# On the Continued Misinterpretation of Stereotype Threat as Accounting for Black-White Differences on Cognitive Tests

**Dana H. Tomeh[1] and Paul R. Sackett[1]**

1. University of Minnesota - Twin Cities

ABSTRACT

Steele and Aronson (1995) showed that stereotype threat affects the test performance of stereotyped groups. A careful reading shows that threat affects test performance but does not eliminate Black–White mean score gaps. Sackett et al. (2004) reviewed characterization of this research in scholarly articles, textbooks, and popular press, and found that many mistakenly inferred that removing stereotype threats eliminated the Black–White performance gap. We examined whether the rate of mischaracterization of Steele and Aronson had decreased in the 15 years since Sackett et al. highlighted the common misinterpretation. We found that the misinterpretation rate dropped from 90.9% to 62.8% in journal articles and from 55.6% to 41.18% in textbooks, though this is only statistically significant in journal articles.

KEYWORDS

stereotype threat, cognitive ability, employment testing, admissions testing

Stereotype threat is a widely studied phenomenon, first examined by Steele and Aronson (1995). The central idea is that in a situation in which a stereotype of a group to which one belongs becomes salient, concerns about being judged according to that stereotype arise and inhibit performance. Steele and Aronson hypothesized that high stakes testing in employment and higher education admission settings is such a situation. There are stereotypes about women's performance in math and stereotypes about racial/ethnic group performance on cognitive tests (such as the verbal ability domain examined by Steele and Aronson). It is posited that when the stereotype is made salient to test takers, concerns about the stereotype consume attentional resources and result in lower test scores than would be observed absent stereotype activation. The phenomenon is widely studied; a recent meta-analysis by Shewach, Sackett, and Quint (2019) found over 200 studies of threat effects on cognitive tests in adult samples.

Steele and Aronson (1995) examined whether stereotype activation affected the test performance of Black students relative to White college students. They placed Black and White students into either a threat or a nonthreat condition and gave them the same set of questions from the Graduate Record Examination (GRE). In the nonthreat condition, the students were told that the questions were a problem-solving task that was not diagnostic of ability. In the threat condition, the participants were told that the questions were a test of verbal ability. After controlling for prior SAT scores, Black students in the threat condition performed worse than Black students in the nonthreat condition, worse than White students in the threat condition, and comparably to White students in the non-threat condition.

This experimental design, used by Steele and Aronson and many subsequent researchers, permits two types of comparisons. The first is within-subgroup comparisons: comparing Black performance in different conditions (e.g., threat vs. no threat conditions). This has been the primary focus in meta-analyses of the stereotype literature (e.g., Shewach et al., 2019; Nguyen & Ryan, 2008). The standardized mean difference (d) between minority group per-

Corresponding author:
Dana H. Tomeh
Author Email: tomeh001@umn.edu

formance in a threat condition versus a non-threat or threat removal condition has been the information extracted from each study and subjected to meta-analysis. The second is between-subgroup conditions, such as comparing majority and minority group mean scores in threat conditions or in non-threat conditions. Between-group comparisons are of interest to researchers with an interest in mean differences between groups, with the central question being the degree to which the presence of absence of stereotype threat affects these mean differences.

Thus, both between-group and within-group comparisons are of interest. It is the between-group comparisons reported by Steele and Aronson (1995) that are of interest to us here, as they are commonly misinterpreted. In the highly visible journal *American Psychologist*, Sackett, Hardison, and Cullen (2004) pointed out the source of misinterpretation, namely, that many readers either ignore or fail to understand the meaning of the fact that the experiment controlled for prior SAT scores. Sackett et al. showed that many interpret the concept of no subgroup difference in the non-threat condition, controlling for prior SAT scores as if it means that subgroup differences have been eliminated in the non-threat condition. The key is understanding the logic of statistical control. Conceptually, Steele and Aronson were asking "given a Black and a White student with the same prior SAT score, what subsequent test performance would be observed in threat vs. non-threat conditions." A finding of no subgroup difference in the non-threat condition means that Black and White students with comparable prior SAT scores perform comparably in a non-threat experimental condition, and finding lower Black student performance in the threat condition than the non-threat condition shows that threat does lower Black student performance in the experimental setting.

Steele and Aronson (1995) reported that the mean prior SAT scores for the Black and White students in their Study 2 differed by just over half a standard deviation (mean of 655 for White students and 603 for Black students on a test with an *SD* of 100). Thus, the reported finding of "no difference in the non-threat condition" does not mean that the Black–White mean difference is eliminated when threat is not present but rather that the mean difference is the same as would be expected given the mean difference on the prior SAT.

Sackett et al. (2004) showed that 90.9% of journal articles and 55.6% of textbooks that described the Steele and Aronson (1995) study incorrectly interpreted the findings as showing that subgroup differences were eliminated in the non-threat condition (we note that Steele and Aronson did not misinterpret their findings; the error is made by others). Sackett et al. included a number of examples; we repeat one here:

> Steele and Aronson (1995) found that when Black and White college students were given a difficult test of verbal ability presented as a diagnostic test of intellectual ability, Blacks performed more poorly on the tests than Whites. However, in another condition, when the exact same test was presented as simply a laboratory problem-solving exercise, Blacks performed equally as well as Whites on the test. One simple adjustment to the situation (changing the description of the test) eliminated the performance differences between Whites and Blacks. (Wolfe & Spencer, 1996, p. 180)

This characterization of the results is not only incorrect but, in our opinion, dangerous. Finding mean differences between Blacks and Whites on cognitive measures is pervasive. The magnitude of these differences averages around 1.0 standard deviation; see Roth et al. (2001) for a meta-analysis of mean differences in employment and educational settings. This is widely viewed as a crucial societal problem, and much research and large investment in interventions (e.g., early childhood programs such as Head Start) have been devoted to understanding causes and developing interventions aimed at the problem (Sackett et al., 2001). An incorrect belief that these differences can be eliminated simply by changing the instructional set under which a test is taken may lead to the reduction of effort and resources aimed at this critical problem.

It has been more than 15 years since Sackett et al.'s (2004) paper, and we set out to address the question of whether there has been a change in how the Steele and Aronson (1995) results have been described in subsequent research. We examine the peer-reviewed literature and a set of introductory psychology, organizational behavior, and human resource management textbooks to see whether or not there has been a decrease in the mischaracterization of the study.

We note that Warne, Astle, and Hill (2018) conducted a related study, focusing solely on psychology textbooks. Their focus was on mischaracterizations stereotype threat findings more generally, rather than on the Steele and Aronson study. They concluded that 9/13 or 69.2% of textbooks mischaracterized stereotype threat. We note that they categorized a study as a mischaracterization if it concluded that threat accounts for at least part of subgroup differences. We see a disconnect in their argumentation: They used Sackett et al. (2004)'s conclusion that threat did not reduce subgroup differences in Steele and Aronson's work as the basis for an overall conclusion about stereotype threat. Other work may indeed show some effect. A recent meta-analysis of over 200 studies by Shewach, Sackett, and Quint (2019) reported that minority group test takers scored an average of .31 *SD* lower in threat conditions across all studies, with this difference reduced to .14 *SD* in studies with features expected in operational testing conditions, and further re-

duced when applying corrections for publication bias. This work suggests that threat may make a small contribution to group differences. Thus, although we would view a textbook statement that removing threat eliminates group differences as a clear error, a statement that it may contribute in part is one we would not view as an error.

There are also other procedural differences between our work and Warne et al.'s. They focused exclusively on sections of textbooks focused on intelligence, whereas we looked at psychology textbooks in their entirety. We found that threat is at times treated in discussions of intelligence, and other times in chapters on social psychology. They focused on a text's broad treatment of threat, whereas our focus was limited to characterization of Steele and Aronson (1995). Thus, on conceptual and procedural grounds we do not believe that meaningful comparisons between Warne et al.'s. and our work are possible.

## METHOD AND RESULTS

We systematically reviewed journal articles, introductory psychology, organizational behavior (OB), and human resource (HR) textbooks that discuss the results of Steele and Aronson (1995). We focused on the 15-year period from 2005 to 2019. We first examined articles and textbooks to determine whether they discussed the results of Steele and Aronson in enough detail to determine whether the conclusions drawn were correct or not; only those with enough detail were included in our analysis. Each article or textbook was initially examined either by the first author or by one of two undergraduate research assistants, who made two determinations: (a) whether there was enough detail to permit an evaluation, and (b) if yes, whether Steele and Aronson's work was accurately or inaccurately described. These initial coding decisions were then reviewed by the first author and the second author, who is the senior researcher on the team. There was 96% agreement between the first and second authors' determinations. Consensus was reached in the case of discrepancies. We coded an article as a misinterpretation either if the article explicitly stated that Black and White students performed the same when threat was removed or implied that the absence of stereotype threat eliminates the racial differences in intelligence scores (ex. "demonstrated that African American college students performed more poorly than European American students on the same set of difficult verbal problems when they believed the goal was to assess their intellectual ability rather than to explore ways to enhance their learning skills."; Krendl et al., 2012). Some examples of both mistaken and correct interpretations are included in the Appendix.

### Journal Articles

A Google Scholar search was done to identify journal articles published between 2005 and 2019 that cite Steele and Aronson (1995). This search resulted in 9,150 articles. This number is very large, and because a full article manual search was needed (i.e., not just a search of abstracts) to ascertain what was said in the article about Steele and Aronson, we concluded that an exhaustive search was not feasible given our resources (e.g., at 5 minutes per article, the task would require 692 hours). We decided to include the first 10 articles we located in a given year that reported enough information to make a decision regarding whether the interpretation was correct or not. The vast majority of articles cite Steele and Aronson as a general reference to stereotype threat and do not describe the studies in detail. We ended up manually examining 2,065 articles in order to obtain a sample of 150 articles, 10 per year, that describe the Steele and Aronson findings. Of the final pool of 150 articles, we observed that 56 articles discussed only within-group findings (e.g., comparing Black student test performance under threat and non-threat conditions). In all of these studies the findings were described accurately. This is consistent with Sackett et al. (2004), who concluded that mischaracterization was limited to between-group comparisons and thus focused on within-group comparisons. Of the 94 studies that discussed between-group findings (e.g., comparing Black and White student test performance), 59 (62.8%, 95% CI 50.4% to 75.1%) were inaccurate. It is in these between-group comparisons that the issue of the implications of controlling for a prior test score are important.

Importantly, the Google Scholar article list that we searched listed articles within each year in the order of number of citations that the article had received to date. Thus, our identified articles are not a random sampling of articles but rather reflect the most influential scholarship on the topic. We made a judgment that it was reasonable to do this, as it sheds light on errors in work that is being used; an argument can be made that errors in uncited work are minimally consequential. We conducted a logistic regression analysis using year of publication and citation count as predictors of an accurate/inaccurate description of the Steele and Aronson (1995). The citation counts are highly skewed, with a mean of 131.27 (SD = 216.58) and a median of 50.

Five articles received over 500 citations; the most cited article received 1580 citations. To address skew in the data, we recoded the five articles with over 500 citations to a value of 500. Of the five articles with high citation counts, three incorrectly characterized Steele and Aronson (1995), whereas two articles correctly described the findings. Citation count did not have a significant relationship with whether Steele and Aronson (1995) was characterized correctly (coded 1) versus incorrectly (coded 0; regression coefficient of .003; $p$ = .15). Thus, influential scholarship does not appear to be related to whether or not Steele and Aronson (1995) was mischaracterized. Additional analyses

also showed that year of publication has a statistically nonsignificant coefficient of .041 ($p$ = .48).

To further address the lack of random sampling, we conducted several analyses to examine other factors. Each article was coded for journal discipline, journal h-index, article content (whether the article focused specifically on stereotype threat or not), and article type (whether the article was an experiment/meta vs. other nonempirical articles).

The correlation between journal h-index and whether an article correctly or incorrectly characterized Steele and Aronson (1995) was nonsignificant ($r$ = .03, $p$ = .77), indicating that there is no relationship with the publishing journal's influence. Likewise, the correlation examining whether article content (was the article focusing stereotype threat [coded 1] or not [coded 0]) related to whether or not Steele and Aronson was correctly characterized was not significant ($r$ = .17, $p$ = .11). Another correlation was used to examine the relationship between article type (coded 1 if the article was an experiment/meta-analysis and 0 if not) and the characterization of Steele and Aronson (1995). This was also nonsignificant ($r$ = -.02, $p$ = .82), indicating there is no relationship between the type of article and mischaracterization of Steele and Aronson (1995).

Finally, a logistic regression was conducted to examine whether the discipline of a journal (psychology, math, education, business, etc.) was related to correctly or incorrectly characterizing Steele and Aronson. Each of the articles was coded into one of six groups: psychology ($N$ = 51), education/ educational psychology ($N$ = 13), business and economics ($N$ = 6), law ($N$ = 4), healthcare, medicine, gerontology ($N$ = 7), and other ($N$ = 12). This information was then dummy coded and used as predictors. The results of the logistic regression showed that there is no significant difference in error rate between articles in psychology journals and any other journal discipline.

To supplement the findings, we examined whether the same authors were consistently mischaracterizing Steele and Aronson. Two-hundred and forty-four unique authors were listed on the articles included in the analyses, disregarding author order. Sixteen authors were associated with more than one paper, again irrespective of author order. Fourteen of them authored two papers, one person authored three papers, and one person authored four. Seven of these authors consistently correctly described the findings, four of them consistently incorrectly described the findings, and five described the findings both correctly and incorrectly on separate occasions. The four authors that consistently incorrectly described the findings accounted for only six unique articles.

We assessed whether there was a significant decrease in the proportion of authors who incorrectly described the results of Steele and Aronson (1995) in the current study, as compared to the findings reported by Sackett et al. (2004). A one-sample z-test was conducted to examine whether the proportions in our sample significantly differed from the population findings of Sackett et al. (2004). The percentage of authors mischaracterizing the results of Steele and Aronson between 2005 and 2019 (62.8) is significantly smaller than the percentage who mischaracterized it in 2004 or earlier (90.9%; $p < .001$). To supplement this, we examined whether the proportion of articles that incorrectly characterized Steele and Aronson each year decreased over time. The regression results did not show a significant relationship between time and the proportion of articles that mischaracterize the findings ($p$ = .57).

A one-sided Fisher's exact test was also conducted to determine whether articles that cited Sackett et al. were more likely to correctly characterized Steele and Aronson's (1995) findings. Of the 17 articles found that cite Sackett et al. (2004), 6 (35.3%) mischaracterize characterize Steele and Aronson's findings. The remaining 77 articles do not cite Sackett et al.; of these 52 (67.5%) incorrectly characterize the findings. For those journal articles that cite Sackett et al., the proportion of articles incorrectly characterizing the results of Steele and Aronson is significantly smaller than the proportion of journal articles that did not cite Sackett et al. and mischaracterized the results ($p$ = .015).

**Textbooks**

As in Sackett, et al. (2004), psychology textbooks were collected from the psychology department at a local university where they had been sent to be considered for course adoption. We note that this is a major research institution, and one of the largest universities in the U.S., with a median composite ACT score at the 88th percentile for admitted students, and with several thousand students per year enrolled in introductory psychology. Publishers are eager to have their texts considered for adoption given class size. Thus, this is not a random sampling of textbooks but rather a reasonably comprehensive collection of texts written for use at such schools. In addition, we accessed introductory psychology textbooks that were available at no cost online. We were able to gather 63 introductory psychology textbooks.

In order to collect introductory HR and OB books, all textbooks in our university's business school Human Resources and Industrial Relations library were examined. In addition, we searched freely accessible online versions of HR and OB textbooks. This process coincidentally resulted in 38 introductory HR textbooks and introductory 38 introductory OB textbooks.

Surprisingly, no HR texts treat stereotype threat and only two OB texts do so; neither of these cite Steele and Aronson. Of the 63 introductory psychology textbooks found, 25 discussed the results of Steele and Aronson (1995) in enough depth to determine whether the results were correctly characterized or not. Of these, 17 address between-group

comparisons. Seven of the 17, or 41.18%, inaccurately characterized the results of Steele and Aronson. The proportion of textbooks misinterpreting Steele and Aronson in the present study was not found to be statistically significantly smaller than the 55.6% figure reported in 2004 by Sackett et al. ($p$ =.18).

A two-sided Fisher's exact test was conducted to examine whether journal articles are more likely to mischaracterize Steele and Aronson (1995) than textbooks. The Fisher's exact test showed that there is no statistically significant difference ($p$ = .79) between the proportion of textbooks mischaracterizing Steele and Aronson (41.18%) and the proportion of journal articles mischaracterizing the results (62.8%).

## DISCUSSION

Sackett et al. (2004) clearly outlined the common misinterpretations of Steele and Aronson (1995) in a notable journal. Since then, the error rate in journal articles (62.8%) has significantly decreased, whereas the error rate in textbooks (41.18%) has not. Although there is a significant decrease in mischaracterization in journal articles, a large proportion of publications are still mischaracterizing the results, and we can only speculate as to the reasons why this continues to be a problem at all. In fact, these reasons do not differ from the speculations presented in the original review of the literature (Sackett et al., 2004). The first is that some people who discuss Steele and Aronson (1995) might not have noticed that the performance of the participants was adjusted for SAT. Sackett et al., discuss the appeal of the misunderstood result (e.g., the elimination of group differences) and how that may lead people to overlook the adjustment. Testing and selection based on cognitive ability would be less controversial if it were possible to eliminate the race gap.

The second possibility is that the authors describing the results of Steele and Aronson (1995) simply did not understand the importance of the adjustment for prior SAT scores. The adjustment fundamentally changes the results and their implications, but if authors did not understand the implications of the adjustment, they might not feel that it is necessary to mention or include.

The third possibility discussed in Sackett et al. (2004) is that the omission of the adjustment in the text or in a copy of the graph could be an accidental omission by authors who know the implications of adjustment. They note that even Aronson et al. (1999) omitted the reference to the SAT adjustment, despite the fact that most of Aronson's other publications do include reference to the adjustment.

A fourth possibility exists that was not discussed in Sackett et al. (2004). It became evident in the review that over time those who referenced the results of the seminal Steele and Aronson (1995) article were not always referencing the original article but rather another author who had previously discussed the results. If one never refers to the original article, and instead cites someone else who discusses the results, it is very possible that they would draw on results that were mistakenly characterized in the article that they are citing. This highlights the importance of referencing original sources in research.

That 35.3% of articles citing Sackett et al. (2004) still mischaracterized the Steele and Aronson (1995) findings was surprising to us. Review of those articles indicated that the treatment of Sackett et al. within the article was generally removed from the treatment of Steel and Aronson, and focused on a different aspect of Sackett et al. For example, Sackett et al. would be cited as part of a discussion of the merits of using a prior test as a control variable in stereotype threat research.

One other factor potentially affecting mischaracterization of Steele and Aronson (1995) is a general change over time in the methodological sophistication of researchers, such that completely independently of Sackett et al. (2004), researchers now more clearly understand the implications of statistical control for prior SAT scores. Aiken, West, and Millsap (2008) surveyed 200 graduate programs in psychology about the quantitative and methodological training provided to graduate students, contrasting results with an earlier survey by Aiken, West, Sechrest, and Reno (1990). Although there was no change over time in the amount of quantitative and methodological training provided, there were shifts in topic focus. More time was devoted to multiple regression in the time period reflected in the more recent survey. As interpreting effects of one variable with other variables controlled is at the heart of multiple regression, the result may be an increase in recognizing the implications of statistical control.

Should there be any question as to whether this mischaracterization has consequences outside discussions within the academic community, we provide some examples of applied situations in which it is relevant. It is common for various individuals or groups to submit amicus curiae ("friend of the court") briefs in an attempt to influence the Supreme Court in their decisions. In a case challenging race conscious college admissions at the University of Texas, a brief from the National Black Law Students Association (2015) mischaracterized Steele and Aronson's (1995) findings as showing the group mean differences were eliminated when threat was removed. We found a similar mischaracterization in a brief from the American Sociological Association (2003) in another case challenging race conscious admission at the University of Michigan. There is no ready way to ascertain whether these briefs influenced the court, but their existence does indicate that incorrect interpretations are not merely a self-contained issue within academia.

This paper calls out the mischaracterization of import-

ant work, which allows us to begin to address the problem. The statistically significantly reduced mischaracterization since the 2004 publication of Sackett et al. is positive; however, the mischaracterization that remains is not. In following up on the previous examination of mischaracterization, we hope to continue to shed light on the problem such that mischaracterization becomes much less common.

**Limitations**

The large number of publications discussing Steele and Aronson (1995) since the publication of Sackett et al. (2004) would have been impractical to examine in its entirety, and as such we sampled 10 articles from each year. Despite the empirical support presented for the sampling procedure, it is important to note that these articles were not randomly sampled from the population but rather represent the first 10 relevant articles in a given year as returned by a Google Scholar search that presented articles in order of the number of citations received. This indicates that there may be less frequently cited articles that also discuss Steele and Aronson's findings. However, more cited articles generally appear in prestigious journals and as such undergo more rigorous review, so it is possible that our findings understate the proportion of misinterpretation.

We also acknowledge that this present research is narrowly focused on the single issue of mischaracterization of the results of a seminal article. However, broader critical examinations of the stereotype threat literature are available elsewhere. Shewach et al. (2019) is a meta-analysis of the threat literature that is not simply a descriptive summary but a critical analysis that leads to the conclusion that stereotype threat effects found in lab settings are reduced to near-zero when conditions expected in high-stakes testing settings are present (e.g., when candidates are motivated to perform well; when tests are scores in the ways used in operational testing settings, namely, number right scoring, as opposed to the proportion correct among items attempted that is used with some frequency in the stereotype threat literature). Sackett and Ryan (2011) offer a critical evaluation of work purporting to show that stereotype threat effects generalize to real-world settings.

**Conclusion**

In the 15 years since the publication of Sackett et al.'s (2004) paper highlighting the high rates of misinterpretation of Steele and Aronson's (1995) work on stereotype threat, researchers (62.8%) and textbook authors (41.18%) have continued to mischaracterize the results. Since its publication, there has been a nominal and statistically significant decrease in the proportion of journal articles misrepresenting the results of Steele and Aronson. This decrease is a significant step in the right direction. However, given the visibility of the publication outlet, and the implications of the misunderstanding, one would hope for a much smaller proportion of errors, particularly in textbooks. Perhaps this paper will serve as a second treatment dose and contribute to reduced mischaracterization.

## REFERENCES

* indicate journal articles included in the analyses
** indicate textbooks included in the analyses


*Abrams, D., Eller, A., & Bryant, J. (2006). An age apart: The effects of intergenerational contact and stereotype threat on performance and intergroup bias. Psychology and Aging, 21(4), 691–702. https://doi.org/10.1037/0882-7974.21.4.691

*Adiredja, A. P., & Andrews-Larson, C. (2017). Taking the sociopolitical turn in postsecondary mathematics education research. International Journal of Research in Undergraduate Mathematics Education, 3(3), 444–465. https://doi.org/10.1007/s40753-017-0054-5

Aiken, L. S., West, S. G., & Millsap, R. E. (2008). Doctoral training in statistics, measurement, and methodology in psychology: Replication and extension of Aiken, West, Sechrest, and Reno's (1990) survey of PhD programs in North America. American Psychologist, 63(1), 32–50. https://doi.org/10.1037/0003-066X.63.1.32

Aiken, L. S., West, S. G., Sechrest, L., Reno, R. R., Roediger III, H. L., Scarr, S., Kazdin, A. E., & Sherman, S. J. (1990). Graduate training in statistics, methodology, and measurement in psychology: A survey of PhD programs in North America. American Psychologist, 45(6), 721. https://doi.org/10.1037/0003-066X.45.6.721

*Alter, A. L., Aronson, J., Darley, J. M., Rodriguez, C., & Ruble, D. N. (2010). Rising to the threat: Reducing stereotype threat by reframing the threat as a challenge. Journal of Experimental Social Psychology, 46(1), 166–171. https://doi.org/10.1016/j.jesp.2009.09.014

*Andretta, J. R., Worrell, F. C., Ramirez, A. M., Barnes, M. E., Odom, T., Brim, S., & Woodland, M. H. (2015). The effects of stigma priming on forensic screening in African American youth. Counseling Psychologist, 43(8), 1162–1189. https://doi.org/10.1177/0011000015611963

*Appel, M., & Kronberger, N. (2012). Stereotypes and the achievement gap: Stereotype threat prior to test taking. Educational Psychology Review, 24(4), 609–635. https://doi.org/10.1007/s10648-012-9200-4

*Armstrong, B., Gallant, S. N., Li, L., Patel, K., & Wong, B. I. (2017). Stereotype threat effects on older adults' episodic and working memory: A meta-analysis. Gerontologist, 57(suppl_2), S193–S205. https://doi.org/10.1093/geront/gnx056

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When White men can't do math: Necessary and sufficient factors in stereotype threat. Journal of Experimental Social Psychology, 35(1), 29–46. https://doi.org/10.1006/jesp.1998.1371

Ballen, C. J., Wieman, C., Salehi, S., Searle, J. B., & Zamudio, K. R. (2017). Enhancing diversity in undergraduate science: Self-efficacy drives performance gains with active learn-

ing. CBE—Life Sciences Education, 16(4), ar56. https://doi.org/10.1187/cbe.16-12-0344

Barber, S. J. (2017). An examination of age-based stereotype threat about cognitive decline: Implications for stereotype-threat research and theory development. Perspectives on Psychological Science, 12(1), 62–90. https://doi.org/10.1177/1745691616656345

*Beasley, M. A., & Fischer, M. J. (2012). Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors. Social Psychology of Education, 15(4), 427–448. https://doi.org/10.1007/s11218-012-9185-3

*Bedyńska, S., & Żołnierczyk-Zreda, D. (2015). Stereotype threat as a determinant of burnout or work engagement. Mediating role of positive and negative emotions. International Journal of Occupational Safety and Ergonomics, 21(1), 1–8. https://doi.org/10.1080/10803548.2015.1017939

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: Can skills that don't rely heavily on working memory still be threatened? Personality and Social Psychology Bulletin, 32(8), 1059–1071. https://doi.org/10.1177/0146167206288489

*Bermudez, G. (2018). The social dreaming matrix as a container for the processing of implicit racial bias and collective racial trauma. International Journal of Group Psychotherapy, 68(4), 538–560. https://doi.org/10.1080/00207284.2018.1469957

*Boschini, A., Muren, A., & Persson, M. (2012). Constructing gender differences in the economics lab. Journal of Economic Behavior & Organization, 84(3), 741–752. https://doi.org/10.1016/j.jebo.2012.09.024

Breedlove, S. M. (2015). Principles of psychology. Sinauer Associates.

Brief for the American Sociological Association, et al., as Amici Curiae in Support of Respondents, Grutter v. Bollinger, 539 U.S. 306 (2003) (No. 02-241)

Brief for the National Black Law Students Association as Amicus Curiae in Support of Respondents, Fisher v. University of Texas at Austin, et al.,, 570 U.S. 297 (2013) (No. 14-981).

Brown, R. P., & Day, E. A. (2006). The difference isn't black and white: Stereotype threat and the race gap on raven's advanced progressive matrices. Journal of Applied Psychology, 91(4), 979–985. https://doi.org/10.1037/0021-9010.91.4.979

**Cacioppo, J. T., & Freberg, L. (2019). Discovering psychology: The science of mind. Cengage Learning.

*Cartledge, G., & Kourea, L. (2008). Culturally responsive classrooms for culturally diverse students with and at risk for disabilities. Exceptional Children, 74(3), 351–371. https://doi.org/10.1177/001440290807400305

**Cervone, D. (2015). Psychology: The science of person, mind and brain. Worth Publishers.

Chalabaev, A., Sarrazin, P., Fontayne, P., Boiché, J., & Clément-Guillotin, C. (2013). The influence of sex stereotypes and gender roles on participation and performance in sport and exercise: Review and future directions. Psychology of Sport and Exercise, 14(2), 136–144. https://doi.org/10.1016/j.psychsport.2012.10.005

*Chang, D. F., & Demyan, A. L. (2007). Teachers' stereotypes of Asian, Black, and White students. School Psychology Quarterly, 22(2), 91–114. https://doi.org/10.1037/1045-3830.22.2.91

Charity Hudley, A. H., & Mallinson, C. (2017). "It's worth our time": A model of culturally and linguistically supportive professional development for K-12 STEM educators. Cultural Studies of Science Education, 12(3), 637–660. https://doi.org/10.1007/s11422-016-9743-7

*Chasteen, A. L., Bhattacharyya, S., Horhota, M., Tam, R., & Hasher, L. (2005). How feelings of stereotype threat influence older adults' memory performance. Experimental Aging Research, 31(3), 235–260. https://doi.org/10.1080/03610730590948177

Chen, Y., Li, S. X., Liu, T. X., & Shih, M. (2014). Which hat to wear? Impact of natural identities on coordination and cooperation. Games and Economic Behavior, 84, 58–86. https://doi.org/10.1016/j.geb.2013.12.002

*Christian, J., Turner, R., Holt, N., Larkin, M., & Cotler, J. H. (2014). Does intergenerational contact reduce ageism? When and how contact interventions actually work? Journal of Arts and Humanities, 1, 15.

*Clark, J. K., Eno, C. A., & Guadagno, R. E. (2011). Southern discomfort: The effects of stereotype threat on the intellectual performance of US southerners. Self and Identity, 10(2), 248–262. https://doi.org/10.1080/15298861003771080

*Clark, J. K., Thiem, K. C., Barden, J., Stuart, J. O., & Evans, A. T. (2015). Stereotype validation: The effects of activating negative stereotypes after intellectual performance. Journal of Personality and Social Psychology, 108(4), 531–552. https://doi.org/10.1037/a0038887

**Coon, D., Mitterer, J. O., & Martini, T. (2018). Psychology: Modules for active learning. Cengage Learning.

*Cooper, J. (2006). The digital divide: The special case of gender: The digital divide. Journal of Computer Assisted Learning, 22(5), 320–334. https://doi.org/10.1111/j.1365-2729.2006.00185.x

Cortland, C. I., & Kinias, Z. (2019). Stereotype threat and women's work satisfaction: The importance of role models. Archives of Scientific Psychology, 7(1), 81–89. https://doi.org/10.1037/arc0000056

Cromley, J. G., Perez, T., Wills, T. W., Tanaka, J. C., Horvat, E. M., & Agbenyega, E. T.-B. (2013). Changes in race and sex stereotype threat among diverse STEM students: Relation to grades and retention in the majors. Contemporary Educational Psychology, 38(3), 247–258. https://doi.org/10.1016/j.cedpsych.2013.04.003

*Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. Journal of Applied Social Psychology, 38(6), 1639–1655. https://doi.org/10.1111/j.1559-1816.2008.00362.x

*Dauphinais, K. A. (2005). Valuating and nurturing multiple intelligences in legal education: A paradigm shift. Washington & Lee Race and Ethnic Ancestry Law Journal, 11(1), 43. Available at SSRN: https://ssrn.com/abstract=889193

*Davies, P. G., Spencer, S. J., & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. Journal of Per-

sonality and Social Psychology, 88(2), 276–287. https://doi.org/10.1037/0022-3514.88.2.276

*Davis, C., Aronson, J., & Salinas, M. (2006). Shades of threat: Racial identity as a moderator of stereotype threat. Journal of Black Psychology, 32(4), 399–417. https://doi.org/10.1177/0095798406292464

Dee, T. S. (2014). Stereotype threat and the student-athlete. Economic Inquiry, 52(1), 173–182. https://doi.org/10.1111/ecin.12006

Delgado, A. R., & Prieto, G. (2008). Stereotype threat as validity threat: The anxiety–sex–threat interaction. Intelligence, 36(6), 635–640. https://doi.org/10.1016/j.intell.2008.01.008

Derks, B., Inzlicht, M., & Kang, S. (2008). The neuroscience of stigma and stereotype threat. Group Processes & Intergroup Relations, 11(2), 163–181. https://doi.org/10.1177/1368430207088036

*Désert, M., Préaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven's progressive matrices. European Journal of Psychology of Education, 24(2), 207–218. https://doi.org/10.1007/BF03173012

*Desombre, C., Anegmar, S., & Delelis, G. (2018). Stereotype threat among students with disabilities: The importance of the evaluative context on their cognitive performance. European Journal of Psychology of Education, 33(2), 201–214. https://doi.org/10.1007/s10212-016-0327-4

*Desombre, C., Jury, M., Bagès, C., & Brasselet, C. (2019). The distinct effect of multiple sources of stereotype threat. Journal of Social Psychology, 159(5), 628–641. https://doi.org/10.1080/00224545.2018.1544540

*Earp, B. D. (2010). Automaticity in the classroom: Unconscious mental processes and the racial achievement gap. Journal of Multiculturalism in Education, 6(1), 1–22.

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. Archives of Sexual Behavior, 41(3), 557–570. https://doi.org/10.1007/s10508-011-9875-5

*Foy, S. L. (2018). Challenges from and beyond symptomatology: Stereotype threat in young adults with ADHD. Journal of Attention Disorders, 22(3), 309–320. https://doi.org/10.1177/1087054715590159

*Freeman, L. (2015). Phenomenology of racial oppression. Knowledge Cultures, 3(1), 24–44.

*Freeman, L. (2017). Embodied harm: A phenomenological engagement with stereotype threat. Human Studies, 40(4), 637–662. https://doi.org/10.1007/s10746-017-9438-4

*Froehlich, L., Martiny, S. E., Deaux, K., Goetz, T., & Mok, S. Y. (2016). Being smart or getting smarter: Implicit theory of intelligence moderates stereotype threat and stereotype lift effects. British Journal of Social Psychology, 55(3), 564–587. https://doi.org/10.1111/bjso.12144

*Gaines, S. O., Bagha, S., Barrie, M., Bhattacharjee, T., Boateng, Y., Briggs, J., Ghezai, H., Gunnoo, K., Hoque, S., Merchant, D., Mehra, K., Noorkhan, N., & Rodriques, L. (2012). Impact of experiences with racism on African-descent persons' susceptibility to stereotype threat within the United Kingdom. Journal of Black Psychology, 38(2), 135–152. https://doi.org/10.1177/0095798411407065

**Gazzaniga, M. S., Heatherton, T. F., & Halpern, D. F. (2016). Psychological science (5th ed.). W. W. Norton & Company.

*Gendler, T. S. (2011). On the epistemic costs of implicit bias. Philosophical Studies, 156(1), 33–63. https://doi.org/10.1007/s11098-011-9801-7

*Gentile, A., Boca, S., & Giammusso, I. (2018). 'You play like a Woman!' Effects of gender stereotype threat on Women's performance in physical and sport activities: A meta-analysis. Psychology of Sport and Exercise, 39, 95–103. https://doi.org/10.1016/j.psychsport.2018.07.013

*Gillespie, J. Z., Converse, P. D., & Kriska, S. D. (2010). Applying recommendations from the literature on stereotype threat: Two field studies. Journal of Business and Psychology, 25(3), 493–504. https://doi.org/10.1007/s10869-010-9178-1

Gleitman, H., Gross, J., & Reisberg. (2011). Psychology (8th ed.). W. W. Norton & Company.

Good, J. J., Woodzicka, J. A., & Wingfield, L. C. (2010). The effects of gender stereotypic and counter-stereotypic textbook images on science performance. The Journal of Social Psychology, 150(2), 132–147. https://doi.org/10.1080/00224540903366552

Gray, P., O., & Bjorklund, D., F. (2018). Psychology (8th ed.). Worth Publishers.

*Grimm, L. R., Markman, A. B., Maddox, W. T., & Baldwin, G. C. (2009). Stereotype threat reinterpreted as a regulatory mismatch. Journal of Personality and Social Psychology, 96(2), 288–304. https://doi.org/10.1037/a0013463

**Grison, S., Heatherton, T. F., & Gazzaniga, M. S. (2015). Psychology in your life. Norton.

Guarneri, J. A., Oberleitner, D. E., & Connolly, S. (2019). Perceived stigma and self-stigma in college students: A literature review and implications for practice and research. Basic and Applied Social Psychology, 41(1), 48–62. https://doi.org/10.1080/01973533.2018.1550723

*Gupta, A., Szymanski, D. M., & Leong, F. T. L. (2011). The "model minority myth": Internalized racialism of positive stereotypes as correlates of psychological distress, and attitudes toward help-seeking. Asian American Journal of Psychology, 2(2), 101–114. https://doi.org/10.1037/a0024183

Harackiewicz, J. M., Canning, E. A., Tibbetts, Y., Giffen, C. J., Blair, S. S., Rouse, D. I., & Hyde, J. S. (2014). Closing the social class achievement gap for first-generation students in undergraduate biology. Journal of Educational Psychology, 106(2), 375–389. https://doi.org/10.1037/a0034679

Harrison, C. K., Stone, J., Shapiro, J., Yee, S., Boyd, J. A., & Rullan, V. (2009). The role of gender identities and stereotype salience with the academic performance of male and female college athletes. Journal of Sport and Social Issues, 33(1), 78–96. https://doi.org/10.1177/0193723508328902

*Harvey, M., Novicevic, M. M., Buckley, M. R., & Fung, H. (2005). Reducing inpatriate managers' 'Liability of Foreignness' by addressing stigmatization and stereotype threats. Journal of World Business, 40(3), 267–280. https://doi.org/10.1016/j.jwb.2005.05.004

*Hoff, K., & Pandey, P. (2005). How belief systems and mistrust shape responses to economic incentives. Economics of Transition, 13(3), 445–472.

Hoff, K., & Pandey, P. (2014). Making up people—The effect of

identity on performance in a modernizing society. Journal of Development Economics, 106, 118–131. https://doi.org/10.1016/j.jdeveco.2013.08.009

*Hoff, K., & Walsh, J. (2018). The whys of social exclusion: Insights from behavioral economics. The World Bank Research Observer, 33(1), 1–33. https://doi.org/10.1093/wbro/lkx010

Horton, S., Baker, J., & Deakin, J. M. (2007). Stereotypes of aging: Their effects on the health of seniors in North American society. Educational Gerontology, 33(12), 1021–1035. https://doi.org/10.1080/03601270701700235

Horton, S., Baker, J., Pearce, G. W., & Deakin, J. M. (2008). On the malleability of performance: Implications for seniors. Journal of Applied Gerontology, 27(4), 446–465. https://doi.org/10.1177/0733464808315291

*Howard, S., & Borgella, A. (2018). "Sinking" or sinking?: Identity salience and shifts in Black women's athletic performance. Psychology of Sport and Exercise, 39, 179–183. https://doi.org/10.1016/j.psychsport.2018.08.016

*Hoyt, C. L., & Murphy, S. E. (2016). Managing to clear the air: Stereotype threat, women, and leadership. The Leadership Quarterly, 27(3), 387–399. https://doi.org/10.1016/j.leaqua.2015.11.002

**Huffman, K. (2012). Psychology in action. Wiley.

*Jacksch, V., & Klehe, U.-C. (2016). Unintended consequences of transparency during personnel selection: Benefitting some candidates, but harming others?: International Journal of Selection and Assessment, 24(1), 4–13. https://doi.org/10.1111/ijsa.12124

*Jones, P. R. (2017). Reducing the impact of stereotype threat on women's math performance: Are two strategies better than one? Electronic Journal of Research in Education Psychology, 9(24), 587–616. https://doi.org/10.25115/ejrep.v9i24.1458

*Jones, P. R., Taylor, D. M., Dampeer-Moore, J., Van Allen, K. L., Saunders, D. R., Snowden, C. B., & Johnson, M. B. (2013). Health-related stereotype threat predicts health services delays among Blacks. Race and Social Problems, 5(2), 121–136. https://doi.org/10.1007/s12552-013-9088-8

*Jordan, A. H., & Lovett, B. J. (2007). Stereotype threat and test performance: A primer for school psychologists. Journal of School Psychology, 45(1), 45–59. https://doi.org/10.1016/j.jsp.2006.09.003

*Jordan, W. J. (2010). Defining equity: Multiple perspectives to analyzing the performance of diverse learners. Review of Research in Education, 34(1), 142–178. https://doi.org/10.3102/0091732X09352898

*Jussim, L., Crawford, J. T., Anglin, S. M., Stevens, S. T., & Duarte, J. L. (2016). Interpretations and methods: Towards a more effectively self-correcting social psychology. Journal of Experimental Social Psychology, 66, 116–133. https://doi.org/10.1016/j.jesp.2015.10.003

Kalat, J. W. (2005). Introduction to psychology (7th ed.). Wadsworth.

*Kalokerinos, E. K., von Hippel, C., & Zacher, H. (2014). Is stereotype threat a useful construct for organizational psychology research and practice? Industrial and Organizational Psychology, 7(3), 381–402. https://doi.org/10.1111/iops.12167

*Kang, J., & Banaji, M. R. (2006). Fair measures: A behavioral realist revision of "Affirmative Action." California Law Review, 94(4), 1063. https://doi.org/10.2307/20439059

Kaplan, J. M. (2015). Race, IQ, and the search for statistical signals associated with so-called "X"-factors: Environments, racism, and the "hereditarian hypothesis." Biology & Philosophy, 30(1), 1–17. https://doi.org/10.1007/s10539-014-9428-0

Kaplan, S., & Berman, M. G. (2010). Directed attention as a common resource for executive functioning and self-regulation. Perspectives on Psychological Science, 5(1), 43–57. https://doi.org/10.1177/1745691609356784

Kellow, J. T., & Jones, B. D. (2008). The effects of stereotypes on the achievement gap: Reexamining the academic performance of African American high school students. Journal of Black Psychology, 34(1), 94–120. https://doi.org/10.1177/0095798407310537

*Kessler, J. B., & Milkman, K. L. (2018). Identity in charitable giving. Management Science, 64(2), 845–859. https://doi.org/10.1287/mnsc.2016.2582

King, A. M. (2011). Stereotype threat and self-perceptions: The impact on college students. Journal of Student Affairs at New York University, 7, 12.

**King, L. A. (2014). The Science of Psychology (3rd ed.). McGraw-Hill.

*Kinias, Z., & Sim, J. (2016). Facilitating women's success in business: Interrupting the process of stereotype threat through affirmation of personal values. Journal of Applied Psychology, 101(11), 1585–1597. https://doi.org/10.1037/apl0000139

**Kosslyn, S. M., & Rosenberg, R. S. (2006). Psychology in context (3rd ed.). Pearson Education.

**Kowalski, R. M., & Westen, D. (2011). Psychology. Wiley & Sons.

*Krendl, A., Gainsburg, I., & Ambady, N. (2012). The effects of stereotypes and observer pressure on athletic performance. Journal of Sport and Exercise Psychology, 34(1), 3–15. https://doi.org/10.1123/jsep.34.1.3

*Kumi Yeboah, A., & Smith, P. (2016). Relationships between minority students online learning experiences and academic performance. Online Learning, 20(4). https://doi.org/10.24059/olj.v20i4.577

Lamberton, C. (2019). Toward a dignity architecture: The critical challenges of stigmatized-identity cues for consumer psychology. Journal of Consumer Psychology, 29(1), 152–159. https://doi.org/10.1002/jcpy.1077

*Laurin, R. (2013). Stereotype threat and lift effects in motor task performance: The mediating role of somatic and cognitive anxiety. The Journal of Social Psychology, 153(6), 687–699. https://doi.org/10.1080/00224545.2013.821098

*Lee, S.-Y., Olszewski-Kubilius, P., & Peternel, G. (2009). Follow-up with students after 6 years of participation in project EXCITE. Gifted Child Quarterly, 53(2), 137–156. https://doi.org/10.1177/0016986208330562

*Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. Sex Roles, 54(1–2), 113–125. https://doi.org/10.1007/s11199-005-8873-2

*Levy, D. J., Heissel, J. A., Richeson, J. A., & Adam, E. K. (2016). Psychological and biological responses to race-based

social stress as pathways to disparities in educational outcomes. American Psychologist, 71(6), 455–473. https://doi.org/10.1037/a0040322

Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., & Smithson, M. (2015). Seepage: Climate change denial and its effect on the scientific community. Global Environmental Change, 33, 1–13. https://doi.org/10.1016/j.gloenvcha.2015.02.013

Lilienfeld, S. O., Lynn, S. J., Namy, L. L., & Woolf, N. J. (2014). Psychology: From inquiry to understanding (3rd ed.). Pearson/Allyn & Bacon.

Linos, E. (2018). More than public service: A field experiment on job advertisements and diversity in the police. Journal of Public Administration Research and Theory, 28(1), 67–85. https://doi.org/10.1093/jopart/mux032

Major, B., & O'Brien, L. T. (2005). The social psychology of stigma. Annual Review of Psychology, 56(1), 393–421. https://doi.org/10.1146/annurev.psych.56.091103.070137

Maloney, E. A., Schaeffer, M. W., & Beilock, S. L. (2013). Mathematics anxiety and stereotype threat: Shared mechanisms, negative consequences and promising interventions. Research in Mathematics Education, 15(2), 115–128. https://doi.org/10.1080/14794802.2013.797744

*Manly, J., & Echemendia, R. (2007). Race-specific norms: Using the model of hypertension to understand issues of race, culture, and education in neuropsychology. Archives of Clinical Neuropsychology, 22(3), 319–325. https://doi.org/10.1016/j.acn.2007.01.006

*Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affirmation on women's intellectual performance. Journal of Experimental Social Psychology, 42(2), 236–243. https://doi.org/10.1016/j.jesp.2005.04.010

Marx, D. M., & Goff, P. A. (2005). Clearing the air: The effect of experimenter race on target's test performance and subjective experience. British Journal of Social Psychology, 44(4), 645–657. https://doi.org/10.1348/014466604X17948

*McClain, R., A. (2018). Bottled at the source: Recapturing the essence of academic support as a primary tool of education equity for minority law students. University of Maryland Law Journal of Race, Religion, Gender and Class, 18(1), 139–188.

McConnell, A. R., & Brown, C. M. (2010). Dissonance averted: Self-concept organization moderates the effect of hypocrisy on attitude change. Journal of Experimental Social Psychology, 46(2), 361–366. https://doi.org/10.1016/j.jesp.2009.10.019

*McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. Journal of Applied Developmental Psychology, 27(5), 486–493. https://doi.org/10.1016/j.appdev.2006.06.003

*McGlone, M. S., & Pfiester, R. A. (2015). Stereotype threat and the evaluative context of communication. Journal of Language and Social Psychology, 34(2), 111–137. https://doi.org/10.1177/0261927X14562609

*Meador, A. (2018). Examining recruitment and retention factors for minority STEM majors through a stereotype threat lens: Minority STEM major. School Science and Mathematics, 118(1–2), 61–69. https://doi.org/10.1111/ssm.12260

Mesa, V., Wladis, C., & Watkins, L. (2014). Research problems in community college mathematics education: Testing the boundaries of K-12 research. Journal for Research in Mathematics Education, 45(2), 173–192. https://doi.org/10.5951/jresematheduc.45.2.0173

Mickelson, R. A., & Nkomo, M. (2012). Integrated schooling, life course outcomes, and social cohesion in multiethnic Democratic societies. Review of Research in Education, 36(1), 197–238. https://doi.org/10.3102/0091732X11422667

Moule, J. (n.d.). Acknowledging our possible biases and working together openly is essential for developing community in our schools. 8.

Mukerji, C. (2010). The territorial state as a figured world of power: Strategics, logistics, and impersonal rule. Sociological Theory, 28(4), 402–424. https://doi.org/10.1111/j.1467-9558.2010.01381.x

Nairne, J. S. (2014). Psychology (6th ed.). Wadsworth.

*Najdowski, C. J., Bottoms, B. L., & Goff, P. A. (2015). Stereotype threat and racial differences in citizens' experiences of police encounters. Law and Human Behavior, 39(5), 463–477. https://doi.org/10.1037/lhb0000140

Newcombe, N. S., Ambady, N., Eccles, J., Gomez, L., Klahr, D., Linn, M., Miller, K., & Mix, K. (2009). Psychology's role in mathematics and science education. American Psychologist, 64(6), 538–550. https://doi.org/10.1037/a0014813

Nguyen, H.-H. D., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. Journal of Applied Psychology, 93(6), 1314–1334. https://doi.org/10.1037/a0012702

*Olson, K., & Jacobson, K. (2015). Cross-cultural considerations in pediatric neuropsychology: A review and call to attention. Applied Neuropsychology: Child, 4(3), 166–177. https://doi.org/10.1080/21622965.2013.830258

Onwuachi-Willig, A., & Fricke, A. (2011). Class, classes, and classic race-baiting: What's in definition. Denver University Law Review, 88(4).

*Orom, H., Semalulu, T., & Underwood, W. (2013). The social and learning environments experienced by underrepresented minority medical students: A narrative review. Academic Medicine, 88(11), 1765–1777. https://doi.org/10.1097/ACM.0b013e3182a7a3af

**Passer, M. W., & Smith, R. E. (2011). Psychology: The science of mind and behavior. McGraw-Hill Higher Education.

Pearson, B. Z., Conner, T., & Jackson, J. E. (2013). Removing obstacles for African American English-speaking children through greater understanding of language difference. Developmental Psychology, 49(1), 31–44. https://doi.org/10.1037/a0028248

*Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. PLOS ONE, 11(1), e0146487. https://doi.org/10.1371/journal.pone.0146487

*Pennington, C. R., Litchfield, D., McLatchie, N., & Heim, D. (2019). Stereotype threat may not impact women's inhibitory control or mathematical performance: Providing support for the null hypothesis. European Journal of Social Psychology,

49(4), 717–734. https://doi.org/10.1002/ejsp.2540

Phuong, A. E., Nguyen, J., & Marie, D. (2017). Conceptualizing an adaptive and data-driven equity-oriented pedagogy. 10(2), 20.

*Pruysers, S., & Blais, J. (2017). Why won't Lola run? An experiment examining stereotype threat and political ambition. Politics & Gender, 13(02), 232–252. https://doi.org/10.1017/S1743923X16000544

*Quinn, K. A., & Rosenthal, H. E. S. (2012). Categorizing others and the self: How social memory structures guide social perception and behavior. Learning and Motivation, 43(4), 247–258. https://doi.org/10.1016/j.lmot.2012.05.008

*Ratan, R., & Sah, Y. J. (2015). Leveling up on stereotype threat: The role of avatar customization and avatar embodiment. Computers in Human Behavior, 50, 367–374. https://doi.org/10.1016/j.chb.2015.04.010

*Rędzio, A. (2019). Gender stereotypes concerning mathematical abilities: Threat or challenge? Studia Ad Didacticam Mathematicae, 273, 157–171.

*Roberson, L., & Kulik, C. T. (2007). Stereotype threat at work. Academy of Management Perspectives, 21(2), 24–40. https://doi.org/10.5465/amp.2007.25356510

Roberts, S. O. (2011). Reconsidering parental involvement: Implications for Black parents. Online Publication of Undergraduate Studies.

*Rogers-Sirin, L., & Sirin, S. R. (2009). Cultural competence as an ethical requirement: Introducing a new educational model. Journal of Diversity in Higher Education, 2(1), 19–29. https://doi.org/10.1037/a0013762

Rosenthal, H. E. S., & Crisp, R. J. (2007). Choking under pressure: When an additional positive stereotype affects performance for domain identified male mathematics students. European Journal of Psychology of Education, 22(3), 317–326. https://doi.org/10.1007/BF03173429

*Rosenthal, H. E. S., Crisp, R. J., & Suen, M.-W. (2007). Improving performance expectancies in stereotypic domains: Task relevance and the reduction of stereotype threat. European Journal of Social Psychology, 37(3), 586–597. https://doi.org/10.1002/ejsp.379

*Rosenthal, L., & Lobel, M. (2011). Explaining racial disparities in adverse birth outcomes: Unique sources of stress for Black American women. Social Science & Medicine, 72(6), 977–983. https://doi.org/10.1016/j.socscimed.2011.01.013

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. Personnel Psychology, 54(2), 297–330. https://doi.org/10.1111/j.1744-6570.2001.tb00094.x

*Ryan, K. E., & Ryan, A. M. (2005). Psychological processes underlying stereotype threat and standardized math test performance. Educational Psychologist, 40(1), 53–63. https://doi.org/10.1207/s15326985ep4001_4

Rydell, R. J., McConnell, A. R., & Beilock, S. L. (2009). Multiple social identities and stereotype threat: Imbalance, accessibility, and working memory. Journal of Personality and Social Psychology, 96(5), 949–966. https://doi.org/10.1037/a0014846

Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotype threat as accounting for African American-White differences on cognitive tests. American Psychologist, 59(1), 7–13. https://doi.org/10.1037/0003-066X.59.1.7

Sackett, P. R., & Ryan, A. M. (2011). Concerns about generalizing stereotype threat research findings to operational high-stakes testing settings. In Stereotype Threat (pp. 246-259). Oxford University Press.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. American Psychologist, 56(4), 302–318. https://doi.org/10.1037/0003-066X.56.4.302

Schacter, D. L., Gilbert, D. T., Nock, M. K., & Wegner, D. M. (2018). Introducing Psychology (4th ed.). Worth Publishers.

*Schmader, T., & Croft, A. (2011). How stereotypes stifle performance potential: Stereotype threat. Social and Personality Psychology Compass, 5(10), 792–806. https://doi.org/10.1111/j.1751-9004.2011.00390.x

*Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. Psychological Review, 115(2), 336–356. https://doi.org/10.1037/0033-295X.115.2.336

*Scholl, J. M., & Sabat, S. R. (2008). Stereotypes, stereotype threat and ageing: Implications for the understanding and treatment of people with Alzheimer's disease. Ageing and Society, 28(1), 103–130. https://doi.org/10.1017/S0144686X07006241

*Schwarz, K. A., Pfister, R., & Büchel, C. (2016). Rethinking explicit expectations: Connecting placebos, social cognition, and contextual perception. Trends in Cognitive Sciences, 20(6), 469–480. https://doi.org/10.1016/j.tics.2016.04.001

*Seider, S., Clark, S., & Graves, D. (2020). The development of critical consciousness and its relation to academic achievement in adolescents of color. Child Development, 91(2). https://doi.org/10.1111/cdev.13262

*Selm, K. R., Peterson, M. N., Hess, G. R., Beck, S. M., & McHale, M. R. (2019). Educational attainment predicts negative perceptions women have of their own climate change knowledge. PLOS ONE, 14(1), e0210149. https://doi.org/10.1371/journal.pone.0210149

Sevo, R., & Chubin, D. E. (n.d.). Bias literacy: A Review of concepts in research on discrimination. 31.

Shapiro, J. R., & Neuberg, S. L. (2007). From stereotype threat to stereotype threats: Implications of a multi-threat framework for causes, moderators, mediators, consequences, and interventions. Personality and Social Psychology Review, 11(2), 107–130. https://doi.org/10.1177/1088868306294790

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. Sex Roles, 66(3–4), 175–183. https://doi.org/10.1007/s11199-011-0051-0

Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely versus unlikely in operational test settings: A meta-analysis. Journal of Applied Psychology, 104(12), 1514–1534. https://doi.org/10.1037/apl0000420

*Smith, C. S., & Hung, L.-C. (2008). Stereotype threat: Effects on education. Social Psychology of Education, 11(3), 243–257.

https://doi.org/10.1007/s11218-008-9053-3

*Smith, J. S., Estudillo, A. G., & Kang, H. (2011). Racial differences in eighth grade students' identification with academics. Education and Urban Society, 43(1), 73–90. https://doi.org/10.1177/0013124510379403

Speight, S. L. (2007). Internalized racism: One more piece of the puzzle. The Counseling Psychologist, 35(1), 126–134. https://doi.org/10.1177/0011000006295119

*Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype Threat. Annual Review of Psychology, 67(1), 415–437. https://doi.org/10.1146/annurev-psych-073115-103235

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. Journal of Personality and Social Psychology, 69(5), 797–811.

Steinecke, A., Beaudreau, J., Bletzinger, R. B., & Terrell, C. (2007). Race-neutral admission approaches: Challenges and opportunities for medical schools. Academic Medicine, 82(2), 117–126. https://doi.org/10.1097/ACM.0b013e31802d85bd

Stricker, L. J., Rock, D. A., & Bridgeman, B. (2015). Stereotype threat, inquiring about test takers' race and gender, and performance on low-stakes tests in a large-scale assessment: Stereotype threat. ETS Research Report Series, 2015(1), 1–12. https://doi.org/10.1002/ets2.12046

Syed, M., Azmitia, M., & Cooper, C. R. (2011). Identity and academic success among underrepresented ethnic minorities: An interdisciplinary review and integration. Journal of Social Issues, 67(3), 442–468. https://doi.org/10.1111/j.1540-4560.2011.01709.x

*Teichman, D. (2005). Sex, shame, and the law: An Economic perspective on Megan's laws. Harvard Journal on Legislation, 42, 63.

*Tempel, T., & Neumann, R. (2014). Stereotype threat, test anxiety, and mathematics performance. Social Psychology of Education, 17(3), 491–501. https://doi.org/10.1007/s11218-014-9263-9

Thames, A. D., Hinkin, C. H., Byrd, D. A., Bilder, R. M., Duff, K. J., Mindt, M. R., Arentoft, A., & Streiff, V. (2013). Effects of stereotype threat, perceived discrimination, and examiner race on neuropsychological performance: Simple as Black and White? Journal of the International Neuropsychological Society, 19(5), 583–593. https://doi.org/10.1017/S1355617713000076

*Tomasetto, C., & Appoloni, S. (2013). A lesson not to be learned? Understanding stereotype threat does not protect women from stereotype threat. Social Psychology of Education, 16(2), 199–213. https://doi.org/10.1007/s11218-012-9210-6

*Trinkner, R., Kerrison, E. M., & Goff, P. A. (2019). The force of fear: Police stereotype threat, self-legitimacy, and support for excessive force. Law and Human Behavior, 43(5), 421–435. https://doi.org/10.1037/lhb0000339

*Tropp, L. R., Smith, A. E., & Crosby, F. J. (2007). The use of research in the Seattle and Jefferson County desegregation cases: Connecting social science and the law. Analyses of Social Issues and Public Policy, 7(1), 93–120. https://doi.org/10.1111/j.1530-2415.2007.00149.x

Urdan, T., & Bruchmann, K. (2018). Examining the academic motivation of a diverse student population: A consideration of methodology. Educational Psychologist, 53(2), 114–130. https://doi.org/10.1080/00461520.2018.1440234

*Walton, G. M., & Spencer, S. J. (2009). Latent ability: Grades and test scores systematically underestimate the intellectual ability of negatively stereotyped students. Psychological Science, 20(9), 1132–1139. https://doi.org/10.1111/j.1467-9280.2009.02417.x

*Walton, G. M., Spencer, S. J., & Erman, S. (2013). Affirmative meritocracy. Social Issues and Policy Review, 7(1), 1–35. https://doi.org/10.1111/j.1751-2409.2012.01041.x

Wang, C. S., Whitson, J. A., Anicich, E. M., Kray, L. J., & Galinsky, A. D. (2017). Challenge your stigma: How to reframe and revalue negative stereotypes and slurs. Current Directions in Psychological Science, 26(1), 75–80. https://doi.org/10.1177/0963721416676578

Warne, R. T., Astle, M. C., & Hill, J. C. (2018). What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. Archives of Scientific Psychology, 6(1), 32. https://doi.org/10.1037/arc0000038

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. Journal of Personality and Social Psychology, 89(5), 696–716. https://doi.org/10.1037/0022-3514.89.5.696

*Williams, D. P., Joseph, N., Hill, L. K., Sollers, J. J., Vasey, M. W., Way, B. M., Koenig, J., & Thayer, J. F. (2019). Stereotype threat, trait perseveration, and vagal activity: Evidence for mechanisms underpinning health disparities in Black Americans. Ethnicity & Health, 24(8), 909–926. https://doi.org/10.1080/13557858.2017.1378803

Wolfe, C. T., & Spencer, S. J. (1996). Stereotypes and prejudice: Their overt and subtle influence in the classroom. American Behavioral Scientist, 40(2), 176–185. https://doi.org/10.1177/0002764296040002008

Woodcock, A., Hernandez, P. R., Estrada, M., & Schultz, P. W. (2012). The consequences of chronic stereotype threat: Domain disidentification and abandonment. Journal of Personality and Social Psychology, 103(4), 635–646. https://doi.org/10.1037/a0029120

Worrell, F. C. (2014). Theories school psychologists should know: Culture and academic achievement. Psychology in the Schools, 51(4), 332–347. https://doi.org/10.1002/pits.21756

*Wout, D. A., Shih, M. J., Jackson, J. S., & Sellers, R. M. (2009). Targets as perceivers: How people determine when they will be negatively stereotyped. Journal of Personality and Social Psychology, 96(2), 349–362. https://doi.org/10.1037/a0012880

RECEIVED 09/16/2021 ACCEPTED 02/08/2022

Appendix

**Examples of Incorrect and Correct Characterizations of Steele and Aronson (1995)**

Examples of Incorrect Characterizations from Journal Articles

"In Steele and Aronson's original study, they randomly assigned Black and white students to complete items similar to those on the Graduate Record Exam after being told that the test was either a diagnostic of their verbal abilities or a measure of their problem-solving strategies. Both groups performed equally when they were told that the task was a measure of their problem-solving strategies, but Blacks performed worse when they were told that the task was a diagnostic of their verbal abilities" – Orom, Semalulu, & Underwood (2013), p. 1769

"Social psychology research has demonstrated that Black students perform poorly on standardized tests when race is perceived as a salient contextual factor. When race is not emphasized Blacks perform as well as White students (Steele & Aronson, 1995)." – Bermudez (2018), p.4

"Additionally, minority respondents may have self-reported lower levels of knowledge due to cultural stereotypes of intellectual ability, similar to those that plague women. For example, Steele and Aronson [24] found that black students only underperform compared to white students when they believe their intellectual ability is being tested. " – Selm et al. (2019), p.6

Examples of Correct Characterizations from Journal Articles

"In support of this hypothesis, their experiments revealed that African American college students performed worse than their White peers on standardized test questions when this task was described to them as being diagnostic of their verbal ability but that their performance was equivalent to that of their White peers when the same questions were simply framed as an exercise in problem solving (and after accounting for prior SAT scores)" – Schmader et al. (2008), p. 2.

"In a seminal set of studies, Steele and Aronson (1995) found that Blacks performed worse than Whites when stereotypes about intellectual ability were activated prior to taking a test (e.g., reporting one's race). However, when the stereotype was less salient (e.g., not reporting race before the test), the performance of Black and White students was equivalent when controlling for previous scores on the Scholastic Aptitude Test (SAT)." – Clark et al. (2015), p. 532

"Stereotype threat was first described by Steele and Aronson (1995) in their study of how this socio-psychological notion affected the intellectual performance of African Americans. Steele and Aronson assigned African American and White students of similar intellectual abilities to two different groups. For one group a stereotype threat was introduced while the other served as a control. Each group was given a diagnostic exam of intellectual ability. Steele and Aronson found that academic performance of the African American students was significantly lower than their White counterparts when a stereotype threat was perceived." – Meador (2018), p. 63

Examples of Incorrect Characterizations from Textbooks

"In the ability condition, Black students performed worse than European Americans. In the problem solving condition, they performed the same as European Americans." – Coon, Mitterer, & Martini (2018), p. 636

"When a test is presented to Black and European American students who have first simply checked a box indicating their ethnicity, the Black Students perform more poorly. When attention is not drawn to ethnicity, no differences in performance emerged" – King (2014), p. 436

Examples of Correct Characterizations from Textbooks

"One group of students, the stereotype threat group, was told that the test accurately assessed their cognitive abilities, a statement that was designed to stimulate thinking about possible racial stereotypes about intelligence. The other group was told that the test was a routine laboratory procedure. Subsequently, the performance of Black and White students on the test

was compared. After controlling for Scholastic Aptitude test (SAT) scores, the black students in the stereotype threat group performed more poorly than did the white students. In contrast, in the nonthreatening condition, Blacks and Whites with equal SAT scores performed similarly on GRE items" – Cacioppo and Freberg (2019), p. 520

“Even when the researchers controlled statistically for preexisting ethnic group difference in verbal ability by using students’ college-entrance SAT scores, the Black–White performance difference on the experimental task was far greater if the students thought that the task measured intelligence than if they were told it was unrelated to intelligence.” – Passer (2011), p. 359-360