GRE

RESEARCH REPORT

# Stereotype Threat, the Test-Center Environment, and Performance on the GRE General Test

Alyssa M. Walters
Soonmook Lee
Catherine Trapani

December 2004

**GRE Board Report No. 01-03R**
**ETS RR-04-37**

# Stereotype Threat, the Test-Center Environment, and Performance on the GRE General Test

Alyssa M. Walters

ETS, Princeton, NJ

Soonmook Lee

Sung Kyun Kwan University, Republic of Korea

Catherine Trapani

ETS, Princeton, NJ

GRE Board Research Report No. 01-03R

December 2004

The report presents the findings of a research project funded by and carried out under the auspices of the Graduate Record Examinations Board.

Educational Testing Service, Princeton, NJ 08541

**********

Researchers are encouraged to express freely their professional judgment. Therefore, points of view or opinions stated in Graduate Record Examinations Board reports do no necessarily represent official Graduate Record Examinations Board position or policy.

**********

The Graduate Record Examinations and Educational Testing Service are dedicated to the principle of equal opportunity, and their programs, services, and employment policies are guided by that principle.



Educational Testing Service
Princeton, NJ 08541

**Abstract**

The study investigated the applicability of previous experimental research on stereotype threat to operational Graduate Record Examinations® (GRE®) General Test testing centers. The goal was to document any relationships between features of the testing environment that might cue stereotype threat as well as any impact on GRE test scores among African American, Hispanic, Asian American, and female test-takers. Among such features were the gender and ethnicity of test proctors and more general factors, such as the size, activity level, and social atmosphere of test centers. Our analyses revealed several relationships among environmental factors and several variations in test performance for all groups. However, we found no direct support for stereotype threat and, in fact, found some effects for proctor ethnicity that ran counter to a stereotype-threat explanation.

Key words: Stereotype threat, testing environment, gender, ethnicity

**Acknowledgements**

The authors wish to thank Venus Mifsud and Joshua Aronson for participating in observations of Prometric Testing Centers; Susan Martin and Alyson Tregidgo for assisting with this draft and coordinating the data collection; Debra Friedman and Shauna Cooper for assisting with the creation of the Test Center Survey; Don Powers, Patrick Kyllonen, Carol Dwyer, and Larry Stricker for commenting on this draft; Candus Hedburg and Edward Winner for providing counsel on test-center environments; Betsy Mayott and Carl Hyman of Prometric Testing Centers for assisting and willingly cooperating in the study; and Frank Jenkins, Brent Bridgeman, Dan Eignor, and Don Rock for providing feedback on the statistical analyses.

## Table of Contents

Page

## List of Tables

Page

## Introduction

A growing body of experimental research has explored the effects of *stereotype threat* within the standardized testing domain. This work posits that social-environmental factors that are apparent during test-taking may contribute to the testing gap that favors males and White test-takers over African American, Hispanic, and female test-takers. Specifically, cues in the test environment are believed to evoke negative stereotypes about ability, which in turn can impair test performance among test-takers who are members of the group or groups associated with that stereotype. The effects of stereotype threat have been demonstrated within the context of nonoperational testing situations for several groups of test-takers. For instance, African American (Steele & Aronson, 1995) and Hispanic students (Aronson & Salinas, 1997) performed worse on verbal items taken from the Graduate Record Examinations® (GRE®) General Test when the items were introduced as diagnostic measures of verbal ability, and women performed worse on GRE quantitative items when reminded of the stereotype that women are poor at math (Spencer, Steele, & Quinn, 1999). Interestingly, in comparison to a control group for whom no identity was made salient, Asian American women performed worse on a test of quantitative ability when *gender* stereotypes were apparent, but performed *better* when their Asian American identity was activated (Shih, Pittinsky, & Ambady, 1999). Taken together, this work highlights the important role certain environmental cues may play in enabling or disabling test-taker performance.

Researchers have experimented with a variety of ways to induce stereotype threat among study participants. A common finding is that subtle cues embedded in the instructions given prior to a test can activate stereotypes (e.g., Steele & Aronson, 1995). For instance, introducing a test as being diagnostic of intellectual ability can prime ability-related stereotypes. More recent work has found that this manipulation impairs test performance by causing a reduction in working memory (Schmader & Johns, 2003). Specifically, when a performance-related stereotype is made salient, working memory capacity is reduced for those participants who are affected by the stereotype, and this reduction, in turn, leads to a decline in performance.

A critical question, therefore, is whether stereotype-activating cues exist in actual operational settings—absent an experimenter's intentional manipulation of the environment. Recent research conducted within environments designed to mimic operational test centers found that variations in the social environment, including the gender and ethnicity of proctors, appeared to activate stereotypes. Walters, Shepperd, and Brown (2003) found that African American

students performed worse than White students on GRE verbal items when the proctor was White, but equal to White students when the proctor was African American. Similarly, women performed worse than men on a difficult math test only when the proctor was male (Marx & Roman, 2002). These studies suggest that stereotype threat may also be induced by the ethnicity and gender of proctors in operational settings.

All of these studies were laboratory investigations. Thus far, the effects of these environmental factors have not been demonstrated in an operational setting. The only previous stereotype-threat studies conducted in an operational setting attempted to induce stereotype threat by asking students to report their gender and ethnicity either prior to or upon completion of a placement test; however, this manipulation was generally unrelated to performance (Stricker, 1998; Stricker & Ward, 1998).

## Overview of Study

We believe that the new evidence about the impact of test administrator ethnicity and gender on student performance in nonoperational settings warrants further investigation in operational settings. Thus, with the current research, we aimed to extend experimental work on environmental cues and stereotype threat to operational GRE testing environments. Our primary objective was to determine whether there is any association between the gender and ethnicity of proctors and the test scores of African American, Hispanic, and female test-takers.

However, an array of variables in the test environment may also prime negative performance-related stereotypes that interfere with test-takers' cognitive processing abilities. Given prior research on stereotype threat, we suspect that such variables could include the size and activity level of a center. Large, active centers may create feelings of anonymity among test-takers, while test-takers in smaller, less-frequented centers may feel like the center of attention. In addition, the center's social atmosphere (as measured by its typical protocol and behaviors of proctors) may affect whether stereotypes are made more or less salient. For instance, the extent to which an atmosphere is more or less comfortable for a given test taker may affect whether a negative stereotype is primed and sustained throughout the testing session. Currently, little is known about these environmental variables and how they relate to performance – or how they may interact with characteristics of proctors. Thus, an important secondary goal of our research was to assess variables that might evoke performance-related stereotypes, so that we could both account

for them in our analyses of proctor gender and ethnicity and also explore their relation to GRE General Test scores.

We divided our research into three related studies. In Study 1, our objective was to determine a viable set of variables that characterized each participating operational test center—including the gender and ethnicity of all proctors as well as the size, activity level, and social atmosphere of the centers. In Studies 2 and 3, we used hierarchical linear modeling (Raudenbush& Bryk, 2002) to assess the relationship between test scores and whether test-takers shared the ethnicity (Study 2) and gender (Study 3) of test-center proctors, as well as to assess the role of other environmental variables in student performance. Each study is described in turn in the sections that follow.

## Study 1: Test-Center Effects

### *Method*

*Participants.* Supervising proctors from 249 Prometric Testing Centers and test centers housed within colleges or universities (which are referred to throughout this report as "institutional" test centers) participated in the studies. The locations of the test centers spanned the continental United States.

*Observations.* The research team observed and met with supervisors at four Prometric Testing Centers and one institutional testing center in June of 2001 to determine a viable set of atmosphere variables to investigate, as well as to learn the general protocol and configuration of each center. A Prometric center was selected from each of four settings: urban, suburban, large, and small. At each center, researchers conducted an extensive interview with the supervising proctor and observed all testing procedures from start to finish for approximately 10 students as they completed the GRE General Test.[1]

*Questionnaire development & administration.* The information garnered from test-center visits led to the creation of a Test-Center Survey (see Appendix A), which was designed to collect information about proctors, center size, typical activity levels and behaviors, the extent to which test-takers' appear to be aware of these behaviors, and test-center atmosphere.[2] The survey was mailed to all 348 centers in the continental United States[3] in August 2001, with a request that the questionnaire be completed by the supervising test administrator at each center. A total of 249 surveys were returned (72% return rate).

*Mean GRE General Test scores.* In order to assess correlations among test-center-level variables and test performance, we accessed GRE verbal and quantitative scores from the GRE data storage warehouse for all test-takers who took the GRE General Test at one of our participating centers during the months of June, July, and August 2001 ($N$ = 28,478). Mean scores were then calculated for each test center.

***Results and Discussion***

*Factor analysis of questionnaire items.* To reduce the number of items identified in the questionnaire to a set of composite environment variables, we performed a principal components analysis on all questionnaire items, except those that asked about proctor gender and ethnicity. Based on the scree plot (see Carroll, 1993; Cattell, 1966), we retained a five-factor solution and subjected these factors to a varimax rotation. We retained all items with factor loadings greater than .40, and considered an item part of a given factor if it loaded greater on that factor than on others. Examination of the factor loadings revealed that a three-factor solution, which accounted for 51% of the total variance, was a more substantive model than the five-factor solution. Test centers that omitted questionnaire items were excluded from the analysis, thus reducing the sample of test centers to $N$ = 182.

After examining the pattern of factor loadings, we called Factor 1 the "warm/friendly" factor and Factor 2 the "formal/professional" factor to reflect settings that appear to be high or low on these social dimensions. The third factor, which we named the "disruptive" factor, indicates the extent to which test-takers appear to notice disruptive elements of the setting. We created the scale scores by summing the high-loading items.[4] Additional items of interest that did not load onto the three factors, but were included in subsequent test-center-level analyses, were test-center size (as measured by number of testing stations at each test center)[5] and activity level (a composite of the perceived, overall level of activity at the center and how many test-takers actually use the center on a typical day [$r$ = .51, $p$ < .001], which were both measured on a 5-point scale). Table 1 displays factor loadings for the final three-factor solution, and Table 2 presents correlations among the predictors and test scores.

**Table 1**

***Factor Analysis of Questionnaire Items After Rotation***

| Questionnaire item | Warm/ friendly | Formal/ professional | Disruptive |
|---|---|---|---|
| *Test-takers notice the following:* | | | |
| Door opening and closing | .02 | .04 | **.62** |
| Other test-takers during the test | -.01 | -.02 | **.58** |
| Other test-takers starting the test | .03 | -.02 | **.54** |
| Surveillance equipment | -.00 | .11 | **.54** |
| Proctor activity behind the window | -.03 | .08 | **.53** |
| Different check-in procedures | .11 | -.04 | **.45** |
| *Typical center atmosphere includes the following:* | | | |
| Fun | **.71** | -.02 | .11 |
| Youthful | **.62** | .07 | .22 |
| Lively | **.63** | .20 | .16 |
| Warm/cozy | **.58** | .09 | -.09 |
| Formal | .01 | **.68** | .06 |
| Library-like | -.01 | **.54** | .04 |
| Serious | .09 | **.50** | .12 |
| Critical | .08 | **.43** | .32 |
| *Proctors should have the following qualities:* | | | |
| Warm | **.71** | .04 | .08 |
| Relaxed | **.55** | -.09 | -.02 |
| Nurturing | **.55** | .20 | -.00 |
| Youthful | **.52** | .18 | .14 |
| Open-minded | **.46** | .18 | -.04 |
| Compassionate | **.45** | .10 | -.11 |
| Mellow | **.41** | -.03 | .13 |
| Formal | .10 | **.70** | .14 |
| Serious | .03 | **.69** | .13 |
| Business-like | .13 | **.66** | .00 |
| Strict | .03 | **.55** | .15 |
| Professional | .11 | **.49** | -.13 |
| **Eigenvalues** | **5.77** | **3.78** | **2.74** |

*Note.* All items were measured on a 5-point scale. Factors greater than .40 are in bold.

**Table 2**

***Correlations Among Test-Center-Level Variables***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | *M (SD)* |
|---|---|---|---|---|---|---|---|---|---|
| 1. Warm/friendly | — | | | | | | | | 34.0 (6.9) |
| 2. Formal/professional | .20** (212) | — | | | | | | | 29.0 (5.5) |
| 3. Disruptive | .08 (217) | .23** (220) | — | | | | | | 18.3 (4.3) |
| 4. Test-center size | .15* (225) | .09 (225) | .15* (233) | — | | | | | 2.6 (1.2) |
| 5. Activity level | .02 (220) | .06 (220) | -.07 (228) | .12 (242) | — | | | | 7.8 (1.2) |
| 6. Median income | .09 (222) | .09 (222) | .05 (230) | .29** (244) | .24** (239) | — | | | 36,688 (11,301) |
| 7. Test-center mean GRE quantitative score | .01 (225) | -.07 (225) | .04 (233) | .22** (247) | .21** (242) | .37** (242) | — | | 529 (146) |
| 8. Test-center mean GRE verbal score | .04 (225) | -.10 (225) | -.01 (233) | .21** (247) | .18* (242) | .28** (244) | .84** (247) | — | 464 (131) |

* $p < .05$. ** $p < .01$.

*Demographic measure: Median income.* It was important to account for differences between test centers that go beyond the environmental variables of interest and might relate to test performance. Using 2000 U. S. Census data and the zip code for each test center, we determined the median family income for the neighborhood in which each center was located—a factor that we believed was likely to relate to differences in standardized test performance. We included median income as a continuous variable in all subsequent analyses as a test-center-level demographic variable (see, e.g., Table 2).[6]

Several test-center-level variables were correlated, and test scores were positively related to median income, test-center size, and test-center activity level. Study 1 allowed us to establish a set of variables that characterize each center in our sample and to explore relationships among those variables and mean test scores. The next step was to investigate these test-center-level variables in more depth by adding test-taker-level variables and also by considering the role of test-taker "match" with the ethnicity and gender of proctors.

## Study 2: Proctor Ethnicity

### *Overview of Analyses*

As noted earlier, the purpose of Studies 2 and 3 was to determine whether stereotype threat, activated by the ethnicity or gender of proctors, affects GRE test scores. Our specific question was whether African American and Hispanic test-takers would perform better on the GRE General Test when proctors shared their ethnicity, and whether female test-takers would perform better on the GRE quantitative test when proctors shared their gender. According to the stereotype-threat hypothesis, they should.

However, one difficulty we encountered in testing this hypothesis was that we did not have information for each of the 28,478 test-takers about the gender or ethnicity of the proctor who was present at the time of their testing. We only had information about the number of White, African American, Hispanic, Asian American, male, and female proctors at each test center, along with information about the centers at which test-takers were tested. From these two pieces of information, we were able to determine for each test-taker the number of proctors in the center that matched his or her ethnicity and gender.

To be certain whether test-takers encountered proctors who were similar or different from them, we limited all subsequent analyses to test-takers who matched either all or none of the proctors in their respective centers. In Study 2, we selected test-takers who either matched or did not match the ethnicity of all of the proctors in their testing center. For example, for an Asian American student to be included in Study 2, s/he would have to have tested at a center staffed by either all Asian American proctors or no Asian American proctors (regardless of his/her gender.) In Study 3, we applied the same selection criterion for gender instead of ethnicity. For instance, for a female to be included in Study 3, she would have to have tested at a center that had either no female proctors or all female proctors (regardless of her ethnicity). This dichotomous variable (All match vs. None match) was treated as an independent variable in Studies 2 and 3.[7]

A second difficulty concerned whether the proper unit of analysis for the stereotype-threat hypothesis test was the individual test-taker or the test center. Although stereotype-threat studies typically treat individual test-takers as the unit of analysis, we also wanted to consider the influence of test-center-level variables. Fortunately, hierarchical linear modeling (HLM) procedures allow for the simultaneous analysis of both the individual test-taker (within-center) and test-center (between-center) levels.

***Method***

*Test-taker sample.* Using information on test-taker gender, ethnicity, and GRE quantitative and verbal scores obtained from the test-taker data set used in Study 1, we matched test-taker data with the returned surveys, yielding a total test-taker sample of $N = 28,478$. However, as noted earlier, we limited our sample to test-takers who either matched all or none of the proctors in the center based on ethnicity or gender. In addition, after applying this criterion, centers with samples of fewer than 10 test-takers were excluded because this sample size was insufficient for estimating center-level parameters. The sample used for the HLM analyses was further reduced to $N = 12,397$, because test-takers for whom data was missing were excluded.

For the analysis of proctor ethnicity, the sample was limited to test-takers identifying themselves as U. S. citizens and as White ($N = 3,720$ males and 6,612 females), Black/African American ($N = 302$ males and 775 females), Hispanic ($N = 165$ males and 307 females identifying as Mexican-American, Puerto Rican, Other Hispanic, or Latin American), or Asian/Pacific American ($N = 232$ males and 284 females). We did not analyze data for test-takers identifying as American Indian or Other due to the small sample size for these groups, and test-takers who did not report their ethnicity were excluded. The total sample was 65% female and 35% male.

*Design and statistical analysis.* We chose HLM (Raudenbush& Bryk, 2002) to assess the relationship between both test-taker-level and center-level variables and GRE General Test scores because this method allowed us to conduct a general linear modeling analysis at multiple levels simultaneously (i.e., test-taker level and test-center level) by treating lower-level regression parameters, such as the regression slope and intercept associated with test-taker characteristics, as dependent variables in higher-level analyses. The key to conducting an HLM analysis is the identification of first- and second-level factors. For our analysis, level-one variables, which pertained to test-takers, were:

- test-taker gender
- match with proctor ethnicity (0 or 100%)
- undergraduate grade-point average (UGPA)[8]
- mother's highest level of education completed[9]
- father's highest level of education completed
- years since bachelor's degree
- graduate education objective[10]

Test-taker-level variables were self-reported from the GRE Background Information Questionnaire[11]. Including this set of control variables was required so that we would not attribute differences in test performance to environmental or proctor variables that might more parsimoniously be attributable to academic-related differences between test-takers.

Level-two factors were the test-center variables described earlier:

- warm/friendly
- formal/professional
- disruptive
- test-center size
- activity level
- median income

Analyses were conducted separately for test-takers who identified themselves as members of one of the four largest ethnicity groups in our sample:

- White
- African American
- Hispanic
- Asian American

The HLM analyses proceeded in two steps for each test-taker ethnicity group. First, we partitioned the variance into the two levels using the fully unconditional model; second, we estimated the effects of predictors at each level using the conditional model. These steps are discussed in more detail in the subsections that follow.

*Fully unconditional model.* We began with the simplest fully unconditional model in which no predictors were specified at either level, which allowed us to assess how the variance in test scores was allocated across the two levels. At the test-taker-level, we modeled a test-taker's GRE score as a function of the respective test-center mean plus a random error, $Y_{ij} = \beta_{oj} + r_{ij}$, where $Y_{ij}$ is the GRE verbal or quantitative score of test-taker i in test-center j, $\beta_{oj}$ is the mean score of test-center j, and $r_{ij}$ is a random test-taker effect (the deviation of test-taker ij's score from the test-center mean). This level-one error, $r_{ij}$, is assumed to be normally distributed with a mean of 0 and a variance of $\sigma^2$. The subscripts i and j denote test-takers and test centers where there are i = 1, 2, …, $n_j$ test-takers within test-center j, and j = 1, 2, …, J test centers in this study.

At the test-center-level, $\beta_{oj} = \gamma_{oo} + u_{oj,}$ we viewed each test-center mean $\beta_{oj}$, as the mean outcome for the jth center. Here, $\gamma_{oo}$ is the average intercept across test centers, and $u_{oj}$ is a random test-center effect associated with test-center j. The $u_{oj}$ is also assumed to follow a normal distribution with a mean of 0 and variance $\tau_{00}$.

*Variance partitioning and reliability estimates.* This simple, two-level model allowed us to partition the total variability of score $Y_{ij}$ into the following two components: (a) variance among test-takers within test centers ($\sigma^2$) for level one, and (b) variance between test centers ($\tau_{00}$) for level two. This partition allowed us to estimate the proportion of within-center variance and between-center variance as follows: $\sigma^2/(\sigma^2+\tau_{00})$ is the proportion of variance in scores among test-takers within centers, and $\tau_{00}/(\sigma^2+\tau_{00})$ is the proportion of variance in scores between centers.

We examined the reliability of test-center mean $\bar{\bar{Y}}_{.j}$ for the parameter $\beta_{0j}$ using the formula:

$$\lambda_j = \text{var}(\beta_{oj})/\text{var}(\bar{\bar{Y}}_{.j})=\tau_{oo}/(\tau_{00}+V_j) \quad (1)$$

Here, we determined the reliability of the sample mean as an estimate of the true mean. We then averaged these reliability estimates across test centers and used the averages as summary measures of the reliability of the test-center means, which in turn indicated the degree to which we could discriminate among level-two units using the random parameter estimates. Low reliabilities did not invalidate the HLM analysis, because they could be fixed in subsequent analyses (Raudenbush, Bryk, Cheong, & Congdon, 2000).

*Conditional models.* The fully unconditional model allowed us to estimate the variability associated with the two levels. However, we anticipated that part of the variability at each level could be explained by measured variables at each level. Thus, we used the series of level-one test-taker variables listed earlier as predictors of individual GRE scores, and we used the level-two test-center-level variables listed earlier as predictors of test-center mean scores.

Further, we speculated that some of the relationships at the test-taker-level may be affected by test-center-level variables. For example, suppose that within centers, test-taker gender was found to be related to GRE quantitative score. The gender effect might depend on certain test-center characteristics (e.g., match with proctor gender or test-center size). If that were the case, the regression coefficient representing the gender effect would vary depending on some characteristics of the center. To understand these multi-level effects, then, we modeled a regression equation to predict level-one coefficients.

For level one, within each test center, we modeled test-taker GRE score as a function of test-taker-level predictors plus a random test-taker-level error:

$$Y_{ij} = \beta_{oj} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \ldots + \beta_{pj}x_{pij} + r_{ij} \quad (2)$$

Here, $Y_{ij}$ is the score of test-taker i in test-center j, $\beta_{oj}$ is the intercept for test-center j, and x is the test-taker variable. $\beta_{pj}$ represents the corresponding level-one coefficients that indicate the direction and strength of association between each test-taker variable, $x_p$, and test-taker i's score in test-center j. The random effect in level one, $r_{ij}$, represents the deviation of test-taker ij's score from the predicted score based on the test-taker model.

Each of the regression coefficients in the test-taker-level model (including the intercept) was then specified as fixed, randomly varying with predictors, or randomly varying without predictors using the following formulas.

Example of a regression coefficient modeled as fixed:

$$\beta_{pj} = \gamma_{pj} \quad (3)$$

Example of a regression coefficient modeled as randomly varying with predictors:

$$\beta_{pj} = \gamma_{po} + \gamma_{p1}w_{1j} + \gamma_{p2}w_{2j} + \ldots + \gamma_{pq}w_{qj} + u_{pj} \quad (4)$$

Example of a regression coefficient modeled as randomly varying without predictors (unstructured):

$$B_{pj} = \gamma_{oo} + u_{oj} \quad (5)$$

Here, p refers to the number of test-taker characteristics, $\gamma_{po}$ is the intercept across test centers, and w is a test-center variable used as a predictor of the level-one coefficient, $\beta_{pj}$. The corresponding coefficient, $\gamma_{pq}$, represents the direction and strength of the association between test-center characteristic $w_{qj}$ and $\beta_{pj}$. The level-two random effect, $u_{pj}$, represents the deviation of test-center j's level-one coefficient $\beta_{pj}$, from its predicted value based on the test-center model. If $u_{pj}$ is not zero, then the $\beta_{pj}$ is said to vary randomly. If only $\gamma_{po}$ is in the equation, then $\beta_{pj}$ is said to be fixed at the same value across test centers. There are p + 1 equations in the level-two model—one for each of the level-one coefficients. The random effects can be correlated.[12]

*Strategy for hypothesis-testing in HLM.* We tested hypotheses about fixed effects, random coefficients, and variance-covariance components at both levels. For each group designation (White, African American, Hispanic, Asian American, and male/female), we first tested the most

complex model. That is, for each group, we conducted regression analyses at level one to determine whether test-taker-level variables were significant predictors of test scores. Since UGPA is a known covariate of test scores, we included UGPA first in our analysis and then added other variables to see if they provided incremental explanation. We then used the final equation for interpretation:

$$Y_{ij} = \beta_{oj} + \beta_{1j}(\text{UGPA})_{1ij} + \beta_{2j}(\text{Gender})_{2ij} + \beta_{3j}(\text{Match with proctor})_{3ij} + \beta_{4j}(\text{Graduate objective})_{4ij} + \beta_{5j}(\text{Father's education})_{5ij} + \beta_{6j}(\text{Mother's education})_{6ij} + \beta_{7j}(\text{Year received bachelor's degree})_{7ij} + r_{ij} \quad (6)$$

We centered UGPA at its grand mean (as in an analysis of covariance) and retained the original metric for all other variables without centering. Only for β's with significant intercepts and random error variances did we attempt to introduce predictors (test-center-level variables) in the level-two model. In cases where random error variances were not significant, we fixed the level-one coefficients across test centers and tested the simpler model.

Finally, we provided a summary of all significant coefficients, effect sizes, and proportions of variance. The formula for the effect size for regression coefficients is $\delta=\gamma/\sigma y$. We divided each fixed effect coefficient( $\hat{\gamma}$ ) by SDy so that the effect size could be interpreted as the change of y in SD units corresponding to 1 unit change in the value of a predictor. In addition to effect size, we calculated the proportion of variance accounted for by adding predictors at the test-taker and test-center levels. We computed the variance at the test-taker level using the following equation:

$$\hat{\sigma}^2(\text{unconditional model}) - \hat{\sigma}^2(\text{level-1 predictors})/\hat{\sigma}^2(\text{unconditional model}) \quad (7)$$

where $\hat{\sigma}^2$ is the estimate of within-center variance. It is reduced by adding predictors to the unconditional model at the test-taker level.

For the test-center level, we computed the variance using the following equation:

$$\hat{\tau}_{qq}(\text{random regression model}) - \hat{\tau}_{qq}(\text{fitted model})/\hat{\tau}_{qq}(\text{random regression model}) \quad (8)$$

where $\hat{\tau}$ is the estimate of between-center variance. It is reduced by fitting predictors to the random regression model at the test-center level.

***Results and Discussion***

In this section we describe the results of our analyses in the following order. For each ethnicity group, we present the descriptive statistics and correlation matrices for test-taker and test-center-level variables. Next, we describe the results of the fully unconditional model (partitioning the variance into each level), followed by the results of the conditional model. Only final models with significant predictors are presented. This procedure was followed first for the verbal component of the GRE General Test and then for the quantitative section. The procedure was identical for all groups of test-takers.

Throughout this section, tables of test-center-level data are presented immediately following tables of test-taker-level variables for the GRE verbal test. Tables of test-center-level data are not repeated following tables of test-taker-level variables for the GRE quantitative test, because level-two data are the same for all test-takers, regardless of test.

*White test-takers—GRE verbal test.* Table 3 and Table 4 display descriptive statistics and correlations for White test-taker-level variables that completed the GRE verbal test. Table 5 and Table 6 present statistics for test-center-level variables.

**Table 3**

***Level-One Descriptive Statistics for White Test-Takers —GRE Verbal Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 10,332 | 0.64 | 0.48 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 10,332 | 0.94 | 0.23 | 0 (None match) | 1 (All match) |
| UGPA | 10,332 | 5.38 | 1.16 | 1 | 7 |
| Mother's highest level of education completed | 10,332 | 5.49 | 2.29 | 1 | 9 |
| Father's highest level of education completed | 10,332 | 5.75 | 2.51 | 1 | 9 |
| Years since bachelor's degree | 10,332 | 6.65 | 7.29 | -2 | 51 |
| Graduate education objective | 10,332 | 3.29 | 0.49 | 1 | 4 |
| GRE verbal score | 10,332 | 477 | 99 | 200 | 800 |

**Table 4**

***Correlations Among Level-One Variables for White Test-Takers—GRE Verbal Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor ethnicity | 0.01 | 1.00 | | | | | | |
| 3. UGPA | 0.10* | -0.01 | 1.00 | | | | | |
| 4. Mother's highest level of education completed | -0.03* | -0.05* | 0.07* | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.05* | -0.06* | 0.06* | 0.58* | 1.00 | | | |
| 6. Years since bachelor's degree | 0.00 | -0.01 | -0.19* | -0.19* | -0.16* | 1.00 | | |
| 7. Graduate education objective | -0.06* | -0.01 | 0.14* | 0.07* | -0.06* | -0.04* | 1.00 | |
| 8. GRE verbal score | -0.10* | -0.07* | 0.23* | 0.16* | 0.20* | 0.01 | 0.18* | 1.00 |

* $p \leq 0.05$.

**Table 5**

***Level-Two Descriptive Statistics for White Test-Takers***

| Variable | *N* of center | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Median Income | 131 | 35,725.91 | 9,326.55 | 19,725 | 71,504 |
| Test-center size | 131 | 2.56 | 1.23 | 1 | 5 |
| Formal/professional | 131 | 29.11 | 5.13 | 18 | 43 |
| Warm/friendly | 131 | 33.94 | 7.31 | 17 | 51 |
| Disruptive | 131 | 17.86 | 3.75 | 12 | 29 |
| Activity level | 131 | 5.67 | 1.10 | 3 | 8 |

**Table 6**

***Correlations Among Level-Two Variables for White Test-Takers***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Median Income | 1.00 | | | | | |
| 2. Test-center size | 0.28* | 1.00 | | | | |
| 3. Formal/professional | 0.10 | 0.01 | 1.00 | | | |
| 4. Warm/friendly | 0.05 | 0.11 | 0.27* | 1.00 | | |
| 5. Disruptive | 0.03 | -0.08 | 0.06 | -0.05 | 1.00 | |
| 6. Activity level | 0.27* | 0.07 | 0.10 | 0.08 | -0.04 | 1.00 |

* $p \leq 0.05$.

When we partitioned the total variance in GRE verbal score into the respective within-center and between-center components using the fully unconditional model, within-center and between-center variance accounted for 93.7% and 6.33% of the score variance, respectively (see Table B1). Next, to investigate the source of the variance using the conditional model, we looked for predictors at the test-taker-level. For the within-center model, test-takers' scores in a test-center j were associated with UGPA ($\beta = 17.72$, $p < .01$), test-taker gender ($\beta = -19.47$, $p < .01$), match with proctors ($\beta = -16.11$, $p < .05$), graduate objective ($\beta = 25.20$, $p < .01$), and father's education ($\beta = 5.67$, $p < .01$). For the between-center model, test-center mean was associated with median income ($\beta = 7.30$, $p < .01$) and test-center size ($\beta = 4.51$, $p < .01$). (Table B2 displays these results.) At the within-center level—after controlling for UGPA—test-taker gender, match with proctor, graduate objective, and father's education were related to GRE-verbal score. However, no test-center-level variables had a cross-level effect on the slope of test-taker-level predictors.

*White test-takers—GRE quantitative test.* Next, we repeated the procedures for White test-takers who took the quantitative portion of the GRE General Test. The sample was reduced by $N = 79$ test-takers. (All discrepancies between GRE verbal and quantitative samples are likely due to the fact that test-takers had the option to cancel a test score upon completion of the test.) Table 7 and Table 8 present level-one descriptive statistics and correlations among variables for White test-takers who completed the GRE quantitative test.

**Table 7**

***Level-One Descriptive Statistics for White Test-Takers—GRE Quantitative Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 10253 | 0.64 | 0.48 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 10253 | 0.94 | 0.23 | 0 (None match) | 1 (All match) |
| UGPA | 10253 | 5.38 | 1.16 | 1 | 7 |
| Mother's highest level of education completed | 10253 | 5.50 | 2.29 | 1 | 9 |
| Father's highest level of education completed | 10253 | 5.76 | 2.52 | 1 | 9 |
| Years since bachelor's degree | 10253 | 6.61 | 7.18 | -2 | 51 |
| Graduate education objective | 10253 | 3.29 | 0.48 | 1 | 4 |
| GRE quantitative score | 10253 | 539 | 131 | 200 | 800 |

**Table 8**

***Correlations Among Level-One Variables for White Test-Takers—GRE Quantitative Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor ethnicity | 0.01 | 1.00 | | | | | | |
| 3. UGPA | 0.09* | -0.01 | 1.00 | | | | | |
| 4. Mother's highest level of education completed | 0.02 | -0.05* | 0.07* | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.04* | -0.06* | 0.06* | 0.57* | 1.00 | | | |
| 6. Years since bachelor's degree | -0.01 | 0.00 | -0.19* | -0.20* | -0.17* | 1.00 | | |
| 7. Graduate education objective | -0.05* | -0.01 | 0.14* | 0.07* | 0.07* | -0.04* | 1.00 | |
| 8. GRE quantitative score | -0.24* | -0.06* | 0.27* | 0.20* | 0.24* | 0.01 | -0.17* | 1.00 |

* $p \leq 0.05$.

When we partitioned the total variance in GRE quantitative scores using the fully unconditional model, within-center variance accounted for 95.4% of the score variance, while between-center variance accounted for 4.59% (see Table B3). Next, using the conditional model, we looked for significant predictors at the test-taker and test-center levels. For the within-center model, test score in a center j was associated with UGPA ($\beta_{ij}$ is a random coefficient), gender ($\beta$ = -65.53, $p < .01$), graduate objective ($\beta$ = 25.64, $p < .01$), and father's education ($\beta$ = 9.31, $p < .01$). In the between-center model, test-center mean and the UGPA slope were random coefficients. Test-center mean was associated with test-center size ($\beta$ = 6.82, $p < .01$), and the UGPA slope was moderated by median income ($\beta$ = 3.17, $p < .05$). (Table B4, displays these results)

After controlling for UGPA, we found that test-taker gender, graduate objective, and father's education were associated with GRE quantitative score. However, in contrast to the analyses of GRE verbal score, match with proctor ethnicity did not relate to GRE quantitative score. Median income moderated the relationship between UGPA and score. When median income was at its grand mean, the UGPA slope was 29.09. However, the UGPA slope increased to 32.26 when median income was $10,000 higher than the grand mean.

*African American test-takers—GRE verbal test.* Table 9 and Table 10 display descriptive statistics and correlations among test-taker-level variables for African American test-takers who completed the GRE verbal test, followed by test-center-level statistics in Table 11 and Table 12.

**Table 9**

***Level-One Descriptive Statistics for African American Test-Takers—GRE Verbal Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 1,077 | 0.72 | 0.45 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 1,077 | 0.16 | 0.36 | 0 (None match) | 1 (All match) |
| UGPA | 1,077 | 4.54 | 1.23 | 1 | 7 |
| Mother's highest level of education completed | 1,077 | 4.87 | 2.47 | 1 | 9 |
| Father's highest level of education completed | 1,077 | 4.43 | 2.53 | 1 | 9 |
| Years since bachelor's degree | 1,077 | 6.52 | 6.17 | -3 | 38 |
| Graduate education objective | 1,077 | 3.32 | 0.48 | 1 | 4 |
| GRE verbal score | 1,077 | 374 | 81 | 200 | 710 |

**Table 10**

***Correlations Among Level-One Variables for African American Test-Takers—GRE Verbal Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor ethnicity | 0.02 | 1.00 | | | | | | |
| 3. UGPA | 0.04 | -0.04 | 1.00 | | | | | |
| 4. Mother's highest level of education completed | 0.04 | 0.00 | 0.01 | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.06* | -0.03 | 0.00 | 0.59* | 1.00 | | | |
| 6. Years since bachelor's degree | 0.02 | 0.01 | -0.04 | -0.13* | -0.13* | 1.00 | | |
| 7. Graduate education objective | -0.06* | -0.01 | 0.11* | 0.15* | 0.13* | 0.01 | 1.00 | |
| 8. GRE verbal score | -0.01 | -0.06* | 0.19* | 0.17* | 0.15* | -0.03 | 0.19* | 1.00 |

* $p \leq 0.05$.

**Table 11**

***Level-Two Descriptive Statistics for African American Test-Takers***

| Variable | *N* of center | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Median income | 30 | 30,988.17 | 8,233.00 | 6,000 | 53,523 |
| Test-center size | 30 | 2.77 | 1.36 | 1 | 5 |
| Formal/professional | 30 | 28.87 | 6.99 | 6 | 43 |
| Warm/friendly | 30 | 30.97 | 8.59 | 6 | 46 |
| Disruptive | 30 | 17.60 | 5.00 | 6 | 30 |
| Activity level | 30 | 5.87 | 1.20 | 2 | 8 |

**Table 12**

***Correlations Among Level-Two Variables for African American Test-Takers***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Median income | 1.00 | | | | | |
| 2. Test-center size | 0.26 | 1.00 | | | | |
| 3. Formal/professional | 0.02 | 0.02 | 1.00 | | | |
| 4. Warm/friendly | 0.28 | 0.06 | 0.39* | 1.00 | | |
| 5. Disruptive | 0.05 | 0.35 | 0.36 | 0.23 | 1.00 | |
| 6. Activity level | 0.37 | 0.28 | 0.08 | 0.01 | 0.02 | 1.00 |

* $p \leq 0.05$.

In terms of the variance partitioning for GRE verbal scores, the between-center portion accounted for only 3.04% of the score variance, indicating that the majority of the variance in scores is attributable to test-taker-level variables rather than test-center-level variables (see Table B5). When we tested the conditional model we found GRE-verbal scores in center j to be related to UGPA ($\beta = 11.16$, $p < .01$), graduate objective ($\beta = 25.21$, $p < .01$), and mother's education ($\beta = 4.66$, $p < .01$). At the between-center level, we found that test-center mean was associated with test-center size ($\beta = 5.30$, $p < .01$). (Table B6, displays these results.) After controlling for UGPA, both graduate objective and mother's education were associated with GRE verbal score. In the test-center-level model, test-center size was significantly related to test-center mean verbal score.

*African American test-takers—GRE quantitative test.* When we repeated these analyses for African American test-takers who took the GRE quantitative test, the sample size was reduced by $N = 4$ test-takers. Table 13 and Table 14 present l test-taker-level descriptive statistics and correlations for African American test-takers who completed the GRE quantitative test.

**Table 13**

***Level-One Descriptive Statistics for African American Test-Takers—GRE Quantitative Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 1,073 | 0.72 | 0.45 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 1,073 | 0.16 | 0.36 | 0 (None match) | 1 (All match) |
| UGPA | 1,073 | 4.52 | 1.23 | 1 | 7 |
| Mother's highest level of education completed | 1,073 | 4.95 | 2.42 | 1 | 9 |
| Father's highest level of education completed | 1,073 | 4.45 | 2.52 | 1 | 9 |
| Years since bachelor's degree | 1,073 | 6.38 | 6.03 | -3 | 38 |
| Graduate education objective | 1,073 | 3.30 | 0.48 | 1 | 4 |
| GRE verbal score | 1,073 | 402 | 120 | 200 | 800 |

**Table 14**

***Correlations Among Level-One Variables for African American Test-Takers—GRE Quantitative Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor ethnicity | 0.03 | 1.00 | | | | | | |
| 3. UGPA | 0.04 | -0.04 | 1.00 | | | | | |
| 4. Mother's highest level of education completed | 0.16 | 0.02 | -0.02 | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.05 | 0.01 | 0.03 | 0.58* | 1.00 | | | |
| 6. Years since bachelor's degree | 0.02 | 0.01 | -0.06 | -0.16* | -0.13* | 1.00 | | |
| 7. Graduate education objective | -0.04 | 0.01 | 0.13* | 0.12* | 0.10* | 0.02 | 1.00 | |
| 8. GRE verbal score | -0.13* | -0.09* | 0.15* | 0.18* | 0.18* | -0.10* | 0.14* | 1.00 |

* $p \leq 0.05$.

Between-center variance accounted for 7.41% of the variance in GRE quantitative score, while within-center variance accounted for 92.6% (see Table B7). When we looked at the conditional model, we obtained the same set of significant predictors for this group's GRE quantitative test scores as we did for their verbal test scores. For African American test-takers, GRE quantitative score was also associated with UGPA ($\beta = 11.19$, $p < .01$), graduate objective ($\beta = 25.95$, $p < .01$), and mother's education ($\beta = 7.93$, $p < .01$). For the between-center model, test-center mean was again related to test-center size ($\beta = 9.66$, $p = .05$). (Table B8, displays these results.) The same set of predictors related to both GRE quantitative score and GRE verbal score. In contrast to a stereotype-threat hypothesis, match with proctor ethnicity was unrelated to test score. Also, unlike White test-takers, both verbal and quantitative test scores were related to mother's education rather than to father's education, and median income was unrelated to GRE scores.

*Hispanic test-takers—GRE verbal test.* Table 15 and Table 16 show descriptive statistics and correlations among test-taker-level variables for Hispanic test-takers who took the GRE-verbal test, while Table 17 and Table 18 report the test-center-level statistics.

**Table 15**

***Level-One Descriptive Statistics for Hispanic Test-Takers—GRE-Verbal Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 472 | 0.65 | 0.48 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 472 | 0.28 | 0.45 | 0 (None match) | 1 (All match) |
| UGPA | 472 | 4.89 | 1.12 | 2 | 7 |
| Mother's highest level of education completed | 472 | 3.79 | 2.50 | 1 | 9 |
| Father's highest level of education completed | 472 | 4.09 | 2.87 | 1 | 9 |
| Years since bachelor's degree | 472 | 5.69 | 5.57 | 1 | 40 |
| Graduate education objective | 472 | 3.25 | 0.48 | 1 | 4 |
| GRE verbal score | 472 | 399 | 95 | 200 | 700 |

**Table 16**

***Correlations Among Level-One Variables for Hispanic Test-Takers—GRE Verbal Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor ethnicity | 0.07 | 1.00 | | | | | | |
| 3. UGPA | 0.11* | -0.00 | 1.00 | | | | | |
| 4. Mother's highest level of education completed | -0.02 | -0.10* | 0.03 | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.01 | -0.09* | 0.09 | 0.63* | 1.00 | | | |
| 6. Years since bachelor's degree | 0.08 | 0.01 | 0.02 | -0.12* | -0.08 | 1.00 | | |
| 7. Graduate education objective | -0.06 | -0.15* | 0.06 | -0.03 | 0.00 | -0.06 | 1.00 | |
| 8. GRE verbal score | -0.10* | -0.23* | 0.11* | 0.21* | 0.25* | -0.03 | 0.10* | 1.00 |

* $p \leq 0.05$.

**Table 17**

***Level-Two Descriptive Statistics for Hispanic Test-Takers***

| Variable | *N* of center | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Median income | 17 | 40,820.06 | 8,233.00 | 25,157 | 64,710 |
| Test-center size | 17 | 3.41 | 1.06 | 1 | 5 |
| Formal\professional | 17 | 30.76 | 4.74 | 23 | 39 |
| Warm\friendly | 17 | 32.94 | 7.69 | 17 | 47 |
| Disruptive | 17 | 18.82 | 4.52 | 12 | 29 |
| Activity level | 17 | 6.00 | 0.94 | 5 | 8 |

**Table 18**

***Correlations Among Level-Two Variables for Hispanic Test-Takers***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Median income | 1.00 | | | | | |
| 2. Test-center size | 0.25 | 1.00 | | | | |
| 3. Formal\professional | 0.53* | 0.24 | 1.00 | | | |
| 4. Warm\friendly | 0.24 | 0.35 | 0.66* | 1.00 | | |
| 5. Disruptive | 0.19 | 0.02 | 0.50* | 0.29 | 1.00 | |
| 6. Activity level | 0.28 | 0.19 | 0.03 | 0.14 | -0.12 | 1.00 |

* $p \leq 0.05$.

When we partitioned the variance in GRE verbal test score into its between-center and within-center components, the between-center portion and within-center portion accounted for 6.59% and 93.41% of the score variance, respectively (see Table B9). When we tested the conditional model, GRE verbal score was related to UGPA ($\beta = 7.53$, $p < .05$), match with proctor ($\beta = -44.99$, $p < .01$), and father's education ($\beta = 7.30$, $p < .01$). At the test-center-level, no variables varied across test centers. (Table B10 displays these results.) Contrary to a stereotype-threat hypothesis, match with proctor ethnicity was negatively related to GRE verbal test score among Hispanic test-takers.

*Hispanic test-takers—GRE quantitative test.* When we repeated these analyses for Hispanic test-takers who took the GRE quantitative test, the sample size was again reduced by $N = 4$ test-takers. Table 19 and Table 20 present descriptive statistics and correlations for test-taker-level variables for Hispanic test-takers who completed the GRE quantitative test.

**Table 19**

***Level-One Descriptive Statistics for Hispanic Test-Takers—GRE Quantitative Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 468 | 0.66 | 0.47 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 468 | 0.28 | 0.45 | 0 (None match) | 1 (All match) |
| UGPA | 468 | 4.89 | 1.12 | 2 | 7 |
| Mother's highest level of education completed | 468 | 3.75 | 2.47 | 1 | 9 |
| Father's highest level of education completed | 468 | 3.85 | 2.76 | 1 | 9 |
| Years since bachelor's degree | 468 | 6.11 | 6.24 | 1 | 40 |
| Graduate education objective | 468 | 3.23 | 0.49 | 1 | 4 |
| GRE quantitative score | 468 | 458 | 134 | 200 | 780 |

**Table 20**

***Correlations Among Level-One Variables for Hispanic Test-Takers—GRE Quantitative Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor ethnicity | 0.08 | 1.00 | | | | | | |
| 3. UGPA | 0.10* | -0.01 | 1.00 | | | | | |
| 4. Mother's highest level of education completed | 0.02 | -0.10* | 0.02 | 1.00 | | | | |
| 5. Father's highest level of education completed | 0.04 | -0.09 | 0.05 | 0.61* | 1.00 | | | |
| 6. Years since bachelor's degree | 0.08 | 0.02 | -0.01 | -0.13* | -0.05 | 1.00 | | |
| 7. Graduate education objective | -0.08 | -0.15* | 0.10* | -0.05 | -0.03 | -0.14* | 1.00 | |
| 8. GRE quantitative score | -0.28* | -0.11* | 0.13* | 0.21* | 0.24* | -0.12* | 0.10* | 1.00 |

Between-center variance accounted for 6.01% of the variance in Hispanic test-takers' GRE quantitative scores, while within-center variance accounted for 93.99% (see Table B11). When we examined the conditional model, GRE quantitative score was associated with UGPA ($\beta = 18.56$, $p < .01$), gender ($\beta_{2j}$ is a random coefficient), and father's education ($\beta = 10.18$, $p < .01$). At the between-center level, the gender-GRE quantitative slope was negatively related to the level of warmth at the test center ($\beta = -4.29$, $p < .01$). (Table B12, displays these results.)

In contrast to the analysis of Hispanic test-takers who took the GRE verbal test, match with proctor was not related to GRE quantitative score for Hispanic test-takers. However, gender was significantly associated with GRE quantitative score, and the level of warmth at the test-center level moderated this relationship. When the level of warmth was at the grand mean, gender slope was –94.42. However, as the level of warmth in a test center increased by one unit, the gender slope decreased by 4.29 points. That is, female test-takers' GRE quantitative score may have decreased by an additional 4.29 points.

*Asian American test-takers—GRE verbal test.* Table 21 and Table 22 present descriptive statistics and correlations for Asian American test-takers. The sample and statistics did not differ for the GRE verbal and quantitative tests, so data are combined in these tables. Table 23 and Table 24 display test-center-level statistics for Asian American test-takers.

**Table 21**

***Level-One Descriptive Statistics for Asian American Test-Takers***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 516 | 0.55 | 0.50 | 0 (Male) | 1 (Female) |
| Match with proctor ethnicity | 516 | 0.00 | 0.00 | 0 (None match) | 1 (All match) |
| UGPA | 516 | 5.14 | 1.19 | 1 | 7 |
| Mother's highest level of education completed | 516 | 5.46 | 2.56 | 1 | 9 |
| Father's highest level of education completed | 516 | 6.10 | 2.60 | 1 | 9 |
| Years since bachelor's degree | 516 | 5.12 | 4.85 | 1 | 28 |
| Graduate education objective | 516 | 3.29 | 0.50 | 1 | 4 |
| GRE verbal score | 516 | 435 | 129 | 200 | 750 |
| GRE quantitative score | 516 | 633 | 133 | 250 | 800 |

**Table 22**

***Correlations Among Level-One Variables for Asian American Test-Takers***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. UGPA | 0.02 | 1.00 | | | | | | |
| 3. Mother's highest level of education completed | 0.00 | 0.11* | 1.00 | | | | | |
| 4. Father's highest level of education completed | 0.01 | 0.06 | 0.72* | 1.00 | | | | |
| 5. Years since bachelor's degree | -0.03 | -0.10* | -0.10* | -0.04 | 1.00 | | | |
| 6. Graduate education objective | 0.00 | 0.17* | 0.10* | 0.10* | 0.03 | 1.00 | | |
| 7. GRE verbal score | -0.01 | 0.23* | 0.22* | 0.22* | -0.08 | 0.20* | 1.00 | |
| 8. GRE quantitative score | -0.24* | 0.26* | 0.06 | 0.09* | -0.16* | 0.11* | .34* | 1.00 |

*Note*. The absence of centers staffed fully by Asian American proctors prevented us from examining the role of match with proctor ethnicity.

* $p \leq 0.05$.

**Table 23**

***Level-Two Descriptive Statistics for Asian American Test-Takers***

| Variable | *N* of center | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Median income | 20 | 43,473.85 | 13,876.68 | 27,662 | 79,970 |
| Test-center size | 20 | 3.35 | 1.31 | 1 | 5 |
| Formal\professional | 20 | 29.45 | 6.36 | 19 | 42 |
| Warm\friendly | 20 | 33.70 | 9.21 | 17 | 49 |
| Disruptive | 20 | 18.45 | 5.43 | 12 | 29 |
| Activity level | 20 | 6.45 | 1.10 | 4 | 8 |

**Table 24**

***Correlations Among Level-Two Variables for Asian American Test-Takers***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Median income | 1.00 | | | | | |
| 2. Test-center size | -0.06 | 1.00 | | | | |
| 3. Formal\professional | 0.33 | 0.09 | 1.00 | | | |
| 4. Warm\friendly | 0.22 | 0.19 | 0.36 | 1.00 | | |
| 5. Disruptive | -0.12 | 0.23 | 0.66* | 0.32 | 1.00 | |
| 6. Activity level | 0.22 | -0.08 | -0.04 | 0.03 | -0.15 | 1.00 |

* $p \leq 0.05$.

Between-center variance accounted for 7.26% of the variance in Asian American test-takers' GRE verbal scores, and within-center variance accounted for 92.74% (see Table B13). The absence of centers staffed fully by Asian American proctors prevented us from examining the role of match with proctor ethnicity; however, we continued with the analyses to explore the potential impact of other environmental factors. For Asian American test-takers, GRE verbal test score was associated with UGPA ($\beta = 21.45$, $p < .01$), graduate objective ($\beta_{2j}$ is a random coefficient), and father's education ($\beta_{3j}$ is a random coefficient). We found that test-center mean verbal score was moderated by formality at the test center ($\beta = 3.21$, $p < .05$). (Table B14 displays these results.)

After controlling for UGPA, both graduate objective and father's education were significantly related to GRE verbal score. At the test-center level, test-center formality was positively related to the test-center mean score. The effect of graduate objective and father's education varied across test centers, but neither were predicted by test-center-level variables. The correlations among the random effects indicate that the graduate objective slope tended to be negatively related to test-center mean score, but the father's education slope tended to be positively related to both the graduate objective

slope and the test-center mean. No test-center-level variables predicted these slopes.

*Asian American test-takers—GRE quantitative test.* Between-center variance accounted for 11.25% of the variance in GRE quantitative scores, while within-center variables accounted for 88.75% (see Table B14). Thus, test-center-level factors appear to have a larger influence on GRE quantitative scores for Asian American test-takers than for any other group across either test.

At the test-taker level, GRE quantitative score was related to UGPA ($\hat{\beta}_{1j}$ is a random coefficient), test-taker gender ($\beta = -64.18$, $p < .01$), and years since bachelor's degree ($\hat{\beta}_{3j}$ is a random coefficient). At the test-center level, test-center mean GRE quantitative score was moderated by median income ($\beta = 12.29$, $p < .05$), and the years since bachelor's degree slope was moderated by the activity level at the test center ($\beta = 3.69$, $p < .01$). (Table B16, displays these results.)

With each year since completing the bachelor's degree, GRE quantitative performance was likely to decrease by 5.17 points, on average. However, this negative relationship was moderated by the activity level of the test center. As activity level increased, it was likely to alleviate the negative impact of years since bachelor's degree on GRE quantitative score. At the test-center level, median income of the test-center was related to the test-center mean. As a test-center's median income increased by $10,000, the test-center mean should have increased by 12.29 points. While the UGPA slope varied over centers, no center-level variables predicted this relationship.

*Overview of effects.* In summary, Study 2 revealed several relationships between test-center variables and test performance. GRE verbal and quantitative test scores among White and Hispanic test-takers increased as the size of the test centers increased. Median income was related positively to test scores among White and Asian American test-takers. Finally, increased test-center activity level moderated the negative influence of the number of years since receiving a bachelor's degree on GRE quantitative score for Asian American test-takers.

Contrary to expectations, however, GRE quantitative scores among Hispanic female test-takers decreased as test-center warmth increased; and test-center formality was positively related to Asian American test-takers' GRE verbal scores. Also unexpectedly, match with proctor ethnicity did not relate to higher scores among African American and Hispanic test-takers. In fact, match with proctor ethnicity was negatively related to GRE verbal score among White and Hispanic test-takers. In order to highlight the significant effects noted for Study 2, Table 25 and Table 26 summarize all significant coefficients, effect sizes, and proportions of variance accounted for at the test-taker and test-center levels, respectively.

**Table 25**

***Significant Level-One Coefficients, Effect Sizes, and Proportions of Variance***

| | Coefficient (effect size) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | White test-takers | | African American test-takers | | Hispanic test-takers | | Asian American test-takers | |
| | GRE verbal | GRE quantitative | GRE verbal | GRE quantitative | GRE verbal | GRE quantitative | GRE verbal | GRE quantitative |
| Test-taker gender | -19.47 (-.20) | -65.53 (-.50) | | | | | | -.64.18 (-.48) |
| Match with proctor ethnicity | -16.22 (-.12) | | | | -44.99 (-.48) | | | |
| UGPA | 17.72 (.18) | | | 11.19 (.09) | 7.53 (.08) | 18.56 (.14) | 21.45 (.17) | |
| Mother's education | | | 4.66 (.06) | 7.93 (.07) | | | | |
| Father's education | 5.67 (.06) | 9.31 (.07) | | | 7.30 (.08) | 10.18 (.08) | | |
| Graduate objective | 25.20 (.25) | 25.64 (.20) | 25.21 (.31) | 25.95 (.22) | | | | |
| Proportion of variance | .10 | .18 | .08 | .33 | .03 | .17 | .17 | .10 |

**Table 26**

***Significant Level-Two Coefficients, Effect Sizes, and Proportions of Variance***

| | Coefficient (effect size) | | | | | |
|---|---|---|---|---|---|---|
| | GRE verbal test | | | GRE quantitative test | | |
| | White test-takers | | | | | |
| | $\hat{\beta}_{oj}$ Test-center Mean | | | $\hat{\beta}_{oj}$ Test-center mean | | $\hat{\beta}_{1j}$ UGPA slope |
| | 388.07 | | | 443.16 | | 29.09 |
| Median income | 7.30 (.07) | | | | | 3.17 (.02) |
| Test-center size | 4.51 (.05) | | | 6.82 (.05) | | |
| Variance | .52 | | | .09 | | .13 |
| | African American test-takers | | | | | |
| | $\hat{\beta}_{oj}$ Test-center mean | | | $\hat{\beta}_{oj}$ Test-center mean | | |
| | 266.83 | | | 266.83 | | |
| Test-center size | 5.30 (.07) | | | 5.30 (.07) | | |
| Variance | .39 | | | .39 | | |
| | Hispanic test-takers | | | | | |
| | No random coefficients modeled at the test-center level. | | | $\hat{\beta}_{oj}$ Test-center mean | | $\hat{\beta}_{2j}$ Gender slope |
| | | | | 484.60 | | -94.42 |
| Warm\friendly | | | | -4.29 (.03) | | |
| Variance | | | | .62 | | |
| | Asian American test-takers | | | | | |
| | $\hat{\beta}_{oj}$ Test-center mean | $\hat{\beta}_{2j}$ Graduate objective slope | $\hat{\beta}_{3j}$ Father's education slope | $\hat{\beta}_{oj}$ Test-center mean | $\hat{\beta}_{1j}$ UGPA slope | $\hat{\beta}_{3j}$ Bachelor's degree slope |
| | 429.49 | 39.10 | 8.19 | 631.92 | 29.08 | -5.17 |
| Formal\ professional | 3.21 (.02) | | | | | |
| Median income | | | | 12.29 (.09) | | |
| Activity level | | | | | | 3.69 (.03) |
| Variance | .27 | | | .11 | | .54 |

## Study 3: Proctor Gender

### *Method*

*Test-taker sample.* Study 3 sought to repeat the analyses completed for Study 2, except that the sample was limited to test-takers who either did or did not match the gender of the proctors at their respective test centers. For the analysis of the match between test-taker gender and proctor gender, the sample was limited to male ($N$ = 2,633) and female ($N$ = 4,891) test-takers who tested in centers staffed fully by male or female proctors so that all test-takers had either a 100% or 0% match on proctor gender.

*Design.* The procedure was identical to that followed for the assessment of the match between test-taker ethnicity and proctor ethnicity in Study 2, with the substitution of gender for ethnicity. Level-one factors were:

- test-taker gender
- match with proctor gender
- UGPA
- mother's highest level of education completed
- father's highest level of education completed
- years since bachelor's degree
- graduate education objective

Level-two factors were the test-center variables:

- median income
- test-center size
- formal\professional
- warm\friendly
- disruptive
- activity level

### *Results and Discussion*

Table 27 and Table 28 display descriptive statistics and correlations among test-taker-level variables for the GRE verbal test, while Table 29 and Table 30 present test-center-level statistics.

**Table 27**

***Level-One Descriptive Statistics for Test-Takers—GRE Verbal Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 7,524 | 0.65 | 0.48 | 0 (Male) | 1 (Female) |
| Match with proctor gender | 7,524 | 0.63 | 0.48 | 0 (None match) | 1 (All match) |
| UGPA | 7,524 | 5.23 | 1.20 | 1 | 7 |
| Mother's highest level of education completed | 7,524 | 5.30 | 2.34 | 1 | 9 |
| Father's highest level of education completed | 7,524 | 5.53 | 2.57 | 1 | 9 |
| Years since bachelor's degree | 7,524 | 6.64 | 7.04 | -1 | 45 |
| Graduate education objective | 7,524 | 3.28 | 0.48 | 1 | 4 |
| GRE verbal score | 7,524 | 455 | 103 | 200 | 800 |

**Table 28**

***Correlations Among Level-One Variables for Test-Takers—GRE Verbal Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor gender | 0.80* | 1.00 | | | | | | |
| 3. UGPA | 0.08* | 0.06* | 1.00 | | | | | |
| 4. Mother's highest level of education completed | -0.03* | -0.02* | 0.08* | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.05* | -0.04* | 0.08* | 0.58* | 1.00 | | | |
| 6. Years since bachelor's degree | -0.01 | 0.02 | -0.15* | -0.16* | -0.14* | 1.00 | | |
| 7. Graduate education objective | -0.05* | -0.05* | 0.12* | 0.08* | 0.05* | -0.02* | 1.00 | |
| 8. GRE verbal score | -0.08* | -0.06* | 0.28* | 0.20* | 0.21* | 0.03* | 0.15* | 1.00 |

* $p \leq 0.05$.

**Table 29**

***Level-Two Descriptive Statistics***

| Variable | *N* of center | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Median income | 85 | 34,509.19 | 9,236.48 | 20,973 | 63,187 |
| Test-center size | 85 | 2.40 | 1.21 | 1 | 5 |
| Formal\professional | 85 | 29.53 | 5.71 | 18 | 43 |
| Warm\friendly | 85 | 34.42 | 6.74 | 17 | 51 |
| Disruptive | 85 | 18.40 | 4.17 | 11 | 29 |
| Activity level | 85 | 5.62 | 1.23 | 1 | 8 |

**Table 30**

***Correlations Among Level-Two Variables***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1. Median income | 1.00 | | | | | |
| 2. Test-center size | 0.31* | 1.00 | | | | |
| 3. Formal\professional | 0.02 | 0.04 | 1.00 | | | |
| 4. Warm\friendly | -0.04 | 0.16 | 0.36* | 1.00 | | |
| 5. Disruptive | 0.29* | 0.06 | 0.22* | 0.16 | 1.00 | |
| 6. Activity level | 0.15 | 0.16 | 0.03 | 0.04 | -0.10 | 1.00 |

* $p \leq 0.05$.

Between-center variance accounted for 6.62% of the variance in GRE verbal scores, while within-center variance accounted for 93.4% (see Table B17). At the test-taker level, GRE verbal score was associated with UGPA ($\beta = 20.46$, $p < .01$), test-taker gender ($\beta = -17.66$, $p < .01$), graduate objective ($\hat{\beta}_{3j}$ is a random coefficient), father's education ($\hat{\beta}_{4j}$ is a random coefficient), and mother's education ($\beta = 3.97$, $p < .01$). At the test-center level, test-center mean verbal score was related to median income ($\beta = 7.60$, $p < .01$). (Table B18, displays these results.)

At the test-taker level, UGPA, test-taker gender, graduate objective, father's education, and mother's education were related to GRE verbal score. Match with proctor gender was not related to test score. At the test-center level, median income was positively related to test-center mean verbal score. The test-center mean was likely to increase by 7.60 points for each $10,000 increase in the median income of the test center. No other test-center-level predictors were

significant. Although the slopes for graduate objective and father's education varied across test centers, no test-center-level variables predicted these relationships.

The analyses were repeated for the sample of test-takers who took the GRE-Quantitative Test; however, the sample size was reduced by $N = 28$ test-takers due to their having opted to cancel test scores upon completion of the test. Table 31 and Table 32 display descriptive statistics and correlations among test-taker-level variables for the GRE quantitative test.

**Table 31**

***Level-One Descriptive Statistics for Test-Takers—GRE-Quantitative Test***

| Variable | Test-taker *N* | M | SD | Min. | Max. |
|---|---|---|---|---|---|
| Test-taker gender | 7,496 | 0.65 | 0.48 | 0 (Male) | 1 (Female) |
| Match with proctor gender | 7,496 | 0.63 | 0.48 | 0 (None match) | 1 (All match) |
| UGPA | 7,496 | 5.23 | 1.19 | 1 | 7 |
| Mother's highest level of education completed | 7,496 | 5.31 | 2.34 | 1 | 9 |
| Father's highest level of education completed | 7,496 | 5.56 | 2.58 | 1 | 9 |
| Years since bachelor's degree | 7,496 | 6.64 | 7.08 | -1 | 45 |
| Graduate education objective | 7,496 | 3.29 | 0.48 | 1 | 4 |
| GRE quantitative score | 7,496 | 517 | 138 | 200 | 800 |

**Table 32**

***Correlations Among Level-One Variables for Test-Takers—GRE-Quantitative Test***

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 1. Test-taker gender | 1.00 | | | | | | | |
| 2. Match with proctor gender | 0.81* | 1.00 | | | | | | |
| 3. UGPA | 0.08* | 0.07* | 1.00 | | | | | |
| 4. Mother's highest level of education completed | -0.02 | -0.02 | 0.08* | 1.00 | | | | |
| 5. Father's highest level of education completed | -0.03* | -0.02 | 0.08* | 0.57* | 1.00 | | | |
| 6. Years since bachelor's degree | -0.02 | -0.02 | -0.13* | -0.18* | -0.15* | 1.00 | | |
| 7. Graduate education objective | -0.04* | -0.03* | 0.13* | 0.08* | 0.06* | -0.02 | 1.00 | |
| 8. GRE quantitative score | -0.23* | -0.18* | 0.31* | 0.21* | 0.24* | -0.11* | 0.14* | 1.00 |

* $p \leq 0.05$.

Between-center variance accounted for 11.2% of the variance in GRE quantitative scores, while within-center variance accounted for 88.8% (see Table B19). The test-taker-level model revealed that GRE quantitative scores were associated with UGPA ($\hat{\beta}_{1j}$ is a random coefficient), test-taker gender ($\beta = -65.80$, $p < .01$), graduate objective ($\beta = 18.59$, $p < .01$), father's education ($\beta = 6.63$, $p < .01$), and mother's education ($\beta = 5.44$, $p < .01$)—the same variables associated with GRE verbal scores. At the test-center level, test-center mean quantitative score was related to median income ($\beta = 18.41$, $p < .05$) and activity level ($\beta = 7.17$, $p < .05$). The slope for UGPA was moderated by median income ($\beta = 6.04$, $p < .01$). (Table B20, displays these results.) As with GRE verbal scores, several test-taker-level variables were related to GRE quantitative score. However, contrary to a stereotype threat-hypothesis, match with proctor gender did not relate to quantitative test score.

*Overview of effects.* In Study 3, GRE verbal and quantitative test score increased as the test-center-level median income increased, and quantitative score increased with the activity level of the center. Also, the positive relationship between UGPA and quantitative test score increased as the median income of the test center increased. However, in contrast to our expectations, match with proctor gender did not relate to GRE score on either test. Table 33 and Table 34 highlight significant coefficients, effect sizes, and proportions of variance accounted for at the test-taker and test-center levels, respectively.

**Table 33**

***Significant Level-One Coefficients, Effect Sizes, and Proportions of Variance***

| | Coefficient (effect size) | |
|---|---|---|
| | GRE verbal | GRE quantitative |
| Test-taker gender | -17.66 (-.17) | -65.80 (-.48) |
| Match with proctor gender | | |
| UGPA | 20.43 (.20) | |
| Mother's highest level of education completed | | 5.44 (.04) |
| Father's highest level of education completed | | 6.63 (.05) |
| Years since bachelor's degree | | |
| Graduate education objective | | 18.59 (.13) |
| Proportion of variance | .12 | .19 |

**Table 34**

***Significant Level-Two Coefficients, Effect Sizes, and Proportions of Variance***

| | Coefficient (effect size) | | | | |
|---|---|---|---|---|---|
| | GRE verbal test | | | GRE quantitative test | |
| | $\hat{\beta}_{oj}$ Test-center mean | $\hat{\beta}_{3j}$ Graduate objective slope | $\hat{\beta}_{4j}$ Father's education slope | $\hat{\beta}_{oj}$ Test-center mean | $\hat{\beta}_{1j}$ UGPA slope |
| | 348.38 | 22.19 | 4.41 | 432.05 | 32.15 |
| Median income | 7.60 (.07) | | | 18.41 (.13) | 6.04 (.04) |
| Activity level | | | | 7.17 (.05) | |
| Variance | .07 | | | .32 | .24 |

## General Discussion

The purpose of this research was to explore whether operational testing environments may inadvertently activate stereotype threat, which in turn, may lead to GRE General Test performance decrements in ways that are consistent with prior experimental studies. Our primary interest was in learning whether previous experimental evidence of a relationship between test performance and proctor gender (Marx & Roman, 2002) and ethnicity (Walters et al., 2003) would generalize to operational settings. The current study extended previous examinations of stereotype threat by sampling a large number of test-takers from the actual GRE testing population and by measuring test-taker and test-center level predictors of test scores.

We expected test-takers to perform best when test-taker gender or ethnicity matched proctor gender or ethnicity—ostensibly alleviating the disruption associated with stereotype threat. However, while we found some evidence for test-center-environment effects on test scores by using multilevel analyses, we found no evidence of stereotype threat as it has been demonstrated in previous work. That is, test-taker match with the ethnicity of the proctors in test centers did not relate to an increase in test score for any group of test-takers, including African American and Hispanic test-takers. Also, test-taker match with the gender of the proctors at each center was unrelated to the relationship between gender and test score.

These findings were inconsistent with what we expected based on previous findings. In fact, Hispanic test-takers actually performed better on the GRE verbal Test when they were paired with non-Hispanic proctors rather than Hispanic proctors. We can think of several general explanations for these findings. With regard to proctor ethnicity, we suspect that the impact of

others of similar gender or ethnicity in the environment may depend on one's "typical environment." That is, if exposure to others who share one's ethnicity is commonplace, the encounter may not relate to attentional processes or to performance. In the experimental investigation of proctor ethnicity (Walters et al., 2003), participants were students from a single university and thus had a similar level of exposure to a predominantly White group of faculty and administrators.

With regard to proctor gender, the mere presence of female proctors may not be enough to alleviate the attentional disruption associated with stereotype threat for female test-takers. Marx and Roman (2002) found that women performed better on quantitative tests only when the female proctor was considered an expert in the quantitative domain. Therefore, their findings may be more applicable to classroom settings, where relationships are more familiar, than to the operational, standardized testing domain.

We can only speculate about why the presence of non-Hispanic proctors was related to higher verbal test scores among Hispanic test-takers. The same pattern emerged for White test-takers—verbal scores were higher in test centers staffed by non-White proctors. Perhaps the homogeny associated with centers staffed by members of a single ethnicity group indicates a specific geographic region. However, variables related to median income and family education-level in our analyses should have accounted for any important regional differences.

Test-center size and activity level did relate to performance for some test-takers. Test-center size was positively related to both verbal and quantitative test score among White and African American test-takers. For instance, African American test-takers' quantitative scores increased by 10 points with each increase in test-center-size. In our analysis of test-taker match with proctor gender, we found that male and female test-takers' scores on the quantitative test increased by seven points with each level of activity at the test center. Also, among Asian American test-takers, an increase in test-center activity level reduced the negative impact of number of years since receiving a bachelor's degree on quantitative performance.

It is feasible that test-center size and activity level relate to performance by way of their influence on stereotype activation. Specifically, we suspect that large and active centers might create a sense of anonymity for test-takers, which in turn, may reduce the likelihood of a stereotypical evaluation. Future research will need to investigate more directly the links among test-center size, activity level, and stereotype activation, as well as any alternative variable for

which these factors may serve as proxies.

Our exploratory social-atmosphere variables—warm/friendly and formal/professional—related to test score in an unexpected way. We expected warm/friendly centers, as opposed to formal/professional centers, to be less threatening and thus less likely to evoke negative self-relevant stereotypes for test-takers. However, for Hispanic test-takers, the reported level of warmth at a test-center moderated the relationship between gender and quantitative score; as the level of warmth increased by one unit, female test-takers' quantitative scores decreased by 4 points. Among Asian-American test-takers, verbal scores increased by three points with each level of center formality. These findings suggest that test-takers may differ in the type of social atmosphere that is most comfortable for them. Given the variability on these atmosphere dimensions across tests centers, future research should explore this possibility in more depth.

While not the primary focus of this research, several notable patterns emerged in our examination of the test-taker-level variables that accounted for the majority of variability in test scores. For instance, test-taker gender was unrelated to test score only among African American test-takers—a trend found in previous analyses of test scores (Coley, 2001). Also, father's highest level of education related to performance for White, Hispanic, and Asian American test-takers, while mother's highest level of education related to test score for African American test-takers. Finally, test-takers' educational objectives related positively to GRE test performance for all groups except Hispanic test-takers. These findings highlight the importance of GRE Background Information Questionnaire data in future investigations of group differences and test performance.

Before discussing the practical implications of our findings, it is important to note the limitations in our study. The most notable limitation was our inability to account for the full range of variables that may have contributed to the test-center atmosphere. For instance, we were unable to account for the characteristics of other test-takers at each test center. Research has shown that the gender of other test-takers can induce stereotype threat among female test-takers (Inzlicht & Ben-Zeev, 2000).

Furthermore, because we did not know exactly with whom individual test-takers interacted, we limited our analyses to test-takers who tested in centers where all proctors either matched or did not match their gender or ethnicity. In doing so, a significant amount of data was discarded, and such heavy selection bias may have affected the results in unforeseen ways. It is

possible that a more precise account of each test-taker's experience may have revealed results more similar to that of laboratory settings. Another limitation was our use of descriptions by supervising proctors at each center to create our test-center-level variables. These descriptions were clearly susceptible to a variety of perceptual biases.

### *Practical Implications*

This research addresses several issues that are relevant to the administration of the GRE General Test at institutional and Prometric Testing Centers. First, our analyses did not reveal the patterns that one would expect given prior stereotype-threat findings. This lack of findings may be a function of limitations in our study, or alternatively, an indicator that the gender and ethnicity of proctors do not induce stereotype threat within operational settings. It is possible that stereotype-threat work is limited to low-stakes, laboratory environments where attitudinal factors may be salient because nothing else is important; but when transferred to a high-stakes setting, the pressure of performing well overcomes any attitudinal effects.

However, while additional research is clearly needed, our findings do suggest that test performance increased somewhat with the size and activity level of testing centers. We can suggest some strategies for addressing these variables. While the size of existing test centers is relatively unchangeable, activity level can be increased by scheduling more test-takers during a single session. Test-takers simply may not perform optimally when they feel that a spotlight is on them. If our reasoning about the role of evaluative pressure and stereotype activation is correct, proctors may need to remain attentive to implicit messages or behaviors that might induce additional stress in test-takers. Proctors may be able to use techniques to reduce stereotype threat for all test-takers. One such successful approach recommended in a review involves conveying positive expectations to test-takers (Aronson, Quinn, & Spencer, 1998). Importantly, this work is exploratory and suggests neither causal relationships nor simple solutions for those engaged in educational and psychological testing. We hope these results provide some direction for future research.

## References


Aronson, J., Quinn, D. M., & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim, & C. Stangor (Eds.), *Prejudice: The target's perspective* (pp. 83-103). San Diego, CA: Academic Press.

Aronson, J., & Salinas, M. F. (1997, April). *Stereotype threat: Is low performance the price of self-esteem for Mexican Americans?* Paper presented at the Western Psychological Association Conference, Seattle, WA.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245-276.

Coley, R. J. (2001). *Differences in the gender gap: Comparisons across racial/ethnic groups in education and work* (Policy Information Center Report). Princeton, NJ: ETS.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem solving deficits in the presence of males. *Psychological Science, 11,* 365-370.

Marx, D. M., & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin, 28,* 1,183-1,193.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S., Bryk, A., Cheong, Y. F., & Congdon, R. (2000). *HLM5: Hierarchical linear and nonlinear modeling* [Computer software]. Lincolnwood, IL: Scientific Software.

Schmader, T., & Johns. M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology, 85,* 440-452.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10,* 80-83.

Spencer, S., Steele, C., & Quinn, D. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4-28.

Steele, C., & Aronson, J. (1995). Stereotype threat and the intellectual 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. 98-01; ETS RR 98-05). New York: College Board.

Stricker, L. J., & Ward, W. C. (1998). *Inquiring about examinees' ethnicity and sex: Effects on computerized placement tests performance* (College Board Rep. No 98-02; ETS RR 98-09). Princeton, NJ: ETS.

Walters, A. M., Shepperd, J. A., & Brown, L. M. (2003). *The effect of test administrator ethnicity on test performance.* Manuscript submitted for publication.

## Notes

[1] During interviews with proctors, researchers took care not to imply that they were conducting evaluations of the test centers—particularly on dimensions such as contract compliance or test-center quality. Rather, they explained that they were interested in conducting observations in an effort to better understand how the centers operate.

[2] During the visits, investigators learned that centers varied very little on dimensions such as physical environment, materials, and protocol. Thus, the survey excluded these topics.

[3] Test centers in Canada and Puerto Rico were excluded from the study. Given our interest in studying the influence of cultural stereotypes, we chose to exclude centers that were not in the continental United States because we assumed that stereotypes would hold a different relevance and meaning for inhabitants of those areas.

[4] After summing the items, score ranges were 1 to 55 for warm/friendly (11 items), 1 to 45 for formal/professional (9 items), 1 to 30 for disruptive (6 items), and 1 to 10 for activity level (2 items).

[5] Size was coded on a 5-point scale using anchors at 1 (3 to 6 testing stations), 2 (7 to 10 testing stations), 3 (11 to 14 testing stations), 4 (15 to 16 testing stations), and 5 (17 or more testing stations).

[6] We do not know the number of test-takers who tested in a center located within a zip code that differed from their home zip code, but we suspect it to be fairly small. Geographic Information Systems technology is used to strategically locate test centers close to where test-takers live.

[7] For 16 test centers, the sum of proctors by gender differed from the sum of proctors by ethnicity. This raised a question as to whether additional proctors were present but not accounted for in our analyses. Thus, we omitted all test centers from the analysis for which the numbers of proctors differed by more than 1 ($N = 8$).

[8] Undergraduate grade point average was coded as 1 (D or lower), 2 (C-), 3 (C), 4 (B-), 5 (B), 6 (A-), or 7 (A).

[9] Highest level of parental education was measured using a 9-point scale: 1 (grade school or less), 2 (some high school), 3 (high school diploma or equivalent), 4 (business or trade school), 5 (some college), 6 (associate degree), 7 (bachelor's degree), 8 (some graduate or professional school), and 9 (graduate or professional degree).

[10] We coded all response options to the question "What is your eventual education objective?" as

1 (not currently planning graduate study), 2 (nondegree graduate study), 3 (master's or intermediate/specialist degree), 4 (doctorate/postdoctoral study).

[11] Mother's and father's highest levels of education completed and graduate education objective were transformed from categorical variables to continuous variables. However, all other variables were left in their original metric to enable future comparisons with background information collected using the GRE Background Information Questionnaire.

[12] When estimating the parameters, Empirical Bayes Estimates were computed for randomly varying level-one coefficients, $\beta$. Generalized least squares estimates were computed for fixed coefficients $\gamma$, and restricted maximum likelihood estimates were computed for variance/covariance components at level one and level two.

## Appendix A

## Test-Center Survey

**Important information: This survey will in no way be used to evaluate your center. Rather we are simply interested in learning about the atmosphere and social environment of testing centers. Your responses will remain strictly confidential, and your participation is completely voluntary. Further, there are no right or wrong answers to the items on the survey. We simply wish to hear your views.**
**In order to ensure that the study is valid, please provide your honest responses to the following questions.**

1. Please indicate your center number. *(If your center has more than one center number, i.e. a mega center, just list one.)* _____
2. What is the total number of workstations at your center site? _____

*Considering all of the Test Center Administrators (TCAs) at your center, please record the approximate number who can be described by the following categories:*

3. Male _____ Female _____

4. 18-24 years _____ 25-30 years _____ 31-40 years _____ Over 40 years _____

5. American Indian or Alaskan Native _____ Puerto Rican _____

Black or African American _____ Other Hispanic or Latin American _____

Mexican, Mexican American, or Chicano _____ White (non-Hispanic) _____

Asian, Asian American, or Pacific Islander _____ Other _____

*Please respond to the following questions by marking the appropriate circle.*

6. Which setting best describes your center's locale?

O Suburban O Urban O Rural

7. Is your center located in a College or University?

O Yes O No

8. Does your center offer services other than testing?

O Yes O No

9. Please describe the level of activity at your center on a typical day.

O Not at all busy O Not very busy O Somewhat busy O Very busy O Extremely busy

10. How would you describe the typical demeanor of students who arrive to take the GRE?

O Not at all nervous O Slightly nervous O Somewhat nervous O Very nervous O Extremely nervous

11. How occupied is your center on a typical day?

o Empty o Mostly empty o Half full o Mostly full o Completely full

12. Can arriving test-takers view the check-in area from the waiting area?

o Yes o No

13. During one GRE testing session, how many times does the testing room door typically open?

o Never o 1-3 times o 4-6 times o 7-9 times o 10+ times

14. On average, how much time do TCAs interact with test-takers?

o 5 minutes or less o 6-10 minutes o 11-20 minutes o 21-30 minutes o More than 30 minutes

15. In your opinion, to what extent do testing sessions vary from one time to the next (i.e., types of interactions with test-takers, questions, time to test completion, etc.)?

o Not at all o Slightly o Some o Quite a bit o Very much

16. In your opinion, please rate the extent to which the following behaviors occur at your center.

| | Never | Seldom | Sometimes | Often | Always |
|---|---|---|---|---|---|
| a. Test-takers are aware that other test-takers are taking different tests. | o | o | o | o | o |
| b. TCAs try to calm down nervous test-takers. | o | o | o | o | o |
| c. During breaks, test-takers interact with other test-takers. | o | o | o | o | o |
| d. During breaks, test-takers interact with TCAs | o | o | o | o | o |
| e. Test-takers who arrive on time, wait for more than 10 minutes to begin their tests. | o | o | o | o | o |
| f. Test-takers comment on having their picture taken at check-in. | o | o | o | o | o |

17. In your opinion, to what extent do GRE students seem to notice the following aspects of their environment?

| | Never notice | Seldom notice | Sometimes notice | Often notice | Always notice |
|---|---|---|---|---|---|
| a. Other test-takers entering and getting started. | O | O | O | O | O |
| b. Other test-takers during the test. | O | O | O | O | O |
| c. Different check-in procedures for fellow test-takers taking different tests. | O | O | O | O | O |
| d. TCA activity behind the observation window | O | O | O | O | O |
| e. Surveillance equipment (i.e., cameras and mirrors) | O | O | O | O | O |
| f. Door to testing area opening and closing. | O | O | O | O | O |

18. In your opinion, please rate the extent to which the following characteristics describe the typical atmosphere at your center. We realize that characteristics, which may sound like opposites, will actually both apply at different points in the day.

| | Not at all characteristic | Slightly characteristic | Somewhat characteristic | Very characteristic | Extremely characteristic |
|---|---|---|---|---|---|
| a. Relaxed | O | O | O | O | O |
| b. Somber | O | O | O | O | O |
| c. Professional | O | O | O | O | O |
| d. Serious | O | O | O | O | O |
| e. Intimidating | O | O | O | O | O |
| f. Hectic | O | O | O | O | O |
| g. Fun | O | O | O | O | O |
| h. Youthful | O | O | O | O | O |
| i. Lively | O | O | O | O | O |
| j. Formal | O | O | O | O | O |
| k. Critical | O | O | O | O | O |
| l. Mellow | O | O | O | O | O |
| m. Warm/cozy | O | O | O | O | O |
| n. Tense | O | O | O | O | O |
| o. Library-like | O | O | O | O | O |
| p. Collegial | O | O | O | O | O |
| q. Other | O | O | O | O | O |

19. How important is it for a TCA in your center to have the following traits?

| | Not at all important | Slightly important | Somewhat important | Very important | Extremely important |
|---|---|---|---|---|---|
| a. Strict | O | O | O | O | O |
| b. Efficient | O | O | O | O | O |
| c. Businesslike | O | O | O | O | O |
| d. Serious | O | O | O | O | O |
| e. Mellow | O | O | O | O | O |
| f. Compassionate | O | O | O | O | O |
| g. Intimidating | O | O | O | O | O |
| h. Fun | O | O | O | O | O |
| i. Professional | O | O | O | O | O |
| j. Youthful | O | O | O | O | O |
| k. Formal | O | O | O | O | O |
| l. Relaxed | O | O | O | O | O |
| m. Nurturing | O | O | O | O | O |
| n. Open-minded | O | O | O | O | O |
| o. Other | O | O | O | O | O |

20. To what degree do you believe that the TCAs in your center possess the traits that you described as extremely important?

O Not at all O A little O Some O Quite a bit O Very much

*Thank you for your time. Please return this survey in the enclosed envelope by September 19.*

## Appendix B

## HLM Tables

**Table B1**

***Fully Unconditional Model for White Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 476.79 | 194.88 | 130 | 0.00 |
| *Random effect* | Variance component (% at level)[a] | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 620.16 (6.33) | 130 | 888.04 | 0.00 |
| Within-center $r_{ij}$ | 9,169.64 (93.67) | | | |

[a] Proportion of variance at each level [e.g., 620.16 / (620.16 + 9169.64) = .0633].

**Table B2**

***Conditional Model for White Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 388.07 | 39.43 | 128 | 0.00 |
| Median income | 7.30 | 3.478 | 128 | 0.00 |
| Test-center size | 4.51 | 2.90 | 128 | 0.00 |
| Model for "UGPA" slope | | | | |
| Intercept | 17.72 | 22.40 | 10,324 | 0.00 |
| Model for "Test-taker gender" slope | | | | |
| Intercept | -19.47 | -10.27 | 10,324 | 0.00 |
| Model for "Match with proctor gender" slope | | | | |
| Intercept | -16.11 | -2.16 | 10,324 | 0.03 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 25.20 | 1.88 | 10,324 | 0.00 |
| Model for "Father's education" slope | | | | |
| Intercept | 5.67 | 15.63 | 10,324 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 300.12 | 128 | 535.03 | 0.00 |
| Within-center $r_{ij}$ | 8,253.92 | | | |

*Note.* $\hat{Y}_{ij}=\beta_{oj}$ + 17.72 (UGPA) - 19.47 (Gender) - 16.11 (Match with proctor) + 25.20 (Graduate objective) + 5.67 (Father's education). $\hat{\beta}_{oj}$=388.07 + 7.3 (Median income) + 4.51 (Test-center size). Here, the metric of median income is $10,000.

**Table B3**

***Fully Unconditional Model for White Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 539.97 | 153.35 | 130 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 1,333.81 (4.59) | 130 | 979.40 | 0.00 |
| Within-center $r_{ij}$ | 15,740.60 (95.41) | | | |

**Table B4**

***Conditional Model for White Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 443.16 | 50.19 | 129 | 0.00 |
| Test-center size | 6.82 | 3.12 | 129 | 0.00 |
| Model for "UGPA" slope | | | | |
| Intercept | 29.09 | 1.22 | 129 | 0.00 |
| Median income | 3.17 | 2.26 | 129 | 0.02 |
| Model for "Test-taker gender" slope | | | | |
| Intercept | -65.53 | 10.74 | 10,246 | 0.00 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 25.64 | 10.74 | 10,246 | 0.00 |
| Model for "Father's education" slope | | | | |
| Intercept | 9.31 | 20.37 | 10,246 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center | | | | |
| Test-center mean $u_{oj}$ | 713.81 | 129 | 689.85 | 0.00 |
| UGPA slope $u_{1j}$ | 46.16 | 129 | 187.68 | 0.00 |
| Within-center $r_{ij}$ | 12,924.28 | | | |

*Note.* $\hat{Y}_{ij}=\hat{\beta}_{oj}+\hat{\beta}_{1j}$ (UGPA) - 65.53 (Gender) + 25.64 (Graduate objective) + 9.31 (Father's education). Test-center mean: $\hat{\beta}_{oj}$= 443.16 + 6.82 (Center size). UGPA slope: $\hat{\beta}_{1j}$ = 29.09 + 3.17 (Median income). The metric of median income is $10,000 here.

**Table B5**

***Fully Unconditional Model for African American Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 372.87 | 100.51 | 29 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 200.53 (3.04) | 130 | 888.04 | 0.00 |
| Within-center $r_{ij}$ | 6,389.02 (96.96) | | | |

**Table B6**

***Conditional Model for African American Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 266.83 | 15.74 | 28 | 0.00 |
| Test-center size | 5.30 | 2.09 | 28 | 0.05 |
| Model for "UGPA" slope | | | | |
| Intercept | 11.16 | 5.84 | 1,072 | 0.00 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 25.21 | 5.03 | 1,072 | 0.00 |
| Model for "Mother's education" slope | | | | |
| Intercept | 4.66 | 4.83 | 1,072 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 121.67 | 28 | 48.19 | 0.01 |
| Within-center $r_{ij}$ | 5,878.29 | | | |

*Note.* $\hat{Y}_{ij} = \hat{\beta}_{oj}$ + 11.16 (UGPA) + 25.21 (Graduate objective) + 4.66 (Mother's education). Test-center mean: $\hat{\beta}_{oj} = 266.83 + 5.30$ (Center size).

**Table B7**

***Fully Unconditional Model for African American Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 399.50 | 55.44 | 29 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 1,077.48 (7.41) | 29 | 112.22 | 0.00 |
| Within-center $r_{ij}$ | 13,463.87 (92.59) | | | |

**Table B8**

***Conditional Model for African American Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 274.62 | 10.71 | 28 | 0.00 |
| Test-center size | 9.66 | 2.02 | 28 | 0.05 |
| Model for "UGPA" slope | | | | |
| Intercept | 11.19 | 3.92 | 1,068 | 0.00 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 25.95 | 3.49 | 1,068 | 0.00 |
| Model for "Mother's education" slope | | | | |
| Intercept | 7.93 | 5.44 | 1,068 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 720.49 | 28 | 82.89 | 0.00 |
| Within-center $r_{ij}$ | 12,747.31 | | | |

*Note.* $\hat{Y}_{ij} = \hat{\beta}_{oj} + 11.19$ (UGPA) + 25.95 (Graduate objective) + 7.93 (Mother's education). $\hat{\beta}_{oj} = 274.62 + 9.66$ (Test-center size).

**Table B9**

***Fully Unconditional Model for Hispanic Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 410.39 | 53.80 | 16 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 584.12 (6.59) | 16 | 61.89 | 0.00 |
| Within-center $r_{ij}$ | 8,277.16 (93.41) | | | |

**Table B10**

***Conditional Model for Hispanic Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 381.88 | 48.78 | 468 | 0.00 |
| Model for "UGPA" slope | | | | |
| Intercept | 7.53 | 2.03 | 468 | 0.04 |
| Model for "Match with proctor gender" slope | | | | |
| Intercept | -44.99 | -4.87 | 468 | 0.00 |
| Model for "Father's education" slope | | | | |
| Intercept | 7.30 | 5.05 | 468 | 0.00 |
| *Random effect* | Variance component | | | |
| Within-center $r_{ij}$ | 8,000.57 | | | |

*Note.* $\hat{Y}_{ij} = 381.88 + 7.53$ (UGPA) - 44.99 (Match with proctor) + 7.30 (Father's education).

**Table B11**

***Fully Unconditional Model for Hispanic Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 465.55 | 43.88 | 16 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 1,081.90 (6.01) | 16 | 44.08 | 0.00 |
| Within-center $r_{ij}$ | 16,931.60 (93.99) | | | |

**Table B12**

***Conditional Model for Hispanic Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 484.66 | 33.16 | 16 | 0.00 |
| Model for "UGPA" slope | | | | |
| Intercept | 18.56 | 3.71 | 463 | 0.00 |
| Model for "Test-taker gender" slope | | | | |
| Intercept | -94.42 | -7.95 | 463 | 0.00 |
| Warm\friendly | -4.29 | -3.28 | 463 | 0.00 |
| Model for "Father's education" slope | | | | |
| Intercept | 10.18 | 4.98 | 463 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 861.33 | 16 | 40.77 | 0.00 |
| Within-center $r_{ij}$ | 14,042.62 | | | |

*Note.* $\hat{Y}_{ij} = \hat{\beta}_{oj} + 18.56$ (UGPA) + $\hat{\beta}_{2j}$ (Gender) + 10.18 (Father's education). $\hat{\beta}_{oj} = 484.6$. $\hat{\beta}_{2j} = -94.42 - 4.29$ (Warm\friendly).

**Table B13**

***Fully Unconditional Model for Asian American Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 429.94 | 43.64 | 19 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 1,241.41 (7.26) | 19 | 53.92 | 0.00 |
| Within-center $r_{ij}$ | 15,854.66 (92.74) | | | |

**Table B14**

***Conditional Model for Asian American Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 429.49 | 49.53 | 18 | 0.00 |
| Formal/professional | 3.21 | 2.38 | 18 | 0.03 |
| Model for "UGPA" slope | | | | |
| Intercept | 21.45 | 4.73 | 511 | 0.00 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 39.10 | 2.51 | 19 | 0.02 |
| Model for "Father's education" slope | | | | |
| Intercept | 8.19 | 2.99 | 19 | 0.01 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center | | | | |
| Test-center mean $u_{oj}$ | 910.81 | 18 | 45.99 | 0.00 |
| Graduate objective slope $u_{2j}$ | 2,380.11 | 19 | 33.95 | 0.02 |
| Father's education slope $u_{3j}$ | 63.77 | 19 | 30.56 | 0.05 |
| Within-center $r_{ij}$ | 13,154.39 | | | |
| *Correlation between random effects* | | 1 | 2 | 3 |
| 1. Test-center mean | | 1.00 | | |
| 2. Graduate objective slope | | -.19 | 1.00 | |
| 3. Father's education slope | | .70 | .41 | 1.00 |

*Note.* $\hat{Y}_{ij} = \hat{\beta}_{oj} + 21.45$ (UGPA) $+ \hat{\beta}_{2j}$ (Graduate objective) $+ \hat{\beta}_{3j}$ (Father's education). $\hat{\beta}_{oj} = 429.49 + 3.21$ (Formal/professional). $\hat{\beta}_{2j} = 39.10$. $\hat{\beta}_{3j} = 8.19$.

**Table B15**

***Fully Unconditional Model for Asian American Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 632.04 | 53.91 | 19 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 2,031.25 (11.25) | 19 | 72.61 | 0.00 |
| Within-center $r_{ij}$ | 16,023.63 (88.75) | | | |

**Table B16**

***Conditional Model for Asian American Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 631.92 | 117.72 | 510 | 0.00 |
| Median income | 12.29 | 2.65 | 510 | 0.04 |
| Model for "UGPA" slope | | | | |
| Intercept | 29.08 | 5.02 | 19 | 0.00 |
| Model for "Test-taker gender" slope | | | | |
| Intercept | -64.18 | -5.98 | 510 | 0.00 |
| Model for "Years since bachelor's degree" slope | | | | |
| Intercept | -5.17 | -4.29 | 510 | 0.00 |
| Activity level | 3.69 | 3.33 | 510 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center | | | | |
| UGPA slope $u_{1j}$ | 213.18 | 19 | 30.51 | 0.05 |
| Within-center $r_{ij}$ | 14,364.81 | | | |

*Note.* $\hat{Y}_{ij} = \hat{\beta}_{oj} + \hat{\beta}_{1j}$ (UGPA) - 64.18 (Gender) + $\hat{\beta}_{3j}$ (Years since bachelor's degree). $\hat{\beta}_{oj}$ = 631.92 + 12.29 (Median income). $\hat{\beta}_{1j}$ = 29.08. $\hat{\beta}_{3j}$ = -5.17 + 3.69 (Activity level).

**Table B17**

***Fully Unconditional Model for Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 455.23 | 141.77 | 84 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 711.09 (6.62) | 84 | 595.86 | 0.00 |
| Within-center $r_{ij}$ | 10,024.69 (93.38) | | | |

**Table B18**

***Conditional Model for Test-Takers—GRE Verbal Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 348.38 | 34.39 | 83 | 0.00 |
| Median income | 7.60 | 2.91 | 83 | 0.00 |
| Model for "UGPA" slope | | | | |
| Intercept | 20.46 | 21.89 | 7517 | 0.00 |
| Model for "Test-taker gender" slope | | | | |
| Intercept | -17.66 | -7.64 | 7517 | 0.00 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 22.19 | 8.07 | 84 | 0.00 |
| Model for "Father's education" slope | | | | |
| Intercept | 4.41 | 7.49 | 84 | 0.00 |
| Model for "Mother's education" slope | | | | |
| Intercept | 3.97 | 6.91 | 7517 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center | | | | |
| Intercept $u_{oj}$ | 2,357.18 | 83 | 116.86 | 0.00 |
| Graduate objective slope $u_{3j}$ | 143.07 | 84 | 108.45 | 0.04 |
| Father's education slope $u_{4j}$ | 4.55 | 84 | 109.46 | 0.03 |
| Within-center $r_{ij}$ | 8,791.79 | | | |

*Note.* $\hat{Y}_{ij}=\hat{\beta}_{oj}$ + 20.46 (UGPA) - 17.66 (Gender) + $\hat{\beta}_{3j}$ (Graduate objective) + $\hat{\beta}_{4j}$ (Father's education) + 3.97 (Mother's education). $\hat{\beta}_{oj}$ = 348.38 + 7.60 (Median income). $\hat{\beta}_{3j}$ = 22.19. $\hat{\beta}_{4j}$ = 4.41.

**Table B19**

***Fully Unconditional Model for Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Average test-center mean $\gamma_{00}$ | 516.92 | 97.13 | 84 | 0.00 |
| *Random effect* | Variance component (% at level) | df | $\chi^2$ | p-value |
| Between-center $u_{oj}$ | 2,123.37 (11.20) | 84 | 1,017.29 | 0.00 |
| Within-center $r_{ij}$ | 16,837.02 (88.80) | | | |

**Table B20**

***Conditional Model for Test-Takers—GRE Quantitative Test***

| *Fixed effect* | Coefficient | t ratio | df | p-value |
|---|---|---|---|---|
| Model for "Test-center mean" | | | | |
| Intercept | 432.05 | 40.35 | 82 | 0.00 |
| Median income | 18.41 | 4.72 | 82 | 0.00 |
| Activity level | 7.17 | 2.41 | 82 | 0.02 |
| Model for "UGPA" slope | | | | |
| Intercept | 32.15 | 20.91 | 83 | 0.00 |
| Median income | 6.04 | 3.56 | 83 | 0.01 |
| Model for "Test-taker gender" slope | | | | |
| Intercept | -65.80 | -22.80 | 7,487 | 0.00 |
| Model for "Graduate objective" slope | | | | |
| Intercept | 18.59 | 6.47 | 7,487 | 0.00 |
| Model for "Father's education" slope | | | | |
| Intercept | 6.63 | 10.21 | 7,487 | 0.00 |
| Model for "Mother's education" slope | | | | |
| Intercept | 5.44 | 7.64 | 7,487 | 0.00 |
| *Random effect* | Variance component | df | $\chi^2$ | p-value |
| Between-center | | | | |
| Test-center mean $u_{oj}$ | 851.20 | 82 | 518.89 | 0.00 |
| UGPA slope $u_{1j}$ | 62.46 | 83 | 129.29 | 0.00 |
| Within-center $r_{ij}$ | 13,640.54 | | | |

*Note.* $\hat{Y}_{ij}=\hat{\beta}_{oj}+\hat{\beta}_{1j}$ (UGPA) - 65.80 (Gender) + 18.59 (Graduate objective) + 6.63 (Father's education) + 5.44 (Mother's education). $\hat{\beta}_{oj}$ = 432.05 + 18.41 (Median income) + 7.17 (Activity level). $\hat{\beta}_{1j}$ = 32.15 + 6.04 (Median income).

**GRE-ETS**
**PO Box 6000**
**Princeton, NJ 08541-6000**
**USA**

---

To obtain more information about GRE programs and services, use one of the following:
Phone: 1-866-473-4373
(U.S., U.S. Territories*, and Canada)
1-609-771-7670
(all other locations)
Web site: www.gre.org

* America Samoa, Guam, Puerto Rico, and US Virgin Islands

I.N. 726109