The Effect of Soliciting Demographic Data on the Performance of Students on Online Tests

Amruth N. Kumar
Ramapo College of New Jersey, amruth@ramapo.edu

Abstract – Stereotype threat was studied in the context of online testing in Computer Science. For the study, the student’s identity was made salient by asking the student to identify sex and race, with no additional allusions to positive or negative stereotypes associated with any of their subgroups. It was found that the negative effect of entering demographic data before answering a test was more readily apparent on harder topics. The negative effect of stereotype threat on female and non-Caucasian students was significant among those who had not already mastered the material. On the harder topic in the study, after excluding those who had mastered the material, stereotype threat was found to negatively affect male and Caucasian students also, possibly because of “choking effect”. Given these results, it is recommended that demographic information should be solicited after rather than before online tests.

Index Terms – Stereotype threat, Online testing, Demographic data, Self-confidence.

INTRODUCTION

Asking students to identify their sex and race makes their sexual and racial identity salient. When these identities are associated with positive or negative stereotypes, making the identities salient before a test can positively or negatively affect the performance of the students. Psychologists have studied this effect on women and minorities in mathematics (e.g., [1, 2]), among other disciplines. The effect has been termed stereotype threat (e.g., [3, 4]). Stereotype threat is listed as one of the factors that could be contributing to problems with recruitment and retention of female and minority students in Computer Science [5]: lower expectations for women and minorities in this technical discipline end up becoming self-fulfilling prophecies.

Does stereotype threat extend to online testing in Computer Science? Would making the sex and race identities salient after rather than before an online test lead to better performance, especially of female and minority students? In order to answer these questions, a study was conducted in fall 2009 using two software tutors on computer programming topics, called problets (www.problets.org). The tutors were administered entirely online and were used by introductory Computer Science students as after-class assignments. In this paper, the methods of the study, and its findings and recommendations will be presented.

METHODS

The study was conducted using two software tutors that were developed to help Computer Science students learn about logical and assignment expression evaluation by solving problems. The software tutors use pre-test-practice-post-test protocol: pre-test to assess the prior knowledge of students, practice to help students learn the concepts that they did not already know, and post-test to verify whether students did indeed learn the targeted concepts. For the purposes of this study, the pre-test administered by the tutors was treated as the online test on which the effect of stereotype threat would be studied.

Stereotype threat depends on the elicitation of student’s identity. When using each tutor, students were asked to enter their demographic information, including their sex and race. This was the only instrument used to make students’ identity salient, without any additional allusions to positive or negative stereotypes associated with subgroups of sex and race.

Given this experimental setup, if stereotype threat extends to online testing in Computer Science, students, especially female and non-Caucasian students entering their demographic information after using the tutor can be expected to score higher on the pre-test than their counterparts who enter their demographic information before using the tutor. So, during data analysis, performance of students on the pre-test was treated as the dependent variable.

In the study, students who used the tutors were randomly assigned to one of two groups: A or B. Partial cross-over design was used, with group A serving as control group on logical tutor and test group on assignment tutor; and vice-versa for group B. Control group entered its demographic information before using the tutor, i.e., before answering the pre-test, whereas test group entered its demographic information after using the tutor, all else being equal. All the students used the logical tutor before the assignment tutor.

For analysis purposes, only those students were considered who had used both logical and assignment tutors, and had identified their sex and/or race. Research shows that the effects of stereotype threat are limited to students who highly value the domain [9]. So, in addition, only those students were considered who had attempted at least 10 out of the 14 pre-test problems on logical tutor and at least 10 out of the 17 pre-test problems on assignment tutor. This
meant 163 students - 107 males, 48 females, 85 Caucasians and 57 non-Caucasians (which included Black/African American, Hispanic/Latino, Asian, and Other categories). In order to eliminate the effect of differences in the number of problems solved, the average score per problem was considered rather than the raw score. The average score per problem ranged from 0 → 1.0.

**ANALYSIS AND RESULTS**

A 2 X 2 mixed factor ANOVA analysis was conducted of the pre-test score, with the topic (logical versus assignment expressions) as the repeated measure and the group (A versus B) as the between-subjects factor. No significant main effect was found for the group [F(1,161) = 1.68, p = 0.197]; the combined performance of the two groups over the two tutors was comparable.

A significant main effect was found for the topic [F(1,161) = 23.322, p < 0.001]: students scored more on the logical expressions pre-test (0.926) than on the assignment expressions pre-test (0.885), suggesting that students were either less familiar with assignment expressions than logical expressions, or they found assignment expressions to be harder than logical expressions. The difference in the scores on the two pre-tests was significant [t(162) = 4.87, p < 0.01].

A significant interaction was found between topic and group [F(1,161) = 4.907, p = 0.028]: whereas there was no significant difference in the score on the logical expressions pre-test, whether students entered their demographic data before or after the tutor [t(161) = 0.247, p = 0.805], there was a significant difference in the score on the assignment expressions pre-test between the two groups [t(161) = 2.080, p = 0.039], with the group that entered the demographic data after using the tutor scoring significantly higher (0.906) than the group that entered the data before using the tutor (0.864). This result, combined with the significant main effect for topic might suggest that any negative effect of entering demographic data before answering a test is more pronounced for the topics with which students are less familiar, or the topics that students find to be harder.

However, stereotype threat is supposed to negatively affect women and minorities, not the entire population. So, the data was analyzed by sex and race. First, t-test was conducted to see whether there was any difference in the scores of the two groups:

- There were no significant difference between males and females on logical [t(153) = 0.363, p = 0.717] or assignment [t(153) = 0.565, p = 0.573] expressions pre-test.

- There were no significant difference between Caucasians and non-Caucasians on logical [t(140) = 0.935, p = 0.351] or assignment [t(140) = 1.870, p = 0.064] expressions pre-test.

Therefore, any differences between the groups would be due to treatment, i.e., asking for demographic information before versus after using the tutor.

In order to study the difference between male and female students given the two different treatments, univariate ANOVA analyses were conducted on the pre-test score with treatment (demographics before versus after) and sex as fixed factors.

- On logical expressions pre-test, although an interaction was observed between treatment and sex as shown in Table I, i.e., male students scored higher when asked for demographic information before using the tutor and female students scored higher when asked for the information after using the tutor, the interaction was not significant [F(1,154) = 2.096, p = 0.15].

- On assignment expressions pre-test, both male and female students scored higher when demographic information was solicited after rather than before using the tutor, as shown in Table II. But, the main effect for treatment was marginally significant [F(1,154) = 2.829, p = 0.095], and the interaction between treatment and sex was not significant [F(1,154) = 0.801, p = 0.372].

Similarly, in order to study the difference between Caucasians and non-Caucasians, the same ANOVA analyses were repeated with race instead of sex as the second fixed factor.

- On logical expressions pre-test, an interaction was observed between treatment and race. Caucasians scored higher when asked for demographic information before using the tutor, whereas non-Caucasians scored higher when asked for the information after using the tutor, as shown in Table III. But, the interaction was not significant [F(1,141) = 1.19, p = 0.277].

- On assignment expressions pre-test, both Caucasians and non-Caucasians scored higher when demographic information was solicited after rather than before using the tutor as shown in Table IV. But, the main effect for treatment was marginally significant [F(1,141) = 3.085, p = 0.081] and the interaction between treatment and race was not significant [F(1,141) = 0.013, p = 0.911].
When differences between male and female students were studied:

- On logical expressions pre-test, male students once again scored higher when asked before rather than after using the tutor, whereas female students scored about the same in both cases, as shown in Table V. However, the interaction between treatment and sex was once again not significant \(F(1,77) = 0.286, p = 0.594\).

- On assignment expressions pre-test, both male and female students scored higher when asked for demographic information after rather than before using the tutor, as shown in Table VI. The main effect for treatment was significant \(F(1,112) = 5.883, p = 0.017\). However, no significant interaction was observed between treatment and sex as would be predicted by stereotype threat \(F(1,112) = 0.412, p = 0.522\).

When differences between Caucasians and non-Caucasians were considered:

- On logical expressions pre-test, Caucasian students scored higher when they entered demographic information before rather than after using the tutor, whereas non-Caucasians scored about the same in both cases, as shown in Table VII. But, the main effect for treatment was not significant \(F(1,70) = 0.584, p = 0.447\).

- On assignment expressions pre-test, both Caucasians and non-Caucasians scored higher when demographic information was solicited after rather than before using the tutor as shown in Table VIII. The main effect for treatment was statistically significant \(F(1,102) = 5.694, p = 0.019\). However, the interaction between treatment and race was not significant \(F(1,102) = 0.526, p = 0.470\), as would be predicted by stereotype threat.

### Table V

<table>
<thead>
<tr>
<th>Treatment Versus Sex on Logical Expressions Tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Tutor</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>

### Table VI

<table>
<thead>
<tr>
<th>Treatment Versus Sex on Assignment Expressions Tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment Tutor</td>
</tr>
<tr>
<td>-------------------</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Female</td>
</tr>
</tbody>
</table>

### Table VII

<table>
<thead>
<tr>
<th>Treatment Versus Race on Logical Expressions Tutor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logical Tutor</td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>Caucasians</td>
</tr>
<tr>
<td>Non-Caucasians</td>
</tr>
</tbody>
</table>
It is clear from the results that stereotype threat is real and applicable to online testing in Computer Science. It could adversely affect the performance of male as well as female students, Caucasian as well as minority students. In light of this, it is recommended that demographic information should be solicited after rather than before online testing or tutoring to reduce any adverse effects of stereotype threat.

ACKNOWLEDGMENT

Partial support for this work was provided by the National Science Foundation under grant DUE-0817187. The author thanks Dr. Nicholas Salter for his consultation on statistical analysis.

REFERENCES