Proceedings of The National Conference
On Undergraduate Research (NCUR) 2014
University of Kentucky, Lexington, KY
April 3-5, 2014

Stereotype Spillover Effects on Women in Mathematics

Laura Jensen
Department of Psychology
Bethel College
North Newton, KS 67117

Faculty Advisor: Dwight Krehbiel

Abstract

This study sought to combine previous literature on stereotype threat in women regarding mathematics testing and stereotype threat spillover. Stereotype threat has been found to occur when a person is put into an environment or situation where they feel they may be negatively stereotyped and as a result, due to increased anxiety and emotional vigilance, perform worse than when in a stereotype-free environment[1]. It has also been found that stereotype threat may have spillover effects that last after the individual leaves the threat-filled environment, leading to a continued loss of executive control[2]. Studies have shown that with an intervention of reappraisal, stereotype threat can be diminished[1]. The current study tested if a reappraisal statement would increase women's mathematics scores, minimize Stroop errors, and increase error-related negativity. Error-related negativity is a negative dip in event-related brain potential that occurs after an individual makes a mistake[3]. Error-related negativity is increased when individuals try for accuracy and is decreased when individuals prioritize speed over accuracy[4]. Forty-six Bethel College undergraduate students served as participants. A between-subjects design was used comparing reappraisal and non-reappraisal conditions. Dependent variables included mathematics test scores, Stroop errors, and error-related negativity measures. Females were randomly assigned into the non-reappraisal or reappraisal group. All participants were given a test of mathematics performance followed by a Stroop task and debriefing sheet. Participants in the reappraisal condition were given instructions that included a reappraisal sentence just prior to the mathematics test. The reappraisal sentence informed participants that they should think of the test objectively rather than personally relevant to themselves. Electroencephalographic data were recorded during the Stroop task to test for error-related negativity. Contrary to our expectations, the results demonstrated that the reappraisal condition (M = 4.83) had significantly lower mathematics test scores than did the non reappraisal (M = 11.92) condition $t(20.46) = 3.87$, $p < 0.001$). While Stroop errors showed no significant difference between groups, females in the non-reappraisal group had significantly higher error-related negativity during the Stroop task, indicating that females in this condition were monitoring their mistakes more closely than females in the reappraisal condition. These surprising results suggest that the reappraisal sentence may have induced rather than diminished stereotype threat. It is possible that in the environment at Bethel College women are not ordinarily subject to negative mathematics stereotypes. It is also possible that the reappraisal sentence prompted individuals to care less about their performance. Further research is needed to determine the conditions under which mathematics stereotype threat is experienced and how it may be evoked during reappraisal.

Keywords: Stereotype Threat, Spillover, Women

1. Introduction

It has been routinely found that when a person is put into an environment or situation where they feel they may be negatively stereotyped, they tend to perform worse than when in a stereotype free environment[1]. This phenomenon has been termed stereotype threat. This threat causes individuals who identify with the stereotyped group to experience stress in a number of different ways. This stress caused by not wanting to be identified with the negative

stereotype seems to increase arousal, vigilance, distracting thoughts, and anxiety[2]. These unconscious reactions cause the individual to increase emotional regulation, consuming limited executive and working memory resources, thus diminishing the cognitive resources left for the task at hand. This reduction in executive control and working memory seem to be the major causes of the reduction in performance produced by stereotype threat. These findings are of immense importance as they may help explain the gap between different groups on standardized tests[5]. This phenomenon can be experienced by any group that is negatively stereotyped. One group that is affected by stereotype threat is women, specifically their performance in mathematics[1]. This threat is found to occur when women are made aware of the negative stereotype that men outperform women in mathematics[6]. Those women with higher self-confidence in their mathematic performance have been found to be affected less by stereotype threat than those with low self-confidence in their mathematic performance[7].

Stereotype threat may also produce spillover -- effects that last after the individual leaves the threat-filled environment[2]. It has been found that stereotype threat may cause such a cognitive drain on individuals that after they have left the threat-filled environment they continue to have less executive control. This continued loss of executive control has been found to then lead to a loss of self-control in tasks relating to the stereotype examined and also in unrelated tasks such as food intake[2]. Self-control is cognitively demanding and when an individual is cognitively drained this seems to lead to a drop in self-control. Thus the stereotype threat spillover effects seem to occur on subsequent tasks that rely on the same type of cognitive resources that the previously experienced stereotype threat drained[8].

Studies have shown that with a simple intervention of reappraisal, these harmful effects can be diminished[1]. Reappraisal seems to cause participants to rethink their emotions. Reappraisal can be done by informing individuals that the stereotype in the said situation has never shown an effect—for example, telling females prior to taking a mathematics test that females and males routinely perform equally well on the test. It can also be done by telling individuals to think of the test objectively and analytically rather than personally or emotionally relevant to them. A study by Jamieson, Mendes, Blackstock, & Schmader[9] found that individuals preparing to take the GRE who were told that performance was improved by feelings of arousal had significantly higher scores than those than were given no reappraisal intervention.

The present study investigated the possibility that stereotype threat experienced by women in mathematics would produce spillover into a subsequent task. Stereotype threat spillover was tested using a Stroop task. Electroencephalographic data were collected in order to investigate potential error-related negativity associated with errors in the Stroop task. Error-related negativity is a negative dip in event-related brain potential (ERP) that occurs after an individual makes a mistake[3]. ERP is increased when individuals try for the most accurate performance and is decreased when individuals prioritize speed over accuracy[4]. Taken together, the findings of this study have the potential to not only further our knowledge of stereotype threat, but additionally work toward explaining the underlying processes contributing to this debilitating phenomenon, leading to the development of more effective reappraisal interventions.

Three main hypotheses were proposed based on the previous research. The first was that females who did not receive reappraisal would have lower mathematics scores and spend a longer time on the test than females who received reappraisal. Females who did not receive reappraisal would experience stereotype threat, resulting in a decrease in test accuracy due to increased emotional vigilance and anxiety. It was also thought that this additional pressure put on the working memory would cause a reduction in its use as well as a reduction in executive functioning for the mathematic test at hand, causing them to take longer to complete it.

The second hypothesis was that females who received the reappraisal intervention would have fewer errors and shorter reaction times on the Stroop task than would those who did not receive the reappraisal intervention. This hypothesis was based on studies previously cited indicating that stereotype threat may use up executive control so much that performance is hindered after the individual leaves the threat-filled environment[2].

The final hypothesis was that females who received reappraisal intervention would have more negative error-related negativity than females who did not. Females who did not receive reappraisal would have fewer cognitive resources left to notice and/or care if they made a mistake, resulting in less error-related negativity.

2. Methods

2.1 Participants

Forty-six Bethel College students (25 females, 21 males, M_{age}= 20, age range: 17-26) were recruited from psychology courses at Bethel College and were compensated with extra credit for their participation. Since the hypotheses were focused on females, the data analyses were strictly concerned with female data. Twelve percent of females had taken 0-3 mathematics courses, twenty percent had taken 4-6, sixteen percent had taken 4-6, and none had taken 7+ mathematic courses. Eight percent had taken a mathematics course in the past year, sixteen percent in the past two years, twenty-eight percent in the past, and thirty-six percent were currently taking a mathematics course.

2.2 Instruments

A mathematics test was created formed from twenty questions taken from previous editions of the ACT[10] and SAT[11] achievement tests. The Stroop task was used which consisted of ten blocks of trials with twenty-four congruent and twenty-four incongruent in each block for a total of 480 trials. Congruent trials consisted of words whose semantic meaning matched the word color while incongruent trials consisted of words whose semantic meaning did not match the word color. EEG data were recorded using ActiVIEW software, running on LabVIEW (National Instruments, Austin, TX) (version 8.6.1) on an iMac computer (MacOS 10.6.8). The software also allowed the recording of an event channel with signals provided by the PsyScope[12] script. PsyScope was also used to create the mathematics test which was graded by the program. EEG recording was completed using a Biosemi 32-channel biopotential measurement system with eight auxiliary channels (http://www.biosemi.com/products.htm). Four auxiliary channels were used as left and right mastoid reference electrodes, and vertical and horizontal eye movement electrodes. Electroencephalographic (EEG) data were recorded to collect brain electrical activity with the use of sixteen electrodes.

2.3 Design

A between-subjects design was used for this experiment with the main independent variable being reappraisal condition. There were two levels of the reappraisal condition: female reappraisal (FR) and female non-reappraisal (FNR). Dependent variables were mathematic test scores, mathematic test completion time, Stroop errors, Stroop reaction time, and error-related negativity measures.

2.4 Procedure

Prior to beginning the experiment, an EEG cap was placed on participants. All sixteen electrodes were put into place, with electrodes placed below and beside the left eye in order to control for eye movement. Electrodes were also placed behind the right and left ears on the mastoid bone in order to collect a baseline for the data. Electroencephalographic data were not recorded during the mathematics test, only during the Stroop task. All participants were first given the mathematics test, comprised of twenty questions. Both groups received instructions first, informing participants that they were to take a test of mathematic intelligence. No mention was made about gender effects, as this would be unrealistic in everyday testing situations. The reappraisal condition then read the reappraisal intervention after the instructions, while the non-reappraisal condition was not given the sentence and began the test. The reappraisal statement was "Think about this test objectively and analytically rather than as personally or in any way emotionally relevant to you." Participants were given as much time as they needed to finish the test. Half of the females were randomly selected to receive reappraisal instructions within their mathematics test.

Immediately following the mathematics test, participants were read instructions pertaining to the Stroop task. Participants were instructed to indicate the semantic meaning of the world shown by either choosing "r" on the keyboard for red or "g" on the keyboard for green. Electroencephalographic data were recorded during this portion of the experiment. When participants finished the Stroop task, EEG recording was stopped, and participants were given a debriefing sheet and dismissed.

2.5 Data Analysis

To begin data analysis of the behavioral data, each participant's data were put into separate Excel spreadsheets. Participants' data were then put into a compiled spreadsheet where errors on the mathematics test, completion time on the mathematics test, errors on the Stroop task for both congruent and incongruent stimuli, and reaction time on the Stroop task for both congruent and incongruent stimuli were calculated. Independent samples t-tests[13] were employed to compare non-reappraisal and reappraisal groups on both mathematics test and Stroop data. Two-way analysis of variance was used to see whether potential differences between reappraisal and non-reappraisal groups may be attributable to differences in mathematics experience (see Participants).

To begin data analysis of the electroencephalographic (EEG) data, data were imported into MATLAB (The MathWorks, Natick, MA) and EEGLAB[14]. EEG data were band-pass filtered from 0.1 to 30 Hz. The data were separated into bins based on correct and incorrect responses to the Stroop task. Epochs of 1.5 seconds, starting at 0.5 seconds pre-response onset, were extracted for both correct and incorrect responses. Epoched data were subjected to independent components analysis and eye movement components were deleted. Then data for each participant were averaged across trials yielding four ERP waveforms: reappraisal-correct, reappraisal-incorrect, non-reappraisal-correct, non-reappraisal-incorrect. Finally ERP data were averaged across participants, and the four experimental conditions were compared statistically.

3. Results

Analyses of the mathematics test scores indicated that, contrary to expectations, the reappraisal group actually performed more poorly than the non-reappraisal group.

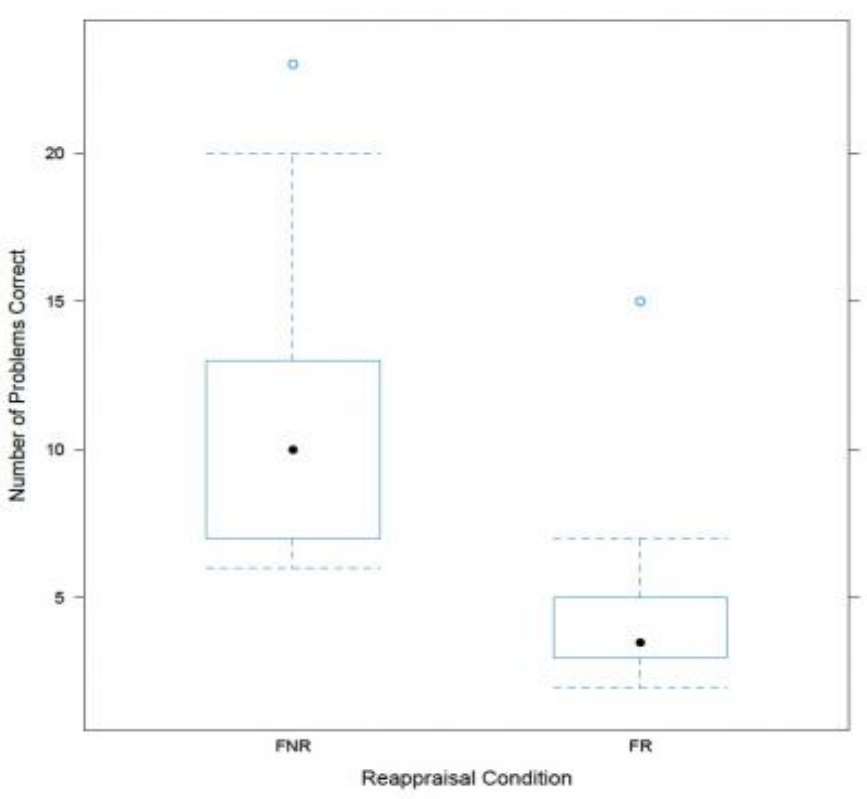

Figure 1. Number of correct answers on mathematics test

Figure 1. Number of correct answers on the mathematics test for each condition. Significant differences were found between the two conditions with an independent samples t-test, t (20.46) = 3.87, p < 0.001. Females in the reappraisal condition (*M* = 4.83) had significantly lower scores than those in the non-reappraisal condition (*M* = 11.92).

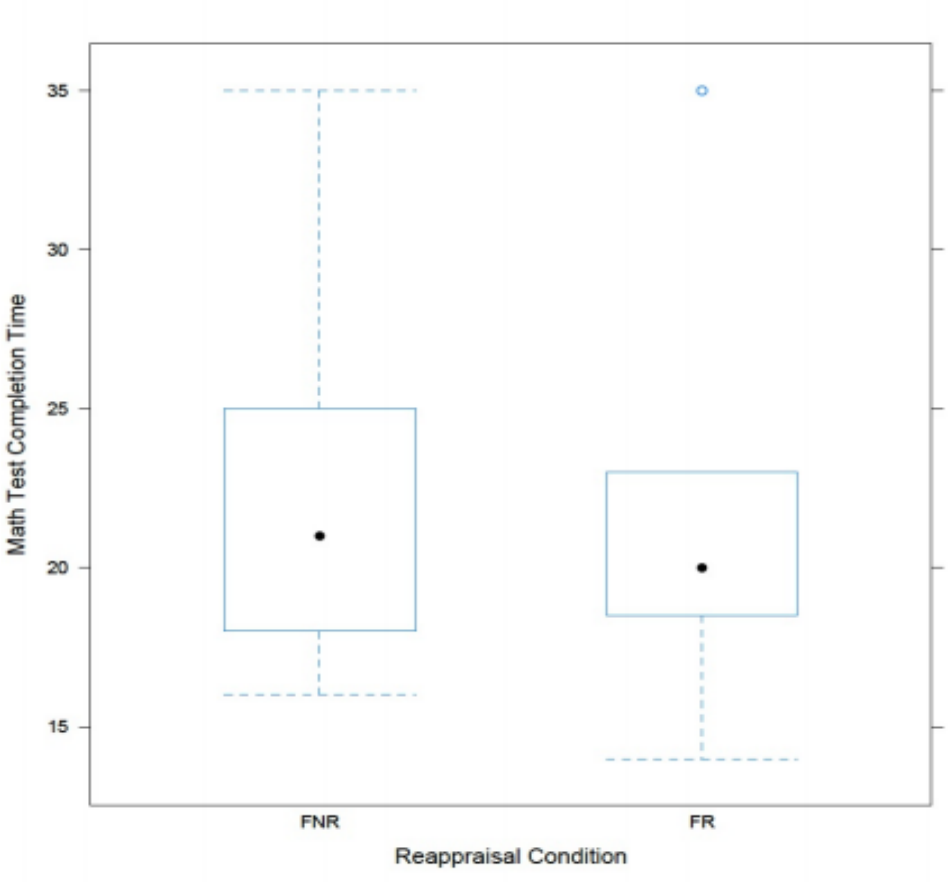

Figure 2. Mathematics test completion time

Figure 2. The time it took each condition to finish the mathematics test. Significant differences were not found between the two conditions when comparing mathematic test completion time t (21.68) = 0.588, p < 0.563.

Females in the non-reappraisal condition had significantly more correct answers on the mathematics test compared to the females that received the reappraisal intervention. The possibility that these t-test results were biased by the lack of normal distribution was examined by conducting a two-sample Wilcoxon test, comparing reappraisal and non-reappraisal groups, and significant differences were again found (p < 0.001). No significant differences were found, however, between the two groups for the time it took participants to complete the mathematics test. The differences between groups were not due to possible group differences in mathematics experience as an analysis of variance indicated significant differences between conditions (p < 0.014), but no significant differences between mathematics experience groups (p > 0.98). Thus, the differences found in mathematics scores between the two groups were not due simply to differences in mathematics experience.

Analyses of Stroop data focused on differences between congruent and incongruent trials. Independent sample t-tests found no significant differences between reappraisal (*M* = 10.17) and non-reappraisal (*M* = 36.59) conditions (p > 0.15) in number of errors on the Stroop task. Wilcoxon test results were again consistent with those of the t-test (p > 0.47). No significant differences were found between the reappraisal (*M* = 89.33) and non-reappraisal groups (*M* = 82.08) when analyzing reaction time in the Stroop task (p > 0.8).

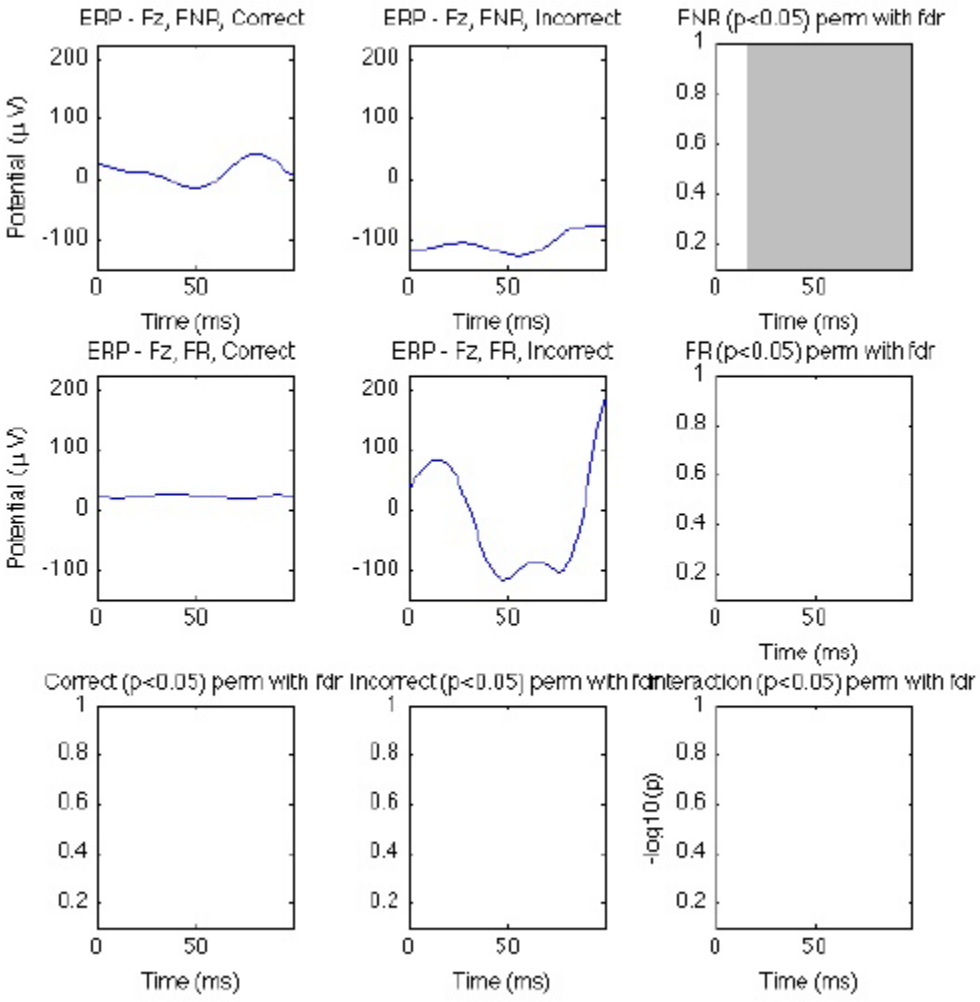

Figure 3. Fz Waveform

Figure 3. ERP waveforms for the Fz electrode from 0-100 milliseconds after the response on the Stroop task in each condition. At 0-100 milliseconds after the response, significant differences were found for females in the non-reappraisal (FNR) condition at the Fz electrode (note shaded area in upper right panel). After incorrect responses, females who did not receive the reappraisal intervention had significantly greater error-related negativity than when answering correctly. Note that these comparisons utilized false discovery rate (fdr) to correct p-values for multiple comparisons.

The Fz electrode was chosen for analysis as previous studies have shown it to demonstrate differences in error-related negativity[4]. Results show that females in the non-reappraisal condition had significantly greater error-related negativity when they answered a question incorrectly on the Stroop task than when answering correctly. Females who received the reappraisal statement, however, did not show a change when answering incorrectly versus correctly on the Stroop task. These findings indicate that reappraisal interventions during the mathematics test had a persisting influence during the Stroop task, indicating spillover effects.

4. Discussion

These results indicate that, contrary to the proposed hypotheses, reappraisal instructions did not improve scores on the mathematics test, nor reduce the time spent on the test. Reappraisal instructions also did not seem to reduce errors made on the Stroop task or reduce reaction time as hypothesized. Instead, the group who received reappraisal produced significantly more errors on the mathematics test (indicating the experience of stereotype threat), and their error-related negativity during the Stroop task was also diminished after incorrect trials (indicating spillover of the previously experienced stereotype threat).

Results seem to indicate that the reappraisal intervention, which previous literature had shown to decrease stereotype threat, actually increased stereotype threat in the female participants. There are a number of possibilities for these opposite findings. One reason may be that all participants were students at Bethel College where one of the main mathematics instructors is a female. Previous studies have shown that identifying with a female role model who is strong in mathematics may decrease the stereotype of women being less capable in mathematics[1]. It is possible that in the environment at Bethel College women are not ordinarily subject to negative mathematic stereotypes. A similar study done at a different small liberal arts college, similar to Bethel College, acquired similar results when trying to induce stereotype threat on women in regards to the mathematics stereotype[16]. They found that under stereotype threat conditions, women did not underperform males on a mathematics test. Therefore, it is also possible that the more personal environment found at a small liberal arts college may work itself to diminish the stereotype that women perform worse than males in mathematics. Perhaps participants were not aware of this stereotype and when they encountered the reappraisal intervention it caused them to become more aware of their emotions and increased their anxiety.

These results suggest that the creation and usage of interventions of reappraisal to undermine stereotype threat are not as straightforward as once thought. Many varieties of interventions have been studied and recent literature suggests that the intervention chosen needs to coordinate with the specific stereotype threat situation one is trying to reduce[15]. Many different factors must be taken into account regarding the specific situation and environment in which stereotype threat may be taking place. The results from this study show what may happen if these factors are not accounted for; the intervention method may in fact induce, rather than combat, stereotype threat.

5. Acknowledgments

The author wishes to express appreciation to Brad Celestin and the 2013 Bethel College Cognitive Psychology lab members for their hard work that made this study possible. She would also like to thank Dr. Dwight Krehbiel for his encouragement and guidance.

6. References Cited

1. Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype Threat and Women's Math Performance. *Journal of Experimental Social Psychology*, 4–28.
2. Inzlicht, M., & Kang, S. K. (2010). Stereotype threat spillover: How coping with threats to social identity affects aggression, eating, decision making, and attention. *Journal of Personality and Social Psychology, 99*(3), 467–481.
3. Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology, 42*, 33–42.
4. Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4*(6), 385–390.
5. Derks, B., Inzlicht, M., & Kang, S. (2008). The neuroscience of stigma and stereotype threat. *Group Processes & Intergroup Relations, 11*(2), 163–181.
6. Jones, P.R. (2012). Reducing the impact of stereotype threat on women's math performance: Are two strategies better than one? *Rev Electron Investig Psicoeduc Psigopedag, 9*(2), 587-616.
7. Gerstenberg, F. R., Imhoff, R., & Schmitt, M. (2012). 'Women are Bad at Math, but I'm Not, am I?' Fragile Mathematical Self-concept Predicts Vulnerability to a Stereotype Threat Effect on Mathematical Performance. *European Journal Of Personality, 26*(6), 588-599. doi:10.1002/per.1836

8. Beilock, S.L., Rydell, R.J., & McConnell, A.R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology, 136*(2), 256-276.
9. Jamieson, J. P., Mendes, W., Blackstock, E., & Schmader, T. (2010). Turning the knots in your stomach into bows: Reappraising arousal improves performance on the GRE. *Journal Of Experimental Social Psychology, 46*(1), 208-212. doi:10.1016/j.jesp.2009.08.015
10. (2014). *Sample mathematics questions.* Retrieved from http://www.actstudent.org/sampletest/math/math_01.html.
11. (2014). *SAT practice questions*. Retrieved from https://sat.collegeboard.org/practice/sat-practice questions.
12. Cohen J.D., MacWhinney B., Flatt M., and Provost J. (1993). PsyScope: A new graphic interactive\environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257-271.
13. R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
14. Delorme, A., & Makeig, S. (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics (pdf, 0.7 MB) *Journal of Neuroscience Methods* 134:9-21
15. Shapiro, J. R., Williams, A. M., & Hambarchyan, M. (2013). Are All Interventions Created Equal? A Multi Threat Approach to Tailoring Stereotype Threat Interventions. *Journal Of Personality & Social Psychology, 104*(2), 277-288. doi:10.1037/a0030461
16. Rivardo, M. G., Rhodes, M. E., & Klein, B. (2008). LACK OF STEREOTYPE THREAT AT A LIBERAL ARTS COLLEGE. *College Student Journal, 42*(3), 832-841.