Comment

Contents

Steele & Aronson on Sackett et al. ..........47
Sackett et al. reply ............................48
Anseel et al. on Adair & Vohra ..........49
Green on Adair & Vohra ....................51
McCormick on Miller & Thoresen ........52
Richmond on Miller & Thoresen .........52
Rayburn on Miller & Thoresen ..........52
Kier & Davenport on Miller & Thoresen ...53
Miller & Thoresen reply ....................54

Sackett et al. (2004) reported an overstatement of the threat in the present instance and may foster misunderstandings of its own. The problem is this: In the universe of stereotype threat material that now includes over 100 articles and dissertations and more media mentions, Sackett et al. (2004) focus on the reporting of only a single experiment from the first published article on stereotype threat. This extremely narrow focus greatly exaggerates (a) that experiment’s impact on people’s understanding of stereotype threat and its role in the race gap, (b) the importance of particular aspects of the experiment, such as its use of covariance analysis, and (c) what its results say about the role of stereotype threat in real-life testing. We address these three issues in turn.

First, to know how important the mischaracterizations of Steele and Aronson’s (1995) Study 2 are, one has to know whether they actually distorted people’s understanding of stereotype threat’s role in the race gap. Sackett et al.’s (2004) review, however, is too narrow to answer this question.

In the literature and reporting on stereotype threat, there are many discussions of its role in the race gap that do not describe Study 2 in Steele and Aronson (1995)—many more than 29. By excluding these from their review, Sackett et al. (2004) never assess how often these discussions make the overclaim they worry about—that stereotype threat is the sole cause of the race gap. The overwhelming majority of these discussions, we believe, get it right; they depict stereotype threat as one of multiple causes of this gap. Or, when they do give stereotype threat a special importance, they are not referring to the race gap in the general population but to the race gap in some subsample in which African Americans and Whites have been equated on other factors that might affect their test performance—either statistically or by selection, as when selective colleges select members of both groups for having high academic skills and, thus, similar educational backgrounds. To the extent that the groups have been equated on other causes of the race gap, it may not be overclaiming to emphasize stereotype threat as a principal cause of any remaining gap (see Massey, Charles, Lundy, & Fischer, 2003).

Finally, accurate discussions of stereotype threat’s role might even be found in some of the articles in which they found mischaracterizations of the particular finding from Steele and Aronson (see Aronson et al., 1999).

In the literature on the race gap itself, Sackett et al. (2004) assessed neither the frequency with which stereotype threat is mentioned nor how often it was distorted when it was mentioned. Between the two of us, we have a respectable knowledge of this literature (e.g., Bowen & Bok, 1998; Jencks & Phillips, 1998; Massey et al., 2003). In these discussions where the focus is on explaining the race gap in large segments of the population, we have never seen a claim that stereotype threat is the gap’s sole cause. Whenever stereotype threat is considered, it is placed, where it belongs, as one of multiple causes of the gap. No attentive reader would come away from this literature with a different view.

Steele and Aronson (1995) is one of the few stereotype threat studies that focused on African Americans. This may be why Sackett et al. (2004) chose to focus on it so exclusively. But it is eight years old. Its characterization of stereotype threat can be checked against dozens of subsequent published stereotype threat studies and discussions. And it is a drop in an ocean of information about the race gap.

Thus, although trying to correct mischaracterizations of research findings can be worthwhile, the importance of doing so depends on evidence that the mischaracterizations in question are affecting general understandings. Without such evidence, one has to assume that the mischaracterizations are unnoticed, that they are understood to be what they are, mischaracterizations, or that, in other ways, evolving literatures have self-correcting capacities. Sackett et al. (2004) offered no systematic evidence that this particular mischaracterization has had any effect—personal communications notwithstanding. Moreover, the narrow design of their review distorts the picture of how stereotype threat and its
role in the race gap are understood in this literature and related material. A broader review would give them much less reason for concern.

Second, Sackett et al.’s (2004) narrow focus may have also led them to worry too much about the use of covariance analysis in Steele and Aronson’s (1995) study. They worried that this analysis led readers to believe that African Americans performed as well as Whites in the nondiagnostic (no stereotype threat) condition of that experiment, when, in fact, without this adjustment, they would be shown to perform still worse than Whites, as predicted by the group difference in their SATs. We, as much as Sackett et al., regret any confusion that this common analysis may have caused. We used it to reduce error variance and thus make the experiment more sensitive to the effect of conditions, especially in light of our small number of participants.

But again, the larger stereotype threat literature is critical. It shows the effect of stereotype threat on an array of tests—SATs, IQ tests, and French language tests to list only a few—sometimes with a covariance adjustment, but many times without. Whatever impression readers got from the use of covariance in Steele and Aronson (1995) would certainly have been corrected by this larger literature. They would know (a) that the skills measured by the SAT can indeed affect subsequent test performance, (b) that under common and important conditions, stereotype threat has powerful effects of its own on test performance, and (c) that detecting an effect of stereotype threat on test performance does not depend on the use of covariance analysis.

We note here that even in Study 2 of Steele and Aronson (1995), the effect of stereotype threat does not depend on the use of the SAT covariate. African Americans in the diagnostic (stereotype threat) condition performed a full standard deviation lower than African Americans in the nondiagnostic (no threat) condition—a 3-item effect on a 26-item test that was significant without the use of a covariate. Also, the interaction that tested whether the effect of stereotype threat was greater for African Americans than for Whites reached a one-way level of significance, $F = 3.75$, $p < .06$, with no covariate and only 10 participants per cell.

Third, Sackett et al. (2004) stated that absent stereotype threat, the African American–White difference is just what one would expect based on the African American–White difference in SAT scores, whereas in the presence of stereotype threat, the difference is larger than would be expected based on the difference in SAT scores. (p. 9)

They seem to be saying that the non-diagnostic (no stereotype threat) condition embodied the conditions of regular testing because it reproduced the African American–White difference observed on the regular SAT (i.e., no mean difference once adjusted for SATs) and that the diagnostic condition imposed an extra threat not typical of regular testing because it caused African Americans to perform worse than their SATs would have predicted.

However, seeing the pattern of African American–White differences in the nondiagnostic condition as more “expected” from SATs is, we believe, over-reading the data. The Graduate Record Examination (GRE) is correlated with the SAT, but not perfectly. And recall our small number of participants. Under these conditions—even under better conditions—SATs could not predict GREs so precisely. Thus, one cannot say which of the two African American–White differences—the threat difference or the no-threat difference—is best expected from the group difference in SATs, let alone which of the two conditions is most like regular testing.

Again, the larger literature is relevant. There (as in Steele & Aronson, 1995) it is the stereotype threat conditions, and not the no-threat conditions, that produce group differences most like those of real-life testing. Stereotype threat conditions represent the test as ability diagnostic, either en passant or by saying nothing at all and relying on participants to know a test when they see one. It is the no-threat conditions that are unlike real-life testing. They present the test as nondiagnostic of the participants’ ability or of their group’s ability—in stark contrast to real-life testing situations. Yet it is the stereotype threat condition that impair performance among the people who are subject to being negatively stereotyped (African Americans in the case of the Steele and Aronson experiments). The big picture, then, rather than guesses based on the pattern of results in a single experiment, should be used to judge which of these conditions—stereotype threat or no stereotype threat—is most like real-life testing.

Twenty-nine mischaracterizations of any research finding are 29 too many. However, using the frequency of these mischaracterizations to signal concern, while ignoring the large amount of information that would allay that concern, only furthers misunderstanding. Sackett et al. (2004) ignored the large number of discussions in the relevant literatures and media reports that do not overattribute the race gap to stereotype threat—discussions that vastly outnumber 29. Thus, rather than these mischaracterizations constituting a gathering danger, they are just mischaracterizations, almost completely ignored and having whatever misunderstanding they do cause constantly corrected by the natural progress of research.

REFERENCES


Correspondence concerning this comment should be addressed to Claude M. Steele, Department of Psychology, Stanford University, Jordan Hall, Building 420, Stanford, CA 94305-2130. E-mail: Steele@psych.stanford.edu

DOI: 10.1037/0003.066X.59.1.48

On the Value of Correcting Mischaracterizations of Stereotype Threat Research

Paul R. Sackett, Chaitra M. Hardison, and Michael J. Cullen
University of Minnesota, Twin Cities Campus

We see no disagreement by Steele and Aronson (2004, this issue) with the key issues that prompted our article (Sackett, Hardison, & Cullen, 2004, this issue). They