may have a limited impact on CPTs and AP Examinations insofar as stereotype threat is already elevated by examinees' perceptions of these tests as diagnostic.[1] Third, Steele and Aronson theorize that stereotype threat only affects examinees who identify with the subject matter being tested. Although AP examinees may be very involved with the academic skills being tested, CPTs examinees are probably less involved. Fourth, research by Spencer et al. suggests that an important element in the operation of stereotype threat is subjects' perceptions of the items as difficult, at the limits of their ability; it is unclear whether CPTs and AP examinees perceive the tests in this way.

A recent study (Stricker, 1998) evaluated applicability of the Steele and Aronson results to the AP Calculus AB examination (College Board, 1994), and to girls as well as African-American examinees. This test was chosen for investigation because it is taken by relatively large numbers of African-American examinees as well as girls; substantial differences exist in the test performance of African Americans and whites and of girls and boys; and the subject matter of the test is pertinent to the stereotype about females' quantitative ability as well as to the stereotype about African-Americans' ability in general. The test administration was modified for a random sample of schools by masking demographic questions on the standard answer sheet and distributing the standard answer sheet after the test to

[1]In an unpublished pilot study by Steele, inquiring about ethnicity did not affect the performance of African-American subjects when they were told that the experimental task was diagnostic (C. M. Steele, personal communication, May 21, 1997).

obtain answers to these questions. Comparisons of the examinees in these classes with examinees in a random sample of classes that received the standard answer sheet generally found no differences for African-American, female, or other subgroups of examinees on the kinds of measures of test performance used by Steele and Aronson. (Time measures could not be obtained for this group-administered, conventional paper-and-pencil test.)

Differences between the AP Calculus AB Examination and CPTs in their content and purpose, as well as in the respective test-taking populations, make the generalizability of the AP results to the CPTs uncertain. Hence the aim of the present study was to replicate the Stricker investigation of the AP Examination, assessing the effects of asking about CPTs examinees' ethnicity and sex on their scores on the tests and the time that they spent on the tests.

Method

Sample

The sample consisted of all incoming students at Central Piedmont Community College, Charlotte, North Carolina, who took the CPTs for the first time during a four-week period from August 12 to September 7, 1996. The total sample was 1,341: 333 white men, 249 African-American men, 65 other men, 391 white women, 219 African-American women, and 84 other women. The experimental group consisted of 632 subjects who took the CPTs during the two middle weeks of August 20 and August 26; the control group consisted of 709 subjects who took the tests the first week, that of August 12, or the last week, that of September 3. (One examinee in the experimental group whose ethnicity could not be ascertained and seven subjects in the control group who took the CPTs with the test administration procedures for the experimental group were excluded from the sample.) The size of the experimental and control groups for each CPTs varies because examinees did not necessarily take all of the CPTs. The sample size for each CPTs was 1,176 for Elementary Algebra, 1,238 for Arithmetic, 1,144 for Reading Comprehension, and 1,073 for Sentence Skills.

The characteristics of the total sample are summarized in Table 1. The experimental and control groups were comparable in ethnicity (52.0 percent and 55.7 percent white, and 38.1 percent and 32.0 percent African American),[2] sex (46.8 percent and 49.5 percent men), age (45.6 percent and 51.5 percent 19-years-old or under), and intended program (27.8 percent and 32.4 percent associate's degree in arts and science, and 34.3 percent and 36.0 percent associate's degree in a vocational field). Over 85 percent took the CPTs quantitative tests (88.8 percent and 86.7 percent for Elementary Algebra; 92.1 percent and 92.5 percent for Arithmetic), and 75 percent to 85 percent took the CPTs verbal tests (77.1 percent and 78.6 percent for Reading Comprehension; 77.2 percent and 82.5 percent for Sentence Skills).

TABLE 1

Summary of Sample Characteristics

Variable	*Group*			
	Experimental (N=632)		*Control (N=709)*	
	N	**%**	**N**	**%**
Ethnicity				
White	329	52.0	395	55.7
African American	241	38.1	227	32.0
Other	62	9.8	87	12.3
Sex				
Male	296	46.8	351	49.5
Female	336	53.2	358	50.5
Age				
19 or under	288	45.6	365	51.5
20 to 24	166	26.3	177	25.0
25 to 29	65	10.3	76	10.7
30 to 34	46	7.3	29	4.1
35 or more	65	10.3	58	8.2
Not ascertained	2	.3	4	.6
Intended Program				
Associate's degree in arts and science	176	27.8	230	32.4
Associate's degree in a vocational field	217	34.3	255	36.0
Undecided	33	5.2	24	3.4
Not ascertained	54	8.5	11.21	20.19
CPTs Test				
Elementary Algebra	561	88.8	615	86.7
Arithmetic	582	92.1	656	92.5
Reading Comprehension	487	77.1	557	78.6
Sentence Skills	488	77.2	585	82.5

Note: Percentages may not add up to 100.0 because of rounding error.

Procedure

Students routinely scheduled to take the CPTs at the college's testing center, before beginning their course

[2] Other ethnic groups were pooled in the study because of their small number.

work in the Fall 1996 semester, were directed to the 16 personal computers regularly used in administering the CPTs. For the experimental group (examinees tested in the weeks of August 20 and August 26), the initial computer screens containing the demographic questions were eliminated on all computers, and a paper-and-pencil questionnaire with these questions was administered after the CPTs were completed. (A copy of the questionnaire appears in the Appendix.) No other changes were made in the test administration. For the control group (examinees tested in the weeks of August 12 and September 3), all the regular test administration procedures were followed, including the presentation on all computers of the initial computer screens with the demographic questions.

Measures

CPTs

The CPTs consist of four tests: Elementary Algebra, Arithmetic, Reading Comprehension, and Sentence Skills. The CPTs are computer adaptive tests, and the same number of items, 12 to 20, depending on the test, are administered to all examinees. Examinees are required to attempt every item presented to them, and there is no penalty for guessing. The DOS 4.5 version of the CPTs was used. Two scores were obtained for each test:

1. The regular Total Right Score. This score is an estimate of the number of items that the examinee would answer correctly in the original pool of 120 items for each test.
2. The total time (in seconds) spent on the items. (Time for Elementary Algebra and Arithmetic were unavailable for one examinee.)

Other measures of test performance in the Stricker (1997) study and the Steele and Aronson (1995) research based on the number of attempted, omitted, or not reached items could not be obtained because examinees must answer all items.

Other variables

Ethnicity, sex, and other background variables were obtained from the CPTs electronic records or the paper-and-pencil questionnaire. In cases where ethnicity and sex were not reported, this information was obtained from school records.

Analysis

The product-moment intercorrelations of the scores and times for the four tests were computed separately for the experimental and control groups, using a pair-wise missing data program.

A series of 2 (Experimental versus Control) x 3 (Ethnicity—White, African American, Other) x 2 (Sex) factorial analyses of covariance of the eight scores and times were carried out, using the least-squares method (Model II error term; Overall and Spiegel, 1969) to deal with unequal *N*s. Sixteen covariates were used. In cases where the data for a covariate were not reported or were unquantifiable (ranging from .4 percent for Age to 21.7 percent for Father's Education), the mean or modal response for examinees of the same ethnic group and sex in the same experimental or control group was substituted. Questions with open-ended response alternatives (e.g., "seven or more years") were dichotomized at the median of the distributions. Intended program: Diploma in vocational field was excluded to eliminate the dependency among the four Intended program dummy variables. The covariates follow:

1. Age (in years)
2. Father's education (high school graduate or less = 0, some college or more = 1)
3. Mother's education (high school graduate or less = 0, some college or more = 1)
4. English is first language (yes = 1, no = 0)
5. Disability (yes = 1, no = 0)
6. Years of English in high school (three years or less = 0, four years or more = 1)
7. Years of mathematics in high school (three years or less = 0, four years or more = 1)
8. Studied algebra in high school (yes = 1, no = 0)
9. Years since mathematics training (less than one year = 0, one year or more = 1)
10. Intended program: Associate's degree in arts and science (this program = 1, all other programs = 0)
11. Intended program: Associate's degree in vocational field (this program = 1, all other programs = 0)
12. Intended program: Undecided (this program = 1, all other programs = 0)
13. CPTs Elementary Algebra test taken (yes = 1, no = 0)
14. CPTs Arithmetic test taken (yes = 1, no = 0)
15. CPTs Reading Comprehension test taken (yes = 1, no = 0)
16. CPTs Sentence Skills test taken (yes = 1, no = 0)

Planned comparisons of simple main effects of the experimental versus control group factor for each ethnic group (e.g., African-American examinees in the experimental group versus African-American examinees in the control group) and each sex (e.g., women in the experimental group versus women in the control group) were also conducted (Howell, 1997).

Note that the analyses of covariance (and comparisons of simple main effects) use unweighted means. Ef-

fect sizes were assessed by the correlation ratio (η). Both statistical and practical significance were considered in evaluating the results. An .05 significance level and an η of .10 (Cohen's, 1988, definition of a "small" effect size) were employed throughout (including the comparisons of simple main effects; Keppel, 1982).

Results and Discussion

Intercorrelations

The intercorrelations of the scores and times on the four tests for the experimental and control groups are reported in Table 2. The correlations were similar for the two groups. Scores on the two quantitative tests correlated highly (.70 and .71) as did scores on the two verbal tests (.73 and .78). Times on the two kinds of tests also correlated highly (.58 and .59 for quantitative tests, .75 and .76 for verbal tests). Apart from a substantial correlation (.51 and .52) between score and time for Elementary Algebra, suggestive of speededness, the correlations between corresponding scores and times for the tests were modest. The correlations between scores and times on different kinds of tests were also generally modest except for substantial correlations of time on Arithmetic with time on Reading Comprehension (.51 and .59) and time on Sentence Skills (.53 and .58).

In short, apart from a few highly related scores or times, most of the variables were relatively independent of each other. The high correlations that were observed demonstrate that all of the scores and times had substantial reliability.

Analyses of Covariance

The analyses of covariance of the scores and times on the tests, as well as the related planned comparisons, are summarized in Table 3; the corresponding means for the subgroups in the experimental and control groups appear in Table 4.

Focusing on differences between the experimental and control groups for each ethnic group and sex, none of the eight two-way interactions of experimental versus control group with ethnicity, none of the eight two-way interactions of experimental versus control group with sex, and none of the eight three-way interactions of experimental versus control group with ethnicity and sex were both statistically and practically significant. In addition, none of the 24 simple main effects for ethnicity (white, African American, other) and none of the 16 simple main effects for sex were significant.

In brief, the test scores and times for an ethnic group or sex were unrelated to whether examinees were asked about their ethnicity or sex.

Conclusions

This study of the CPTs not only replicated the Stricker (1998) investigation of the AP Examination by failing to find a connection between inquiring about examinees' ethnicity and sex and how well they did on the tests but also extended the initial investigation by failing to find a connection with how long examinees worked on the tests. The convergence between the two studies, which differed in tests and test-taking populations, supports the generalizability of these negative outcomes and contrasts with the Steele and Aronson (1995) findings of

TABLE 2

Intercorrelations of Scores and Times on Tests

	Variable							
Variable	***(1)***	***(2)***	***(3)***	***(4)***	***(5)***	***(6)***	***(7)***	***(8)***
1. Elementary Algebra—Score	—	.51	.71	.05	.24	-.08	.28	-.11
2. Elementary Algebra—Time	.52	—	.48	.59	.19	.34	.20	.30
3. Arithmetic—Score	.70	.44	—	.20	.48	.01	.47	-.05
4. Arithmetic—Time	.12	.58	.18	—	.03	.59	.04	.58
5. Reading Comprehension—Score	.37	.28	.57	.12	—	-.07	.73	-.22
6. Reading Comprehension—Time	-.08	.35	-.12	.51	-.06	—	-.12	.75
7. Sentence Skills—Score	.44	.38	.57	.19	.78	-.05	—	-.22
8. Sentence Skills—Time	-.17	.26	-.21	.53	-.22	.76	-.13	—

Note: Correlations for the control group appear above the diagonal; correlations for the experimental group appear below it. *N*s vary from 499 to 656 for the control group, and from 439 to 581 for the experimental group. For *N*s of 656 and 581, correlations of .07 and .08 are significant at the .05 level (two-tail) respectively, and correlations of .10 are significant at the .01 level for both *N*s.

TABLE 3

Summary of Analyses of Covariance of Scores and Times on Tests

		F							
		Elementary Algebra		*Arithmetic*		*Reading Comprehension*		*Sentence Skills*	
Source	df	*Score*	*Time*	*Score*	*Time*	*Score*	*Time*	*Score*	*Time*
Experimental Control (E-C)	1	1.68	1.20	.12	2.20	.00	4.10*	.42	3.84*
Sex	1	7.41*	.13	11.94**	.32	.58	.61	.89	2.13
E-C x Sex	1	1.37	.36	.31	.04	.75	1.22	.87	.90
Male	1	3.62	.15	.48	.96	.42	.50	.05	.66
Female	1	.01	1.24	.02	1.25	.33	4.29	1.03	3.49
Ethnicity	2	28.17**[a]	.97	73.09**[a]	2.92*	58.18**[a]	9.06**[a]	51.10**[a]	22.63**[a]
E-C x Ethnicity	2	1.08	.23	1.51	.18	.30	1.19	1.08	.54
White	1	.08	2.08	2.17	2.08	.15	1.84	.03	1.25
African American	1	.18	.10	.08	2.95	.41	.02	.77	.52
Other	1	2.15	.29	1.48	.08	.03	3.75	1.36	2.31
Ethnicity x Sex	2	2.37	.46	3.28*	.71	.32	1.67	1.18	3.28*
E-C x Ethnicity x Sex	2	.70	.07	.24	.48	1.62	1.07	4.14*	3.56*

Note: The *df* for Error and Mean Square Error are 1150 and 542.99 for Elementary Algebra—Score, 1,150 and 132,605.00 for Elementary Algebra—Time, 1,211 and 614.13 for Arithmetic—Score, 1,210 and 188,970.50 for Arithmetic—Time, 1,017 and 386.82 for Reading Comprehension—Score, 1,017 and 280,482.20 for Reading Comprehension—Time, 1,046 and 401.35 for Sentence Skills—Score, and 1,046 and 222,828.70 for Sentence Skills—Time. *$p < .05$; ** $p < .01$; [a]$\eta > .10$

TABLE 4

Mean Scores and Times on Tests for Ethnic Groups and Men and Women

	Ethnicity						*Sex*				
	White		*African American*		*Other*		*Men*		*Women*		
Variable	*Exp*	*Con*	*Exp*	*Con*	*Exp*	*Con*	*Exp*	*Con*	*Exp*	*Con*	S.D.[a]
Elementary Algebra											
Score	52.31	51.80	40.17	41.15	45.51	51.86	47.37	51.69	44.63	44.86	22.91
Time	676.95	633.93	630.13	618.52	684.82	647.40	660.73	646.64	667.21	619.93	364.15
Arithmetic											
Score	68.44	65.51	48.01	47.32	54.68	60.23	59.75	61.41	54.34	53.96	24.78
Time	1029.96	979.46	1103.87	1031.60	1083.82	1061.80	1059.93	1018.47	1085.18	1030.10	434.71
Reading Comprehension											
Score	79.62	80.29	65.83	64.47	70.96	71.60	72.06	73.44	72.21	70.79	19.67
Time	1452.16	1389.49	1580.01	1572.10	1629.25	1434.45	1546.98	1506.54	1560.63	14424.15	529.61
Sentence Skills											
Score	86.37	86.68	74.00	72.14	72.11	76.87	77.48	77.02	77.51	80.11	20.03
Time	1086.11	1040.62	1301.87	1265.73	1247.53	1101.29	1222.05	1182.74	1201.62	1089.03	472.05

[a] Calculated from the Mean Square Errors in the analyses of covariance.

differences in test performance produced by inquiring about ethnicity.

The present study, like the previous one, differed from the Steele and Aronson research in some respects that may account for the divergent findings, as already mentioned and as discussed in detail by Stricker. The Steele and Aronson research employed subjects in a laboratory study, whereas this investigation used examinees taking an operational test with real-life consequences. As a result, the CPTs examinees may be more motivated to do well on the tests, offsetting the adverse effects of stereotype threat.

Relatedly, the Steele and Aronson subjects were led to believe that they were engaged in an innocuous problem solving task whereas the examinees in this study were aware that they were being tested for their academic skills. If examinees perceive the CPTs as diagnostic of their cognitive resources, thereby generating stereotype threat, it is entirely conceivable that inquiring about ethnicity and sex cannot further increase the stereotype threat. However, it is not at all certain that stereotype threat is actually at its maximum on these tests.

Several processes suggested by Stricker that might account for the difference between his AP study and the Steele and Aronson research are made less plausible by differences in the research designs of the Stricker study and the present one. It was argued that attributions for poor performance may not be the same for the AP Examination and the GRE verbal items used by Steele and Aronson because the AP Examination is linked to a particular course. Thus, examinees may attribute their poor performance on the AP Examination to an inadequate course, not to their own characteristics or those of their ethnic group or sex, thereby blocking the effects of stereotype threat. But the CPTs are not tied to particular courses, making such attributions less likely, though examinees might still attribute their poor performance to substandard schooling in general. These remote attributions could also be made, though, by the subjects in the Steele and Aronson, and Spencer et al. (1997) research, for the test items that they used were similar in content to most of the CPTs (Steele and Aronson's GRE verbal items and CPTs Reading Comprehension, and Spencer et al.'s GRE quantitative and mathematics items and CPTs Elementary Algebra and CPTs Arithmetic). Nevertheless, the Steele and Aronson, and Spencer et al., test items were able to elicit stereotype threat.

It was also suggested that stereotype threat in the Stricker study may have been vitiated by feedback during the course that either inoculated the examinees against stereotype threat or caused them to disidentify with the course material and thereby eliminated the ego involvement that stereotype threat requires to be effective. But no such feedback exists for the CPTs.

Finally, it was proposed that examinees may perceive quantitative tests, such as the AP Calculus Examination in the Stricker study, as more difficult than verbal tests, such as the verbal items in the Steele and Aronson research. Insofar as a test is seen as beyond the examinee's ability level, stereotype threat may not operate. But both verbal and quantitative tests were used in the present study, and the same results were obtained with both kinds of tests.

Other differences exist between the present study and the Steele and Aronson research but are unlikely to account for the divergent results. First, the two-year college students are probably less ego involved in the academic skills assessed by the tests and less able than the Stanford undergraduates in the Steele and Aronson research and the AP students in the Stricker investigation. This difference is unlikely to be important, for the Steele and Aronson and the Stricker findings diverged despite being based on essentially the same population of students.

Second, the sample was large (totaling 1,341 examinees, including 468 African-American and 694 female test takers), roughly comparable in size to the sample in the Stricker research (though the sample of African-American examinees was much greater in the present study), and substantially larger than the samples in the Steele and Aronson experiments (44 and 20 African-American subjects). Hence, the statistical power to detect mean differences was appreciable in this study.

Third, the CPTs, because they are computer adaptive, are geared to administer items at each examinee's ability level, with the result that he or she should be able to answer about 60 percent correctly (allowing for chance success). The Steele and Aronson research used a conventional testing approach, with all examinees being given the same items. The average item difficulty (mean percent correct) in the Steele and Aronson research was about 50 percent, comparable to the difficulty in the present study.

Fourth, unlike the Steele and Aronson research, which adjusted for ability differences by covarying on the SAT®'s Verbal score (Donlon, 1984), no direct control for ability was employed. However, ability was indirectly controlled by covarying on amount of high school course work in English and mathematics, for such course work is substantially related to performance on ability tests (e.g., College Board, 1997; Laing, Engen, and Maxey, 1987; Morgan, 1989). Using course work as a covariate sidesteps the interpretive complexities inherent in employing as a covariate performance on a test that may also be susceptible to stereotype threat. Although control for ability is unneeded to eval-

uate differences between the experimental and control conditions within an ethnic group or sex, it is useful in comparing the interaction between ethnic group or sex and experimental versus control group in the Steele and Aronson research and the present study.

Fifth, in contrast to the Steele and Aronson research (and the Stricker study), examinees were not randomly assigned to the experimental and control groups. However, randomization was approximated by assigning examinees tested at different time periods to groups, and a large number of covariates were used to adjust for any secular trends in the nature of examinees that might exist. It is still possible, though improbable, that unadjusted but relevant differences are present between the examinees in the experimental and control groups.

Finally, a point of similarity between this study and the Steele and Aronson research—examinees were individually tested (in this study and in one of the Steele and Aronson experiments, the testing was done by computer)—lends support to the argument that any depersonalization associated with the group test administration in the Stricker study was unlikely to produce the different results in that study and in the Steele and Aronson research.

The present study reinforces the findings of the Stricker research and rules out some, but not all, of the alternative explanations for the differences between these investigations and the seminal experiments by Steele and Aronson. It is becoming increasingly clear that simply asking about ethnicity and sex is unlikely to degrade the performance of examinees who take standardized tests in real-life settings. However, the broader consequences of stereotype threat for the functioning of these tests remains a matter of speculation that needs to be documented in field research.

References

Briel, J. B., O'Neill, K. A., and Schueneman, J. D. (1993). *GRE technical manual*. Princeton, NJ: Educational Testing Service.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

College Board. (1994). *Advanced placement course description, mathematics, calculus AB, calculus BC—May 1995, May 1996.* New York: College Entrance Examination Board.

College Board. (1995). *ACCUPLACER program overview: Coordinator's guide.* New York: College Entrance Examination Board.

College Board. (1997). *College-bound seniors, 1997* (national report). New York: College Entrance Examination Board.

College Board and Educational Testing Service. (1995). *A guide to the advanced placement program, May 1996.* Princeton, NJ: Educational Testing Service

Donlon, T. F. (Ed.). (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests.* New York: College Entrance Examination Board.

Howell, D.C. (1997). *Statistical methods for psychology* (4th ed.). Belmont, CA: Wadsworth.

Keppel, G. (1982). *Design and analysis—A researcher's handbook* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Laing, J. H., Engen, H. B., and Maxey, J. (1987). *Relationships between ACT test scores and high school courses* (ACT Research Rep. No. 87-3). Iowa City, IA: American College Testing Program.

Morgan, R. (1989). *An examination of the relationships of academic coursework with admissions test performance* (College Board Rep. No. 89-6; ETS Research Rep. No. 89-38). New York: College Entrance Examination Board.

Overall, J.E., and Spiegel, D. K. (1969). Concerning least squares analysis of experimental data. *Psychological Bulletin, 72*, 311–322.

Spencer, S. J., Steele, C. M., and Quinn, D. (1997). *Stereotype threat and women's math performance.* Manuscript submitted for publication.

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

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797–811.

Stricker, L. J. (1998). *Inquiring about examinees' ethnicity and sex: Effects on AP Calculus AB Examination performance* (College Board Rep. No. 98-1; ETS Research Rep. No. 98-5). New York: College Entrance Examination Board.

Ward, W. C. (1988). The College Board Computerized Placement Tests: An application of computerized adaptive testing. *Machine-Mediated Learning, 2*, 271–282.

Appendix

BACKGROUND INFORMATION

_____ LAST NAME

_____ FIRST NAME

____ MIDDLE INITIAL

_____ - _____ - _____ STUDENT IDENTIFICATION NUMBER
(Social Security Number)

___ - ___ - ___ DATE OF BIRTH
Month, followed by the day, followed by the year
Examples: 10-23-70 for October 23, 1970
01-01-70 for January 1, 1970

___ - ___ - ___ TODAY'S DATE

The following questions ask for information that will be useful in research and evaluation of the test. Your responses to these questions are voluntary. If you choose not to answer a question, select Omit as your response.

1. What is the total number of years you studied English in high school (grades 9-12)? Count less than a full year of English as a full year, but do not count a repeated year of the same course as an additional year of study.

____ One year or the equivalent
____ Two years or the equivalent
____ Three years or the equivalent
____ Four years or the equivalent
____ More than four years or the equivalent
____ I did not take any courses in English
____ Omit

2. What is the total number of years you studied mathematics in high school (grades 9-12)? Count less than a full year of mathematics as a full year, but do not count a repeated year of the same course as an additional year of study.

____ One year or the equivalent
____ Two years or the equivalent
____ Three years or the equivalent
____ Four years or the equivalent
____ More than four years or the equivalent
____ I did not take any courses in mathematics
____ Omit

3. Did you study algebra for at least one semester in high school?

____ Yes
____ No
____ Omit

4. How long has it been since you have taken a mathematics course or other formal mathematics training?

____ Less than one year
____ One to three years
____ Four to six years
____ Seven or more years
____ Omit

5. What is your sex?

____ Female
____ Male
____ Omit

6. How do you describe yourself?

____ Native American, American Indian, or Alaskan Native
____ Black or African American
____ Mexican American or Chicano
____ Puerto Rican
____ Other Hispanic, Latino, Central American, or South American
____ Asian or Pacific American
____ White (non-Hispanic) or Caucasian
____ Other
____ Omit

7. Is English the first language you learned?

____ Yes
____ No
____ Omit

8. What documented disabling condition do you have, if any, that might affect the usefulness of your test scores as measures of your skills? (Select only one.) Upon receiving your results you may wish to contact student services for advice.

____ None
____ Blindness or other visual impairment
____ Deafness or other hearing impairment
____ Paraplegia
____ Learning Disability
____ Other neurological or orthopedic impairment
____ Other
____ Omit

9. What is the highest level of education completed by your father or male guardian?

____ Grade school or less
____ Some high school
____ High school diploma or equivalent
____ Business or trade school
____ Some college
____ Associate degree
____ Bachelor's degree
____ Some graduate or professional school
____ Completed graduate or professional school
____ Omit

10. What is the highest level of education completed by your mother or female guardian?

____ Grade school or less
____ Some high school
____ High school diploma or equivalent
____ Business or trade school
____ Some college
____ Associate degree
____ Bachelor's degree
____ Some graduate or professional school
____ Completed graduate or professional school
____ Omit