agree that it is a misinterpretation of the Steele and Aronson (1995) results to conclude that eliminating stereotype threat eliminates the African American–White test-score gap. They agree that we have identified multiple mischaracterizations of their work in media reports, journal articles, and textbooks, which wrongly interpret their work as finding that eliminating stereotype threat did indeed eliminate the score gap. They agree that these mischaracterizations are regrettable.

However, Steele and Aronson (2004) assert that there is no need to worry about mischaracterizations of their findings in the absence of evidence that these mischaracterizations have led to widespread misunderstanding of the role stereotype threat plays in explaining the African American–White test-score gap. We disagree. Although evidence of such misunderstanding would certainly be grave cause for concern, we believe it is sufficiently worrisome when one of the seminal studies on stereotype threat is commonly wrongly interpreted—by the popular media, textbook publishers, and academics alike—to mean that the African American–White test-score gap disappears when stereotype threat is eliminated. Steele and Aronson assert that their 1995 study is “a drop in an ocean of information about the race gap” (Steele & Aronson, 2004, p. 47). We believe they are unduly modest about the impact of their paper; that the Social Sciences Citation Index reports that it has been cited more than 300 times is one indicator of its prominence.

Steele and Aronson (2004) assert that because there are now over 100 research studies on stereotype threat, our focus on the first article on the topic results in a serious bias. However, they later acknowledge that their article is one of few stereotype threat studies focusing on African Americans. As the African American–White score gap was the topic of our article, we see our focus on this pivotal and highly cited article as entirely appropriate.

Steele and Aronson (2004) also assert that no attentive reader of the literature on the race gap would conclude that stereotype threat is its sole cause. However, our concern is with broader audiences than the serious scholar working on issues of race. We are concerned about students who are being initially exposed to issues of psychological testing and the race gap in their introductory psychology courses. We are concerned about managers responsible for personnel selection systems in their organizations. We are concerned about psychologists who do not follow testing issues closely and whose only exposure to stereotype threat may be through an American Psychological Association *Monitor on Psychology* column making the interpretive error that is the focus of our article. We are concerned about the large audience watching *Frontline* and hearing that the score gap is eliminated in the no-threat condition.

Steele and Aronson (2004) address the use of a prior SAT score as a covariate, claiming that we overworried about readers being misled by this analysis. They argue that a larger literature shows the stereotype threat effect, sometimes with the use of a prior test as a covariate and sometimes without. However, in our article, we noted clearly that we are not questioning the finding of a stereotype threat effect (i.e., the finding of a Race $\times$ Diagnostic Condition interaction) in Steele and Aronson (1995). Our concern is with misinterpreting the graphical presentation of findings as suggesting that group differences can be eliminated.

Steele and Aronson (2004) take issue with our comparison of African American–White differences on the prior SAT and on GRE-based scores in the two experimental conditions. Steele and Aronson assert that these are not comparable because the pretest SAT and the experimental GRE-based test are not perfectly correlated and because N is small. Given the extensive data on the similarity of the score gaps between the two tests and the correlation between the two, we see it as reasonable to posit that two groups that do not differ on the SAT would also be expected not to differ on the GRE.

We share with Steele and Aronson the beliefs that single experiments do not answer all questions and that it is important to examine the role of stereotype threat in real-life testing settings. We certainly agree with their position that evolving literatures have self-correcting capacities, and we view our article as fulfilling exactly such a role. Most crucially, we note that the disagreement between us is about the consequences of the mischaracterization we documented, not about whether the work has been mischaracterized.

REFERENCES

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,* 7–13.

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

Steele, C. M., & Aronson, J. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist, 59,* 47–48.

Correspondence concerning this comment should be addressed to Paul R. Sackett, Department of Psychology, University of Minnesota, Elliott Hall, 75 E. River Road, Minneapolis, MN 55455. E-mail: psackett@tc.umn.edu

DOI: 10.1037/0003.066X.59.1.49

Journal Impact Factors and Self-Citations: Implications for Psychology Journals

Frederik Anseel, Wouter Duyck, and Wouter De Baene
Ghent University

Marc Brysbaert
Royal Holloway University of London

Recently, Adair and Vohra (January 2003) analyzed changes in the number of references and citations in psychology journals as a consequence of the current knowledge explosion. In their study, the authors made a striking observation of the sometimes excessive number of self-citations in psychology journals. However, after this illustration, no further attention was paid to the issue of self-citation. This is unfortunate because little is known about self-citing practices in psychology. Early research on self-citations in psychology journals indicated that about 10% of citations were self-citations, and one author concluded that “it is apparent that controlling for self-citation is not necessary” (Gottfredson, 1978, p. 932). Similarly, although the *Publication Manual of the American Psychological Association* (American Psychological Association, 2001) provides clear guidelines on the form citations should take, it does not indicate when it is appropriate to cite one’s own work.

Recent figures urge more caution when dealing with self-citations. A multidisciplinary study found that 36% of all citations represent author self-citations (Aksnes, 2003; see also McGarty, 2000, for a similar finding in social psychology). Especially troublesome is the finding that self-citations peak during the first three years after publication, thereby strongly influencing impact factors of journals that are based on two-year periods.

Although the use of citation counts (and impact factors) has been criticized in all disciplines (see, e.g., Boor, 1982), it has become the main quantitative measure of the quality of a journal. Accordingly, these figures are used to make decisions about journal subscriptions, selection of journals for paper submission, rankings of authors, author tenure, individual grants, and even funding of entire research groups and institutions, insofar that researchers have expressed their concern about the current obsession of the academic world with impact factors (Lawrence, 2003).

Therefore, an important underexplored question is to what extent impact factors of psychology journals are artificially inflated or deflated by self-citations. There are good reasons to expect more self-citations from authors publishing in high-impact journals (e.g., *Psychological Bulletin*) than from authors publishing in low-impact journals (e.g., *Computers in Human Behavior*). This is because the former authors in general are more experienced and more successful. In addition, they are often part of a larger research group, so that they can coauthor more articles. For example, researchers who published in *Psychological Bulletin* during 1998–1999 had on average 2.83 publications in 2000, in comparison with 0.96 publications for authors who published in *Computers in Human Behavior*. As a consequence, these researchers have more opportunities to cite their own work. The main question, however, is whether this difference in self-citation rate is strong enough to affect the journal impact scores: Do journals with high citation counts get an extra "citation boost" due to self-citations?

We used the Institute for Scientific Information (ISI) databases Web of Science (WoS) and Journal Citation Reports (JCR) as the basis of our analyses. From each article (including empirical articles and literature reviews) in five high-, five middle-, and five low-ranked journals in psychology published in 1998 and 1999, we collected the number of self-citations and other-citations in 2000 from the WoS. All journals were included in either the *psychology* or *psychology, experimental* subsection of the JCR Social Science Edition 2000. Self-citations were assessed using the criterion that at least one author (first or coauthor) was also an author (first or coauthor) of the citing paper. On the basis of these data, we recalculated the impact factor for each journal with and without self-citations. As can be seen in Table 1, small discrepancies between the official impact factors (ISI) and our own calculations appeared. These discrepancies are caused by inaccuracies in the ISI databases and have been reported before ("Errors in Citation Statistics," 2002).

We found that articles in high-impact journals received more (M = 0.91) self-citations than articles in middle- (M = 0.44) and low-impact journals (M = 0.11), thus confirming our expectation that authors publishing in high-impact journals have more opportunities to cite their own work. To test whether these differences in numbers of self-citations change the differences in journal impact factors, we calculated the ratio of self-citations to the total number of citations for each article that received citations in 2000. A one-way analysis of variance was calculated with this ratio as a dependent variable and impact factor group (high, middle, low) as the independent variable. There was a significant effect of impact factor group, $F(2, 650) = 17.43$, $p < .001$, $\eta^2 = .05$. Tukey's post hoc analyses revealed that the high-impact factor group ($M = .20$, $SD = .02$) had a significantly lower ratio of self-citations to total citations, $p < .01$, than the other two groups, which did not differ from each other, $p = .85$ (low group $M = .39$, $SD = .05$; middle group $M = .36$, $SD = .02$). So, contrary to our expectations, the ratio of self-citations to total citations in high-impact journals was about half that of low-impact journals.

This is good news for the high-impact psychology journals: Compared with low- and middle-impact journals, their true citation counts are actually underestimated. When we adjusted our calculated impact factors for self-citations (see the last column of Table 1), we found that high-impact journals dropped on average by

Table 1

Citations, Self-Citations, and Different Impact Factors for Selected Psychology Journals

Impact factor group & journal	ISI impact factor	No. cited 2000	No. self-citations 2000	No. articles 1998–1999	Computed impact factor	Adjusted impact factor
Low-impact factor group						
Psychology	0.16 (15)	3	3	44	0.07 (15)	0.00 (15)
Journal of Economic Psychology	0.25 (14)	10	3	61	0.16 (14)	0.11 (13)
American Journal of Psychology	0.29 (13)	13	5	49	0.27 (13)	0.16 (12)
Computers in Human Behavior	0.33 (12)	41	11	78	0.53 (11)	0.38 (11)
Psychologica Belgica	0.42 (11)	8	6	25	0.32 (12)	0.08 (14)
Middle-impact factor group						
Animal Learning and Behavior	1.11 (10)	92	39	86	1.07 (10)	0.62 (10)
Journal of Motor Behavior	1.14 (9)	80	37	65	1.23 (7)	0.66 (9)
Acta Psychologica	1.27 (8)	120	35	101	1.19 (9)	0.84 (8)
Journal of Experimental Child Psychology	1.28 (7)	107	30	87	1.23 (8)	0.89 (7)
Journal of Experimental Psychology: Animal Behavior	1.37 (6)	131	40	74	1.77 (6)	1.23 (6)
High-impact factor group						
Journal of Cognitive Neuroscience	5.12 (5)	520	133	104	5.00 (5)	3.72 (5)
Annual Review of Psychology	5.85 (4)	272	26	47	5.79 (4)	5.23 (3)
Psychological Review	6.07 (3)	426	84	72	5.92 (3)	4.75 (4)
American Psychologist	6.86 (2)	740	60	119	6.22 (2)	5.71 (1)
Psychological Bulletin	6.91 (1)	459	69	69	6.65 (1)	5.65 (2)

Note. Journal rankings are reported in parentheses. ISI = Institute for Scientific Information.

15%, whereas middle- and low-impact journals dropped by 35% and 45%, respectively. As can be seen in Table 1, adjusting for self-citations also produced changes in journal rankings. For instance, adjusted impact factors indicate that the *American Psychologist* takes over the place of *Psychological Bulletin* as the most cited journal in psychology.

However, the good news for the high-impact journals should not make one forget that the journal impact scores are to a large extent determined by self-citations. This is problematic when the impact factors are used as measures of journal quality. A first way to improve the situation would be for editors and reviewers to keep the number of self-citations within limits. Editors already have expressed concern about the rising number of references and citations in psychology journals (Adair & Vohra, 2003), so a self-citing restriction would probably be welcomed. Unfortunately, the implementation of such a restriction may be difficult to achieve because it is not easy to determine where the border is between gratuitous self-citations and self-citations as a consequence of the cumulative nature of one's research. A second, more plausible solution is to compute impact factors adjusted for self-citations, as we did in this study. This procedure can be defended on the basis of the observation that self-citations do not reveal much about the impact of an article in the wider scientific community (Aksnes, 2003). A third possibility is to use longer citation windows (e.g., five years) when calculating impact factors. This is known to reduce the proportion of self-citations (Moed, van Leeuwen, & Reedijk, 1999).

REFERENCES

Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist, 58,* 15–23.

Aksnes, D. W. (2003). A macro study of self-citation. *Scientometrics, 56,* 235–246.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Boor, M. (1982). The citation impact factor: Another dubious index of journal quality. *American Psychologist, 37,* 975–977.

Errors in citation statistics: Opinion of the *Nature* editors. (2002, January 10). *Nature, 415,* 101.

Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist, 33,* 920–934.

Lawrence, P. A. (2003, March 20). The politics of publication. *Nature, 422,* 259–261.

McGarty, C. (2000). The citation impact factor in social psychology: A bad statistic that encourages bad science? *Current Research in Social Psychology, 5*(1), 1–16. Retrieved October 16, 2003, from http://www.uiowa.edu/~grpproc/crisp/crisp.5.1.htm

Moed, H. F., van Leeuwen, T. N., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics, 46,* 575–589.

Correspondence concerning this comment should be addressed to Frederik Anseel, Department of Personnel Management and Work and Organizational Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium. E-mail: frederik.anseel@UGent.be

DOI: 10.1037/0003.066X.59.1.51

On (Not) Trimming One's Toenails With a Bazooka

Christopher D. Green
York University

Adair and Vohra (January 2003) showed that the number of references used in psychological articles has increased over the past few decades. This increase constituted such a crisis in their view that they advised psychologists to cut back on the number of references they use as a way of responding to the information explosion.

From my perspective, however, there is no crisis. First, a simple plot of the information pertaining to the four psychology journals given in their Table 1 (Adair & Vohra, 2003, p. 17) shows the increase to have been linear over the past three decades, not explosive, as they claimed.

Second, for many of us who have grown up in the information age, the amount of material to which we have ready access does not constitute a crisis but a bounty. To be sure, navigation can be difficult at times, but electronic technologies are making it easier by the day.

Third, what exactly is the cost of additional references? Surely it does not constitute an academic, intellectual, or scholarly crisis. Adair and Vohra (2003) attempted to buttress their argument with a comparison of the increase in references used in psychology articles to that of those used in physics, but surely we psychologists are now beyond the stage where we measure our progress by how faithfully we ape physicists. Every discipline develops its own scholarly culture over time. In psychology, we use more references than in physics. Why? Perhaps because we have more respect for our history. Perhaps, by contrast, because we cannot assume that our colleagues know their disciplinary history as well as physicists do. Perhaps because we have more competing theoretical strands running side by side. Perhaps because we just like it that way.

In point of fact, given that additional references can only be advantageous to the reader, the only party that could legitimately object to them would be the publisher, who must set and print extra pages to accommodate them. I am tempted to dismiss this as irrelevant. Publishers should not trump matters of intellectual quality with economic concerns—if they try to, we should get different publishers (or abandon them in favor of the Internet, but that is another argument). Even if we accept this problem as legitimate, it can be dealt with in ways much less drastic than demanding that reference lists be cut: Smaller fonts, standard abbreviations for commonly used journal titles, truncated article titles, and the like would all shrink the physical size of reference lists without actually deleting items.

Even more important, however, is that we should abandon the lingering Gutenberg-era assumptions that permeate Adair and Vohra's (2003) article. None of their worries will matter one whit once psychologists move into the era of electronic scholarly publication. References will no longer take up literal page space but occupy only a few hundred extra bytes in a digital file. If the user chooses not to print the references to save paper, so be it. Stevan Harnad has shown psychologists the way of the probable future (indeed, the way of the present in physics) with his electronic journal *Psycoloquy*, with his e-print archive *CogPrints*, and with his many articles and addresses on the topic of electronic scholarly publication.

Whatever else, let us not take drastic measures to resolve a matter that is problematic only for an antiquated technology that we are now (gradually) in the process of abandoning.

REFERENCES

Adair, J. G., & Vohra, N. (2003). The explosion of knowledge, references, and citations: Psychology's unique response to a crisis. *American Psychologist, 58,* 15–23.

Correspondence concerning this comment should be addressed to Christopher D. Green, Department of Psychology, York University, Toronto, Ontario, Canada M3J 1P3. E-mail: christo@yorku.ca