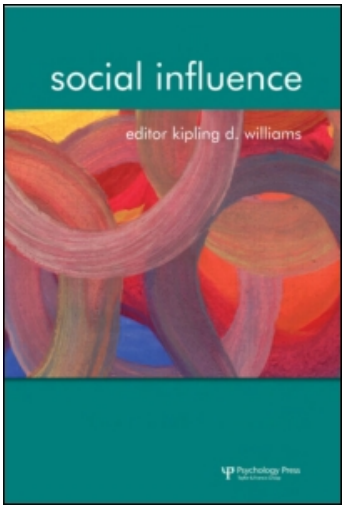
This article was downloaded by: *[Jamieson, Jeremy P.]*
On: *17 February 2010*
Access details: *Access Details: [subscription number 919326556]*
Publisher *Psychology Press*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Social Influence
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100705

Evaluation is necessary to produce stereotype threat performance effects

Jeremy P. Jamieson [a]; Stephen G. Harkins [b]
[a] Harvard University, Cambridge, MA, USA [b] Northeastern University, Evanston, IL, USA

First published on: 17 February 2010

To cite this Article Jamieson, Jeremy P. and Harkins, Stephen G.(2010) 'Evaluation is necessary to produce stereotype threat performance effects', Social Influence,, First published on: 17 February 2010 (iFirst)
To link to this Article: DOI: 10.1080/15534510903512409
URL: http://dx.doi.org/10.1080/15534510903512409

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

SOCIAL INFLUENCE
2010, 1–12, iFirst

Psychology Press
Taylor & Francis Group

Evaluation is necessary to produce stereotype threat performance effects

Jeremy P. Jamieson
Harvard University, Cambridge, MA, USA

Stephen G. Harkins
Northeastern University, Evanston, IL, USA

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

In the first direct test of the assumption that an evaluative source is necessary to produce stereotype threat effects, we combined a traditional stereotype threat manipulation with a scoring manipulation. Participants in a "pooled" condition believed that their scores would be averaged across gender, whereas in a "segregated" condition participants were led to believe that their scores would be segregated by gender. Thus we manipulated the potential for the experimenter to evaluate performance at the level of the group. Even though the gender stereotype was equally salient in the pooled and segregated threat conditions, performance was debilitated only in the segregated/threat condition, suggesting that evaluation is indeed necessary to produce stereotype threat performance effects.

Keywords: Stereotype threat; Evaluation; GRE; Group identity.

Stereotype threat, the concern about confirming a negative stereotype about one's group, has received a great deal of attention over the past decade. Work in this area has produced a large corpus of research demonstrating the reliability and generalizability of the effect of stereotype threat on performance. Steele, Spencer, and Aronson (2002) have argued that the debilitating effects of threat on performance are mediated in multiple ways—cognitively, affectively, and motivationally. Consistent with this view, Schmader, Johns, and Forbes (2008) identified the role of working memory as a core cognitive faculty implicated in stereotype threat effects, whereas other research has demonstrated the role of affect (e.g., anxiety: Bosson,

Address correspondence to: Jeremy P. Jamieson, 295.02 Northwest Building, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA. E-mail: jamieson@wjh.harvard.edu

© 2010 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business
http://www.psypress.com/socinf DOI: 10.1080/15534510903512409

Haymovitz, & Pinel 2004), and motivation (e.g., Jamieson & Harkins, 2007, 2009) in producing stereotype threat performance effects. Thus, since its conceptualization, the focus of threat research has changed from simply documenting the effects of stereotype threat to identifying the mechanisms through which threat produces performance effects.

Researchers have also made efforts to establish the conditions that are necessary and sufficient to produce the experience of stereotype threat in stigmatized individuals. For example, to produce threat effects, the relevant stereotype must be made salient. A common manipulation of stereotype threat explicitly informs stigmatized participants that group differences exist on tasks relevant to the stereotyped domain (e.g., Brown & Pinel, 2003; Jamieson & Harkins, 2007; Keller & Dauenheimer, 2003; O'Brien & Crandall, 2003; Spencer, Steele, & Quinn, 1999). Subtle manipulations of stereotype threat (e.g., solo status) can also increase the salience of the stereotype (Ben-Zeev, Fein, & Inzlicht, 2005; Inzlicht & Ben-Zeev, 2000; Schmader & Johns, 2003). Thus the salience of the belief that one's group is thought to be inferior in performance in the particular task domain is a central to the experience of stereotype threat. Research has also shown that participants must see the task as diagnostic of ability in the stereotyped domain (e.g., Gonzales, Blanton, & Williams, 2002; McKay, Doverspike, Bowen-Hilton, & Martin, 2002), and must identify with performing well in the test domain (Smith & White, 2001) to produce threat effects.

In recent work Marx and Stapel (2006) have further refined our understanding of the conditions that are necessary and sufficient to produce stereotype to stereotype threat by showing that, despite the fact that stereotype threat and priming can each debilitate performance, the processes that produce these effects are different. Specifically, they show that stereotype threat effects stem from the fact that the target of the stereotype knows about the stereotype and is a member of the stigmatized group, whereas priming effects simply require knowledge of the stereotype. As they write: "in stereotype threat situations, targets (but not nontargets) are affected because they know the group stereotype ('women are bad at math') and because they are members of the group that is targeted by the stereotype ('I am a woman')" (p. 244). It is this combination of "knowing and being" that gives rise to threat-based concerns, which distinguishes stereotype threat effects from priming effects.

Although this previous work identifies the roles played by a number of factors in producing stereotype threat, we would argue that another factor, the potential for evaluation, also plays an integral, albeit implicit, role in the production of stereotype threat performance effects. In their seminal paper Steele and Aronson (1995) posited that threat was produced by the fear of confirming the negative stereotype directed at one's group, but stereotypes cannot be confirmed if performance cannot be evaluated. In fact, even

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

though researchers have employed different types of stereotype threat manipulations, every stereotype threat study to date has included some potential source of evaluation as part of the manipulation (e.g., Inzlict & Ben-Zeev, 2003; Jamieson & Harkins, 2007; Stone & McWhinnie, 2008).

In their multi-threat framework, Shapiro and Neuberg (2007) argue that six qualitatively distinct stereotype threats arise from the intersection of two dimensions—the target (the self vs one's group) and source (the self vs outgroup others vs ingroup others) of threat. According to this model, stereotype threat arises from the knowledge that one (or one's group) is stereotypically inferior in a domain, the task is diagnostic of ability in the stereotyped domain, and an audience (self, ingroup other, or outgroup other) can evaluate performance, whether at the level of the self or group. Thus, in this typology, there is no threat without the potential for evaluation (i.e., source).

The goal of the current research is to directly test the assumption that an evaluative source is necessary to produce stereotype threat performance effects. This work focused on gender-math ability stereotypes. Thus, females were told that they would either be completing a math test on which gender differences had been found (gender difference condition), or they were told that the performance of men and women did not differ (no gender difference condition). Unlike previous research, we also introduced a scoring manipulation. In the ''pooled'' condition females were led to believe that their score would be averaged across gender, but in the ''segregated'' condition participants believed scores would be segregated by gender. Thus, we varied whether or not there was the potential for the experimenter to make a group-level evaluation of the performance of males versus females.

This source/target combination (outgroup member as source, group as target) was chosen because it minimizes differences between the control and experimental conditions, while maximizing the salience of group identity, an important component of the experience of stereotype threat. In each scoring condition females were led to believe that neither they, nor the experimenter, would be able to evaluate their performances as individuals. Thus, it was their performance as a representative of their group (females), not as an individual, that was implicated. In each condition participants were also led to believe that they would be unable to compare performances at the level of groups (female vs male). Thus the only difference between the scoring conditions was that, in the segregated condition, females were told that their scores would be averaged but segregated by gender, which would allow the experimenter to evaluate the performances of males and females, whereas in the pooled condition females were told that scores would be pooled across gender, which minimized the potential for evaluation.

If the potential for evaluation is necessary to produce threat, then the gender difference manipulation should only impact the performance of

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

females in the segregated condition, but not in the pooled condition. On the other hand, evaluation may not be necessary to elicit stereotype threat performance effects if stigmatized individuals are concerned about performing poorly on any task that is diagnostic of ability in the stereotyped domain when their stereotyped status is made salient. That is, the threat may be rooted solely in the knowledge of one's stereotyped status and the fact that one is a target of the stereotype. If this is the case, then females' performance should not be impacted by the scoring manipulation.

To test these predictions we examined performance on problems taken from the quantitative section of the GRE general test (GRE-Q). We crossed a typical stereotype threat manipulation, which we refer to as the gender difference manipulation, with a scoring manipulation (pooled vs segregated). If the potential for evaluation is required to produce the experience of threat, we should observe an interaction between the gender difference and scoring conditions. That is, performance debilitation would only be expected in the gender difference/segregated condition. However, if evaluation is not a necessary component of the stereotype threat experience, then we should find a main effect for the gender difference manipulation. Participants in the gender difference condition performance more poorly than those in the no gender difference condition, regardless of the manner in which scores are treated.

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

METHOD

Participants

A total of 52 Northeastern University undergraduate students (all female) participated in this experiment for partial fulfillment of a course requirement.

Materials

Participants were given 15 multiple-choice comparison type math problems taken from actual GRE-Q tests. Problems were selected such that overall accuracy averaged 50% (range: 38–60%). Each problem required the test taker to compare the quantity in a column marked "A" to a column marked "B," and select one of four standardized answer choices: (a) the quantity in Column A is greater, (b) the quantity in Column B is greater, (c) the two quantities are equal, or (d) the answer cannot be determined from the information provided. This type of problem has successfully used by previous stereotype threat research to produce performance deficits (Jamieson & Harkins, 2009). Participants were given two practice problems prior to beginning the test. Participants worked on the test for 10 minutes

and were instructed to complete as many problems as they could as accurately as possible.

Procedure and manipulations

Verbal and written instructions were given, and after participants completed the practice trials the experimenter implemented the gender difference and source manipulations. The gender difference manipulation preceded the scoring manipulation. In the gender difference condition, participants were instructed: "*The task you are about to complete is a test of math ability. Previous research has demonstrated that gender differences exist on some math tests, but not on others. The test on which you are about to perform has been shown to produce gender differences.*"

The control condition instructions were identical to the gender difference instructions except the last sentence read: "*The task on which you are about to participate has not been shown to produce gender differences.*" No specific mention was made as to whether men outperformed women or vice versa, only that gender differences did or did not exist on the task. Participants were expected to infer that women would perform more poorly than men based on the societal stereotype that men are superior to women in mathematical and spatial ability. Thus, in the gender difference condition, participants were told that the test was diagnostic of ability in the stereotyped domain and were also informed of their stereotyped status.

The scoring manipulation then followed. All of the participants were told: "*To ensure confidentiality, no one on our research team will know how well you performed. At the end of the task, you will be asked to place your test in an envelope that contains the tests of participants who completed this experiment before you.*" In the pooled condition the participants were then told: "*Because we already know that there are*/ [gender difference condition] *are no* [no gender difference condition] *gender differences, the envelope contains the scores of all participants, both males and females.*" In this condition there was only one envelope in the experiment room labeled "math tests." Prior to leaving the room the experimenter also emphasized the fact that the scores of males and females were combined into one average such that scores could not be evaluated either at the level of the individual or the group.

In contrast, participants in the segregated condition were told: "*Because only tests of participants of one gender are contained within each envelope, at the end of the experiment, we will be able to compare the average performance of males and females to see the presence*/ (gender difference) *absence* (no gender difference) *of gender differences on this task.*" In this condition participants saw two envelopes labeled "female tests" and "male tests." This condition eliminates the potential for the experimenter to assess any one individual's score, but preserves his/her ability to compare the average score

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

for the females against the average score for the males because the gender means remain segregated.

After 10-minutes the experimenter knocked on the door and, as instructed, participants added their sheets to the requisite envelopes so that the experimenter could examine the average scores of females (segregated) or all participants (pooled) without knowing any one individual's score.

Upon the completion of the test participants completed a questionnaire. Two questions allowed us to evaluate the effectiveness of the gender difference manipulation: "To what extent are there gender differences in performance on this task?" (1 = "no gender differences" and 11 = "gender differences"); and "Who do you believe performs better on this task?" (1 = "males perform better," 6 = "males and females perform the same," and 11 = "females perform better"). We were also able to asses the effectiveness of the scoring manipulation, as participants indicated the extent to which they could evaluate their own performance (self as source, self as target), the extent to which they could compare the scores of males and females (self as source, group as target), the extent to the experimenter knew how well they performed (outgroup member as source, self as target), and the extent to which the experimenter could evaluate the scores of males and females (outgroup member as source, group as target), all on 11-point scales (1 = "not at all" and 11 = "to a great extent").

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

RESULTS

Unless otherwise noted, the data were analyzed in 2 (condition: gender difference vs no gender difference) × 2 (scoring: pooled vs segregated) ANOVAs with condition and scoring as between-participants factors. Pairwise contrasts were used to decompose significant interactions (Kirk, 1995).

Questionnaire findings

Gender difference manipulation. Participants in the gender difference condition reported that gender differences existed to a greater extent ($M = 7.65, SD = 1.77$) than controls ($M = 2.54, SD = 1.79$), $F(1, 48) = 110.56$, $p < .001$, $d = 3.03$. Participants in the gender difference condition also indicated that males outperform females ($M = 3.35, SD = 1.38$) to a greater extent than no gender difference participants ($M = 5.69$, $SD = .88$), $F(1, 48) = 51.80$, $p < .001$, $d = 2.08$. Taken together, these findings indicate the success of the gender difference manipulation. Furthermore, the scoring manipulation did not impact ratings of gender differences or whether males outperformed females, $ps > .30$.

Scoring manipulation. We then analyzed participants' response to the questions regarding potential sources of evaluation. All means and standard deviations are presented in Table 1. Analysis of the questionnaire item asking participants whether they could evaluate their own performance produced no reliable effects. A one-sample *t*-test illustrates that ratings of self-evaluation ($M = 3.40, SD = 2.41$) were significantly lower than the midpoint of the scale (6), $t(51) = -7.77$, $p < .001$, $d = 2.18$. No significant effects emerged for ratings of how well the experimenter could evaluate individual-level performance, and the overall mean ($M = 2.75, SD = 2.50$) fell below the midpoint of the scale, $t(51) = -9.36$, $p < .001$, $d = 2.61$.

Neither the gender difference nor the scoring manipulations impacted ratings of the extent to which participants could compare the scores of males and females, and the overall mean ($M = 2.79, SD = 2.26$) fell below the scale's midpoint (6), $t(51) = -10.24$, $p < .001$, $d = 2.80$.

Consistent with the success of the scoring manipulation, analysis of participants' ratings of the extent to which the experimenter could make group-level comparisons produced a main effect for scoring, $F(1, 48) = 36.62$, $p < .001$, $d = 1.75$. Participants in the segregated condition indicated the experimenter was more able to compare males' and females' scores ($M = 8.39, SD = 2.45$) than participants in the pooled condition ($M = 3.57, SD = 3.22$). In addition, the segregated participants' rating exceeded the scale's midpoint (6), $t(25) = 4.96$, $p < .001$, $d = 1.98$, whereas the pooled participants' rating fell below the scale's midpoint (6), $t(25) = -3.83$, $p = .001$, $d = 1.53$.

These data indicate the scoring manipulation was successful. Neither participants nor the experimenter were rated as being able to evaluate individual-level performance. Participants also reported they could not

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

TABLE 1
Participants' responses to the evaluation questionnaire items as a function of scoring condition

Scoring Condition	*Pooled*		*Segregated*	
	M	*SD*	*M*	*SD*
Self: Individual-level evaluation	3.69_a	2.49	3.11_a	2.34
Experimenter: Individual-level evaluation	3.00_a	2.91	2.50_a	2.04
Self: Group-level evaluation	2.85_a	2.63	2.73_a	1.87
Experimenter: Group-level evaluation	3.57_a	3.22	8.39_b	2.45

Different subscript letters indicate significant mean differences within the row.
All questions were presented on 11-point scales (1 = not at all, 11 = to a great extent).

evaluate group scores, and only in the segregated condition was the experimenter rated as being able to compare the scores of males and females.

GRE performance

To test for the effects of stereotype threat on GRE problems we examined the performance of participants on two measures: the number of problems attempted and the percentage of these problems that were correctly answered.

Problems attempted. Analysis of the number of problems attempted revealed no significant effects. Overall, participants attempted 12.29 problems.

Percentage correct. We observed main effects for both gender difference, $F(1, 48) = 7.12$, $p = .010$, $d = 0.77$, and scoring, $F(1, 48) = 9.71$, $p = .003$, $d = 0.90$. However, these main effects must be interpreted in the context of the Condition × Scoring interaction, $F(1, 48) = 9.23$, $p = .004$, $d = 0.88$ (see Figure 1).

In the pooled condition, gender difference participants ($M = 44.57\%$, $SD = 15.61\%$) performed as well as control participants ($M = 43.20\%$, $SD = 12.15\%$), $F < 1$. However, when participants' scores were segregated

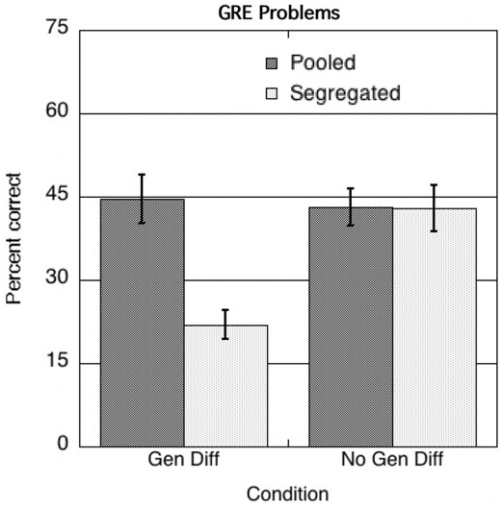


Figure 1. Percentage of quantitative GRE problems answered correctly as a function of the gender difference and scoring manipulations. Gen Diff = gender difference condition, No Gen Diff = no gender difference condition. Error bars = ±standard error of the mean.

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

by gender, females in the gender difference condition performed more poorly ($M = 21.94\%, SD = 9.46\%$) than their no gender difference counterparts ($M = 42.92\%, SD = 14.90\%$), $F(1, 48) = 16.28$, $p < .001$, $d = 1.16$. In addition, participants in the gender difference/segregated condition performed more poorly than participants in the gender difference/pooled condition, $F(1, 48) = 18.94$, $p < .001$, $d = 1.25$.

DISCUSSION

The present research tested the assumption that the potential for evaluation is necessary to produce stereotype threat performance effects. Females were told that a test of math ability either did or did not produce gender differences. Manipulation checks indicated that, in the gender difference condition, females were aware of the existence of the female-math ability stereotype and believed that females performed more poorly than males to a greater extent than control participants on the math test. Thus females in the gender difference condition had knowledge of their stereotyped status and the diagnosticity of the math test. However, unlike previous stereotype threat research, we crossed this typical stereotype threat manipulation with a scoring manipulation, which varied whether or not performance was subject to evaluation. In each condition participants were instructed that individual scores would remain anonymous. In the segregated condition females believed their scores would be averaged with those of other females and the experimenter would compare the group average of females against that of males, whereas in the pooled condition they believed that their scores would be averaged with those of all other participants (both male and female). Only the performance of females in the segregated condition was impacted by the gender difference manipulation (see Figure 1).

Although this research was the first to directly test the role of evaluation in stereotype threat, Wout, Danso, Jackson, and Spencer (2008) also argued that stereotype threat impacts performance when only group evaluation is possible. However, in their research the treatment of participants in the control group differed from the treatment of participants in the group-evaluation condition in several ways: females in the group-evaluation condition were told that the task was diagnostic, whereas females in the control condition was told that the task was non-diagnostic; females in the group-evaluation condition were told that their individual scores were to be averaged, whereas females in the non-diagnostic condition were told nothing about how their individual scores would be treated; and females in the group-evaluation condition were told that the average scores of males would be compared to the average scores of females, whereas females in the non-diagnostic condition were told nothing about a gender comparison. Of course, Wout et al. (2008) intended to look at only the effect of the last

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

difference (comparing average scores as a function of gender), but the other differences could also have impacted the pattern of findings. In the current research the only difference between the pooled/gender difference condition and the segregated/gender difference condition was whether the scores were pooled or segregated. Thus the current research isolates the effect of one of Shapiro and Neuberg's (2007) six core threats: Group-Reputation Threat (Outgroup).

This form of threat is particularly interesting when put in the context of previous research that has examined the effect of evaluation on group performance. Harkins and Szymanski (1989) have demonstrated that the potential for group-level evaluation motivates performance but, in this minimal groups paradigm, it was only the potential for the participants themselves to compare the groups that motivated performance, not the potential for evaluation by the experimenter. That is, participants were told that they were a member of a group, the average for which would be compared to the average of another group. Matters were arranged such that individual evaluation was not possible, either by self or by experimenter. Under these conditions participants were motivated by the potential to compare their group's average to that of another group, but not by the potential for the experimenter to make this group-level comparison.

In contrast, in the current research participants in the segregated condition were motivated by the fact that the experimenter would be able to compare the average female score against the average male score, even though they could not. This finding is consistent with Marx, Stapel, and Muller's (2005) research, which argues that stigmatized individuals experience an increase in the salience of group identity (i.e., "we-ness") in stereotype threat situations, as well as Schmader's (2002) research that indicates females' gender identity moderates the impact of stereotype threat on performance. In sum, the current demonstrates the importance of group identity to the experience of stereotype threat, an effect that is not seen in the minimal groups paradigm.

Although the current research went to lengths to isolate Group-Reputation Threat (Outgroup), this type of threat is only one of six core threats proposed by Shaprio and Neuberg's multi-threat framework. Previous research (Inzlicht & Ben-Zeev, 2003; Wout et al., 2008) has also produced stereotype threat performance effects under Self-Concept Threat conditions, in which the self is both the target and source of the threat. However, research has yet to examine the four remaining core threats: Group-Reputation Threat (Ingroup) (source: ingroup member; target: group), Own-Reputation Threat (Outgroup) (source: outgroup member, target: self), Own-Reputation Threat (Ingroup) (source: ingroup member, target: self), and Group-Concept Threat (source: self; target: group).

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

Future studies should seek to isolate these other forms of stereotype threat as a more comprehensive test of the multi-threat framework model.

Before this work, the field had yet to test whether the potential for evaluation was necessary for the experience of stereotype threat. That is, researchers assumed that stigmatized individuals feared confirming the negative stereotype to some meaningful audience. However, it was possible that stigmatized individuals were threatened simply by performing a stereotyped task and being targets of the stereotype. In other words, the experience of threat may have been rooted solely in the knowledge of one's stereotyped status and diagnosticity of the task at hand. The research presented here, however, illustrates that the performance of stigmatized individuals is only impacted by a manipulation of stereotype threat when an evaluative source can judge performance.

Manuscript received 28 September 2009
Manuscript accepted 17 November 2009
First published online 16 February 2010

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010

REFERENCES

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. *Journal of Experimental Social Psychology*, *41*, 174–181.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, *40*, 247–255.

Brown, R. P., & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, *39*, 626–633.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effect of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, *28*, 659–670.

Harkins, S. G., & Szymanski, K. (1989). Social loafing and group evaluation. *Journal of Personality and Social Psychology*, *56*, 934–941.

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*, 365–371.

Inzlicht, M., & Ben-Zeev, T. (2003). Do high achieving female students underperform in private? The implications of threatening environments on intellectual processing. *Journal of Educational Psychology*, *95*, 796–805.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. *Journal of Personality and Social Psychology*, *93*, 544–564.

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on quantitative GRE problems: A mere effort interpretation. *Personality and Social Psychology Bulletin*, *35*, 1301–1314.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin*, *29*, 371–381.

Kirk, R. (1995). *Experimental design*. Pacific Grove, CA: Brooks/Cole.

Marx, D. M., & Stapel, D. A. (2006). Distinguishing stereotype threat from priming effects: On the role of the social self and threat-based concerns. *Journal of Personality and Social Psychology*, *91*, 243–254.

Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: The interplay of construal orientation and social comparison under threat. *Journal of Personality and Social Psychology*, *88*, 432–446.

McKay, P. F., Doverspike, D., Bowen-Hilton, D., & Martin, Q. D. (2002). Stereotype threat effects on the Raven Advanced Progressive Matrices scores of African Americans. *Journal of Applied Social Psychology*, *32*, 767–787.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, *29*, 782–789.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, *38*, 195–201.

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

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356.

Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications for a multi-threat framework for causes, moderators, mediators, consequences, and interventions. *Personality and Social Psychology Review*, *11*, 107–130.

Smith, J. L., & White, P. H. (2001). Development of the domain identification measure: A tool for investigating stereotype threat effects. *Educational and Psychological Measurement*, *61*, 1040–1057.

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

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

Steele, C. M., Spencer, S. J., & Aronson, J. A. (2002). Contending with group image: The psychology of stereotype and social identity threat. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 379–440). New York: Academic Press.

Stone, J., & McWhinnie, C. (2008). Evidence that blatant versus subtle stereotype threat cues impact performance through dual processes. *Journal of Experimental Social Psychology*, *44*, 445–452.

Wout, D., Danso, H., Jackson, J., & Spencer, S. (2008). The many faces of stereotype threat: Group- and self-threat. *Journal of Experimental Social Psychology*, *44*, 792–799.

Downloaded By: [Jamieson, Jeremy P.] At: 14:47 17 February 2010