

DEMONSTRATING THE CUMULATIVE EFFECTS OF UNCONSCIOUS BIAS WITH WAGES-ACADEMIC (WORKSHOP ACTIVITY FOR GENDER EQUITY SIMULATION): SHORT- AND LONG-TERM IMPACT ON FACULTY AND ADMINISTRATORS

Stephanie A. Shields,^{1,*} Kaitlin T. McCormick,¹
Elaine C. Diccico,¹ & Matthew J. Zawadzki²

¹Pennsylvania State University, University Park, Pennsylvania 16801, USA

²University of California, Merced, Merced, California 95340, USA

*Address all correspondence to: Stephanie A. Shields, Pennsylvania State University, 140 Moore Bldg., University Park, PA, United States, 16801; Tel.: 814-863-1729, E-mail: sashields@psu.edu.

The long-term effectiveness of WAGES-Academic, a brief intervention that illustrates the cumulative negative effect of minor disadvantages, is reported. University faculty and academic administrators (n = 69) in six sessions at four different universities completed assessments at two time points: a pre/post questionnaire at intervention and open-ended questions in response to email between two and four years after the WAGES session. Pre/post evaluations replicate and extend results obtained in randomized trials. Specifically, after playing WAGES compared to before, participants were more likely to endorse statements that the effect of many small incidents of gender inequity are cumulatively harmful, that case-by-case comparisons of individual applicants are difficult to do objectively, and that masked evaluations are effective in making unbiased hiring decisions. No change occurred in participants' agreement with standardized evaluation forms or accountability of decision-makers as effective. Open-ended questions indicated that WAGES validated many participants' experiences and observations about subtle bias. Long-term follow-up responses were obtained for 23 of 60 individuals with working email addresses. All except two indicated that they remembered participating and recalled inequity as WAGES' focus. Fifty-seven percent indicated that WAGES had led to changes in their behavior, insights into gender biases in their institutions' policies and practices, or policy change at their institutions. We discuss the implications of using WAGES-Academic as a primary or supplemental intervention to educate regarding unconscious bias, (i.e., systematic errors in judgment due to ordinary cognitive processes rather than conscious decision).

KEY WORDS: experiential learning, academic culture, salary equity, university administration, STEM faculty

1. INTRODUCTION

The proportion of women earning PhDs in science, technology, engineering, and mathematics

(STEM) fields in the U.S. is steadily increasing, with few exceptions. According to the National Science Foundation (2014), women earned 20% of all PhDs in computer science, 31% in the physical sciences, and 22% across engineering fields in 2011. That said, the proportion of women tenured faculty remains well below the proportion of women PhDs available for most science and engineering fields; furthermore, the proportion of women in academic positions decreases as a function of professorial rank and institutional prestige (West and Curtis, 2006).

The lag between women's success in obtaining PhDs in STEM fields and successful advancement to tenure and senior leadership positions can be attributed to diverse factors (Duch et al., 2012; Settles et al., 2006; Banerjee and Pawley, 2013), including the persistence of deliberate discrimination. One factor that stands out, however, is the cumulative effect of *unconscious bias* (Wennerås and Wold, 1997; Budden et al., 2008), that is, systematic errors in judgment (e.g., activation of stereotypes) that are due to ordinary, relatively automatic cognitive processes rather than conscious decision. Unconscious bias operates in a way that is hard to notice, identify, and change unless one is actively seeking out evidence of its influence (Valian, 1998). To reduce the impact of unconscious bias, it must be made visible.

Various types of brief interventions and educational modules have been developed to educate faculty and administrators about the operation of unconscious bias (see <http://www.portal.advance.vt.edu/> for examples), but there is sparse information on the long-term effects of these efforts at either the individual or institutional levels. In this paper, we report on the short- and long-term effectiveness of the Workshop Activity for Gender Equity Simulation academic version (WAGES-Academic; <http://wages.la.psu.edu/>) which is designed for use with college and university faculty and academic administrators. WAGES-Academic is a brief intervention that illustrates the cumulative negative effect of minor disadvantages that accrue through unconscious bias in an academic context. It is designed to be used alone or in conjunction with existing training programs. Previous randomized trials with undergraduate research participants have shown that WAGES increases knowledge of gender issues (Shields et al., 2011) without evoking reactance but while instilling self-efficacy (Zawadzki et al., 2012). WAGES has been shown to reduce the endorsement of sexist beliefs (Zawadzki et al., 2013), increase the recognition of everyday sexism as harmful and promote behavioral intentions to reduce everyday sexism (Cundiff et al., 2014), and increase the ability to detect subtle sexism (Danube et al., in preparation).

In the present study, we measured the effectiveness of WAGES-Academic at two time points in four universities with WAGES' target audience, namely, university faculty and academic administrators, in the academic settings for which WAGES was designed to be used. A brief pre/post questionnaire was administered at the time of the intervention, and then responses to a follow-up email questionnaire were collected between two to four years after the WAGES session. Questions addressed whether playing WAGES increased understanding of biased hiring systems and knowledge about gender bias. Thus, the study evaluated both the immediate impact of participating in WAGES and the long-term effects of participation (e.g., does WAGES promote support for and adoption of gender-fair procedures?).

2. WHY WAGES?

Instances of sexist prejudice and blatant discrimination continue to occur in the academic environment (Heilman and Eagly, 2008; London et al., 2012). More insidious and broad-reaching in its effects, however, is the interlocking set of structural, cultural, and environmental factors, such as inadequate family leave policies and the tenure clock itself, that produce disproportionately

detrimental outcomes for women in terms of attracting them to academic positions, retaining them as faculty, and promoting their advancement to leadership positions (Bagillhole and Goode, 2001; Rosser, 2004; Blackwell et al., 2009; Yost et al., 2013). Unconscious bias is the “glue” that binds these factors together and makes them difficult to displace. People want to be fair and believe they are being fair, even though research has shown again and again (as detailed in the following section) that good intentions are not sufficient to bring about change. Thus, true change requires addressing structural factors, but structural factors are at least partly held in place by unconscious bias, making interventions at both the individual and institutional level necessary to work toward equity (see, for example, Bird, 2011).

2.1 The Harms of Unconscious Bias

A growing body of research concerned with forms of unconscious gender bias (Fletcher, 1999; Glick and Fiske, 1996; Sayers, 2012) such as gender stereotypes (Burgess and Borgida, 1999) reveals their cumulative harms (Lincoln et al., 2012). For example, Spalter-Roth and Erskine’s (2005) and Williams’ (2005) research underscores that women often hit a “maternal wall” that has little to do with actual performance and more to do with stereotypes about mothers’ suitability for serious science careers (Heilman and Okimoto, 2008; King, 2008; Lips and Lawson, 2009). In another vein, Eagly and colleagues’ research has shown that women in leadership positions are evaluated in a negatively biased way compared to men in those positions (e.g., Eagly, 2005; Eagly et al., 1995; Scott and Brown, 2006).

Information about patterns of bias is important because of the dramatic cumulative effects of even a small rate of gender bias (Martell et al., 1996). First, bias materially disadvantages women in terms of time sinks, lower pay, and lost opportunities (Valian, 1998). Second, there is a measurable psychological and emotional toll of dealing with everyday stress associated with outgroup status (e.g., Biernat and Fuegen, 2001; Lincoln et al., 2012; Swim et al., 2001; Trix and Psenka, 2003). Third, this pattern of exclusion and bias feeds back to institutional devaluation of women and women’s performance with the result that women are less likely to be chosen for leadership positions (Hollenshead, 2003; Ferber, 2003) and/or less likely to be able to take them (Long, 2003).

These incidents take a considerable cumulative toll on individuals, particularly in conjunction with other stressors of daily life (e.g., Klonoff and Landrine, 1995; Lincoln et al., 2012; Swim et al., 2001) and when the individual is multiply marginalized, as is the case for women of color (De Welde and Stepnick, 2014; Gutiérrez y Muhs et al., 2012) and/or other marginalized statuses (e.g., disability or sexual orientation). In particular, the pervasiveness and persistence of the effects of unconscious bias negatively impacts morale, performance, and, ultimately, willingness to continue working within the academic setting (Kogan and Laursen, 2011; Xu, 2008).

2.1.1 *Difficulty in Correcting Unconscious Bias. Reducing the Influence of Unconscious Bias*

It takes more than good intentions and resolve on the part of administrators and faculty decision-makers.

Unconscious biases, which can be thought of as one form of *cognitive shortcut* (e.g., Bargh, 1997; Kahneman, 2003; Keren and Teigen, 2004), persist because even well-intentioned people are susceptible to them. Susceptibility is exacerbated when people lack the time to carefully

and objectively process the information available to them. Cognitive shortcuts are judgments deployed automatically and without deliberation that make it possible to process large quantities of complex information without becoming caught up in detail. The very same advantage of cognitive shortcuts, however, can lead to errors in judgment, such as failing to consciously consider available information. Processing of social information, especially information about other people, is influenced by stereotypes even when a person does not consciously endorse the stereotype (e.g., Bonnot and Croizet, 2007; Burkley and Blanton, 2008; Delisle et al., 2009; Logel et al., 2009; Sinclair et al., 2009; Wessel and Ryan, 2008). For example, if job candidates are compared on a case-by-case basis, it is easy to overlook or explain away qualifications that do not match our expectations or, conversely, give undue weight to stereotype-confirming information. These cognitive shortcuts (which reflect the operation of unconscious bias) interfere with objective decision-making. Use of masked review or standardized evaluation forms are two strategies that can reduce the impact of unconscious bias in these instances (Bornmann et al., 2007; Knobloch-Westerwick et al., 2013). In sum, people are susceptible to unintentional bias even when they try not to be or when they are themselves disadvantaged by the bias; interventions are needed to reveal the power and impact of cognitive shortcuts.

3. WAGES AS AN INTERVENTION

WAGES (for additional information see Shields et al., 2011) illustrates the varied sources and cumulative effect of apparently minor disadvantages that negatively impact women faculty in STEM disciplines. WAGES draws on demographic data and a broad body of social/behavioral science research relevant to understanding the nature and operation of factors that contribute to the cumulative effect of unconscious bias in the academic STEM workplace. To name but a few, these include sexism (Becker et al., 2014), effects of “token” or solo status (King et al., 2010; Turner et al., 2011), and evaluation of women’s versus men’s emotion (Shields, 2013).

3.1 The WAGES Demonstration

WAGES is a 75–90 minute interactive demonstration that consists of a game portion followed by a discussion (Shields et al., 2011). For the game, four to eight individuals are randomly divided into two teams (Green and White). The goal of the game is to earn credit chips that allow members of each team to advance up the academic career ladder. The same standards for advancement are required for all players, but the accumulation of credit chips depends on the game cards that give a small overall advantage to the White Team, with the gendered nature of the team experience emerging over time. Following game play, a facilitator leads guided discussion that makes WAGES’ purpose and learning objectives clear. Discussion also explicitly addresses intersections of social identity and how women with multiple marginalized statuses may experience different or compounded negative outcomes in academic STEM (e.g., De Welde and Stepnick, 2014; Gutiérrez y Muhs et al., 2012). Discussion concludes with ways that unconscious bias can be counteracted by the institution (e.g., transparency in promotion criteria) and by individuals (e.g., evaluation tools that promote fair evaluation in the promotion and tenure process).

3.2 Theory Underlying the Intervention

The framework for WAGES is derived from educational research, which shows that to develop

competence in an area of inquiry, the learner must have a foundation of factual knowledge and must understand facts and ideas within a conceptual framework so as to organize knowledge in ways that facilitate retrieval and application (National Research Council, 2000). Experiential learning (Kolb, 1984; Kolb and Kolb, 2005; Garvin and Ramsier, 2003) has demonstrated effectiveness for enhanced learning (Cantor, 1997; Springer et al., 1997), particularly in contexts in which complex information must be processed (Burke et al., 2011) and contexts in which deeply ingrained behavioral attitudes are challenged (Eubank et al., 2011).

4. THE PRESENT STUDY

Short- and long-term effectiveness of WAGES-Academic, with its target audience of university faculty and academic administrators, was tested at four different universities and at two time points: a brief pre/post questionnaire at intervention (immediate follow-up) and open-ended questions in response to email between two and four years after the WAGES session (long-term follow-up).

At immediate follow-up, we hypothesized that if WAGES is effective, we would find a significant postintervention increase on closed-ended items assessing knowledge of gender bias and equitable recruitment methods. On open-ended questions, we expected that participants would report more positive than negative features of WAGES at immediate follow-up and that they would remember WAGES and the goals of WAGES at long-term follow-up.

5. METHOD

5.1 Overview of the Study

Responses were collected from university faculty, staff, and administrators at four different university locations who participated in WAGES either as a stand-alone intervention or in conjunction with other leadership or diversity-training workshops. The study was conducted in two phases. *Immediate follow-up* consisted of a brief pre/post questionnaire that was administered at the time of participation in a WAGES session. *Long-term follow-up* was a five-item questionnaire emailed to participants who had indicated willingness to be contacted again.

5.2 Participants

All individuals who participated in the WAGES sessions ($n = 94$) completed the immediate follow-up. Of these, 69 completed both quantitative and qualitative portions of the pre/post questionnaire, and 25 participants completed only the qualitative portion.

Demographic data were not collected from participants, but based on facilitators' written reports, the large majority of participants were perceived to be White by facilitators, and most participants played WAGES as part of diversity workshops at their institutions as indicated on their response sheets. University positions of participants are listed in Table 1.

A total of eight WAGES sessions were conducted between November 2010 and July 2012 and ranged in size from 4 to 25 individuals (which entailed multiple games simultaneously being played with a single group discussion); all except one were mixed-gender groups. Sessions were conducted at Penn State (by MJZ), two Midwestern universities, and a Mid-Atlantic liberal arts college. All sessions except those at Penn State were conducted by individuals unaffiliated with the WAGES project who used the WAGES-Academic Facilitator's Manual as a guide to

TABLE 1: University positions of participants

University Position	n = 96	Percent of Sample
Professor	52	55.3
Academic Dean	19	20.2
Staff	8	8.5
Department Chair	6	6.4
Lecturer	4	4.3
Diversity Planner	2	2.1
Associate Vice Chancellor	1	1.1
Director of Graduate Program	1	1.1
Emeritus Professor	1	1.1

conducting the session (game play and discussion) and who mailed completed questionnaires to the first author.

The long-term follow-up was comprised of participants who indicated that they were willing to be contacted at a later date ($n = 64$ provided an email address). Of the 60 who met employment criteria as present/former faculty members or academic administrators, and for whom we had a functioning email address, 23 responded (38%) and were included in analyses. This is within the average response rate to email surveys (Shih and Fan, 2009).

6. MATERIALS AND PROCEDURE

6.1 Immediate Follow-up

Before playing WAGES and immediately following postplay discussion, participants completed brief questionnaires. Prior to play, participants indicated the group with which they were participating in the intervention, their role in the group, their academic/administrative title, how much responsibility they had for equitable treatment of job candidates, and previous diversity training and its usefulness to them. They then rated their agreement with four statements related to sources of bias that contribute to gender inequity and four statements related to the effectiveness of specific evaluation procedures to reduce bias on a 1 (*do not agree at all* or *not at all effective*) to 7 (*agree strongly* or *very much effective*) scale (see Appendix). They were also asked if they currently used any of these evaluation procedures and, if so, to describe. Immediately following postplay discussion, participants responded to two open-ended questions: (1) whether they would recommend the activity to others, and (2) what they found to be effective/ineffective about the activity. They then rated their agreement with the same four statements related to sources of bias and four statements related to the effectiveness of specific evaluation procedures to reduce bias but listed in a different order. Last, they were asked if they would use any of the evaluation procedures if available, if they had any additional comments, and whether they would be willing to be contacted in the future.

6.2 Long-term Follow-up

Participants who indicated they were willing to be contacted were sent a brief follow-up ques-

tionnaire via email between 24 and 45 months after they had participated in a WAGES-Academic session. Up to three reminders were sent to nonresponders. The email included the following questions: (1) Do you remember participating in this activity? (2) If you do recall WAGES, do you recall one or more of the main message(s) of WAGES? If so, please describe. (3) Since participating, have you noticed issues raised by the activity occurring at work or in everyday life? If yes, please describe. (4) Has your participation in the activity influenced your own decision-making in hiring or advancement decisions? If yes, please briefly describe in what way. And, last, (5) Is there anything else you would like to share with us about WAGES?

7. RESULTS

7.1 Immediate Follow-up

7.1.1 Quantitative Responses

We conducted individual repeated-measures ANOVAs for each of the pre/post quantitative items ($n = 69$) (Table 2).

Regarding items related to sources of bias that contribute to gender inequity, participants' agreement that the discrepancy between women and men in career success can be explained by the cumulative effect of small incidents significantly increased postactivity, $F(1, 63) = 5.20, p = 0.03, \eta_p^2 = .08$, as did agreement that case-by-case comparisons of individual applicants are difficult to do objectively, $F(1, 62) = 14.99, p < .001, \eta_p^2 = .20$. In contrast, agreement that masked evaluations are biased (reverse-scored before analysis) increased opposite to the predicted direction, $F(1, 63) = 11.01, p < .001, \eta_p^2 = 0.15$. Participant agreement with stereotypes influencing decision-making even when trying to be fair did not increase, $F(1, 63) = 0.00, p = 1.00, \eta_p^2 = 0.00$.

For items asking about the objectivity of specific evaluation techniques, there was significant increase in agreement that masked evaluations are an effective tool to make unbiased hiring decisions, $F(1, 59) = 26.88, p < .001, \eta_p^2 = .31$, whereas endorsement of standardized evaluations as reducing bias only approached significance, $F(1, 59) = 2.74, p = .10$. Counter to hypotheses, agreement with the effectiveness of comparing CVs directly (reverse-scored before analysis) did not significantly increase, $F(1, 60) = 0.60, p = 0.44, \eta_p^2 = 0.10$, nor did agreement with the effectiveness of accountability increase, $F(1, 58) = 1.66, p = 0.20, \eta_p^2 = 0.03$.

On the pretest we also asked "Do you currently use any of these evaluation procedures?" and 49 participants responded. Four said no and nine said yes without specifying which procedures. Of 36 participants who responded yes and specified procedures, 16 mentioned using standardized evaluations; 10 mentioned using direct CV comparison; six mentioned using both standardized evaluations and direct comparison of CVs; two mentioned using standardized evaluations, direct CV comparison, and systems of accountability; one mentioned using direct CV comparison and systems of accountability; and one mentioned using all procedures, standardized evaluation forms, direct CV comparison, masked evaluations, and systems of accountability. On the post-test we asked, "If available, would you want to use any of these evaluation procedures?" and 45 responded. One said no, one said maybe, and four expressed feasibility concerns. Twelve responded yes without specifying, and 27 responded yes and listed procedures: eight mentioned masked evaluations, six mentioned systems of accountability, four mentioned standardized evaluations, one mentioned direct CV comparison, and eight mentioned a combination of two or more procedures.

TABLE 2: Means (standard deviations) of agreement with gender inequity items

	Pre-WAGES	Post-WAGES	Effect Size (d)
Sources of Gender Inequity			
Cumulative bias	5.73 (1.16)	6.05 (0.09)	.39*
Case-by-case evaluations	4.79 (1.66)	5.54 (1.66)	.45**
Masked evaluations	5.95 (1.05)	5.20 (1.84)	.50**
Stereotypes	6.34 (0.88)	6.34 (0.90)	0
Effectiveness of Evaluation Procedures			
Standardized evaluation forms	4.45 (1.45)	4.65 (1.70)	.13 ⁺
Masked evaluations	5.00 (1.14)	5.65 (1.16)	.57**
Direct CV comparison	3.77 (1.51)	3.66 (1.69)	.07
Accountability for decisions	5.58 (1.33)	5.80 (1.19)	.17

Note: Items rated on a 1 to 7 scale, with higher numbers indicating greater agreement with the sources of gender inequity and/or effectiveness of evaluation procedures.

⁺ $p = .01$, * $p < .05$, ** $p < .001$.

7.1.2 Qualitative Responses

In response to whether they would recommend WAGES to others, 95% of participants indicated that they would. In addition to indicating “yes,” many added comments similar to the participant who wrote, “Good activity [active learning] to show biases in academia. Clear method for illustrating the impact of subtle differences over time.” Another participant on a diversity committee wrote, “Bias is often portrayed as being obvious, [WAGES] demonstrates that it is usually not.” Other participants in similar positions described WAGES as a “useful conversation starter,” as a way to “stay current and get remotivated to combat sexism, racism, etc.,” and as a “good way to learn to relate” and “facilitate empathy.”

The question “Was there anything particularly effective or ineffective about the activity?” was answered by 73 participants, 13 of whom mentioned both effective and ineffective aspects. For this question and other open-ended questions, consensus coding was used. Responses were examined independently by two authors (KTM and ECD) for themes that were mentioned by at least three participants. The two coders then met to compare results and disagreements were discussed until consensus was reached. The authors’ independent coding was overall strongly in agreement with one another, with the exception of one or two items which were discussed until consensus was reached.

Of 53 participants who indicated effective aspects of WAGES, the most common theme ($n = 20$) was WAGES’ powerful illustration of the cumulative effect of subtle bias over time. Eight additional individuals mentioned that the game was engaging in a nonthreatening way given the content. For example, one participant wrote, “Having a game feeling helped keep the discussion going without personal threat.” Of the 31 participants who identified ineffective aspects of WAGES, the most common theme was that some aspect of the game was unrealistic ($n = 11$). One of several respondents, for example, wrote, “We wondered why everyone got tenure.” The second most named limitation was that game instructions were not sufficiently clear ($n = 5$).

7.2 Long-term Follow-up

All questions were open-ended. Of those who responded ($n = 23$), all except two indicated that they remembered participating and that gender inequity in academia was the theme of the intervention, and seven specifically mentioned that the activity was concerned with the cumulative effect of subtle bias.

In response to the question of whether they had noticed issues raised by WAGES occurring at work or in everyday life, over half (12 of 21) indicated that they had and that this had led to changes in their behavior, policy change at their institutions, or new insights into gender biases in their institutions' policies and practices. A female participant, for example, described recognizing gender bias after WAGES, "especially at department meetings seeing the reactions to what we say, or who talks more, or who dominates the meeting. . . with senior male faculty advocating more for my male colleague." Of those who recalled WAGES, but indicated no change, nine said that this was due to policies already being in place and six did not elaborate.

In response to the question on whether WAGES had influenced their decision-making, seven noted WAGES' influence on their professional behavior and their examination of their own and others' biases. Some broadly described becoming more generally "aware of bias in myself and others." Others mentioned specific actions taken after the intervention, either at the institutional level or individual level. For instance, one participant wrote: "I am more sensitive to climate issues, and I did get bold enough to take action...I wrote to the chief to point out that it came across as favoritism." Another woman faculty member, apparently influenced by WAGES items that reveal the undue service load that many women faculty experience, described "decid[ing] to step down after four years as one of the coadvisors of the student organization and be[coming] more strategic about the forms of service work I engage in." Of the 14 who said WAGES was not an influence, seven indicated that they are not in a position to make hiring decisions, and seven said no/did not answer, with two of these stating it was because they already had policies in place for equity in hiring.

In response to whether they had anything else to add about the intervention, 11 provided an additional comment. Five suggested other possible versions of WAGES to develop, including a graduate student/postdoc version, race/ethnicity version, online version, a modified version for faculty hiring, and a version specifically for chairs and faculty when evaluating candidates for promotion and advancement. Three commented that they would like to recommend the intervention be used in their own departments or to other leaders and students early in their careers. The remaining three commented that the intervention was enjoyable, effective, and not representative of faculty's experiences in the humanities, respectively.

8. DISCUSSION

The short- and long-term effectiveness of WAGES-Academic was demonstrated for the intervention's target audience of university faculty and academic administrators. We had hypothesized that on immediate pre/post intervention measures, we would find a significant postintervention increase on closed-ended items assessing knowledge of gender bias and equitable recruitment methods. We also expected that participants would report more positive than negative features of WAGES in the immediate follow-up, and that they would remember WAGES and the goals of WAGES in the long-term follow-up.

The pre/post questionnaire at intervention extended results we have obtained in published

randomized trials with undergraduate research participants. Specifically, after playing WAGES, faculty and administrators were more likely to endorse that the cumulative effect of many small incidents of gender inequity are harmful, that case-by-case comparisons of individual applicants are difficult to do objectively, and that masked evaluations help in making unbiased hiring decisions. No change was seen, however, in participants' agreement that we are influenced by stereotypes even when we try to be fair, or that fair evaluation procedures are enhanced by using standardized evaluation forms or holding decision-makers accountable for the fairness of their decisions.

Importantly, one of the two items most directly related to the information in the game-play portion of WAGES (the cumulative effects of minor biases) showed a significant increase in endorsement from pre- to post-play assessment. Many participants were already engaged in diversity activities, such as serving on committees or as trainers, or had previous diversity training. The fact that there was significant pre/post change in seeing minor bias as a source of gender inequity suggests that training about unconscious bias and its cumulative negative effects is neither a feature of, nor is sufficiently emphasized in, conventional diversity-training methods. A second item directly addressed in the game-play portion of WAGES (the power of stereotypes and cognitive shortcuts) did not change, but lack of change is likely due to ceiling effects. On both pre- and postmeasures, the mean endorsement was 6.34 on a 7-point scale. This likely indicates that participants were already aware of the harmful effects of stereotypes preintervention, although WAGES may have increased their attentiveness to the under-the-radar way in which stereotypes and biases exert an increasing effect over time.

One particularly puzzling finding relates to two items concerned with the effectiveness of masked evaluations. One was an item included in the four concerned with sources of bias. It, counter to predictions, showed a significant decrease in endorsement pre/post. The other item was included in the four items related to the effectiveness of specific evaluation procedures. It showed the predicted significant increase in endorsement pre/post. Review of the wording of each suggests that the former question ("Rate agreement with statement: 'Masked evaluations—evaluations where we don't know a person's gender, race, etc.—tend to be more biased than those where we know the individual's identity'") may have been confusing, given that it was stated in the negative (i.e., reverse coded). In contrast, the item that yielded the predicted pre/post increase in endorsement was stated in a straightforward manner ["Rate effectiveness of evaluation procedure: 'Masked evaluation (i.e., where the person's identity is not known to the evaluator)']". In addition, a related item concerning bias in case-by-case comparisons also yielded a pre/post change in the predicted direction.

Overall, open-ended questions in the immediate follow-up indicated that WAGES validated many participants' experiences and observations about subtle bias without making them feel uncomfortable about engaging in discussion. WAGES-Academic's structure as an interactive board game relies on active participation for its success. Participant comments underscore the importance of face-to-face discussion in WAGES' effectiveness. First, it encourages engagement in discussion, which is essential to the experiential learning process. Second, the activity helps to validate some faculty members' experiences, while also helping others grasp the validity of the game (e.g., if my colleague says "This happened to me," I should take it seriously). Discussion during the game is typically lively and often includes players sharing their own observations and experiences. These give-and-take conversations about individual items during the course of play then set the stage for a more fruitful postgame discussion. Additionally, discussion often highlights the intersectionality of social identities, most often when members of multiply mar-

ginalized groups describe how their experience of a game-card scenario differs from other participants' experience because of their intersectional position. In future research, it will be useful to gather information about these dynamics of game-play and postplay discussion as it relates to WAGES' downstream effectiveness.

For long-term follow-up, we were able to contact 23 of 64 people who had provided email addresses. Although our long-term follow-up generated a reasonable response, we cannot estimate how many participants remembered playing WAGES because we do not know why people did not respond to the email (either they did not remember, and therefore did not respond, or they did not respond for other reasons). About two-thirds of participants had provided contact information and indicated willingness to be contacted, suggesting potential selection effects. Importantly, 95% of all 94 participants had said they would recommend WAGES to others, so not volunteering for follow-up should not be interpreted as dissatisfaction with the experience. That said, to determine whether our respondents were representative of the original sample, we compared responses of those who indicated willingness to be contacted to those who did not on the cumulative bias item on the pre/post questionnaire. This item most directly reflects the "message" of WAGES regarding the cumulative negative effect of small incidents of gender inequity on women's careers. The increase in agreement with the harmfulness of cumulative bias from pre to post remained significant regardless of whether or not participants agreed to be contacted in the future, $F(1, 63) = 4.13, p = .05, \eta_p^2 = .06$.

Responses to long-term follow-up were striking given that participants were contacted long after the WAGES session, at least 24 months and up to 45 months later. Participants described noticing bias in their everyday lives many months after playing WAGES and cited WAGES' influence on their hiring decisions. A participant remarked noticing "gendered expectations for teaching, service, and student organizations" after WAGES, and another commented on noticing men receiving "more attention/praise for filling the same role." Participants also credited WAGES with positively impacting their own hiring decisions. One participant commented that WAGES "helped inform a hiring decision" and another described the intervention as "influencing decisions since." A third participant attributed WAGES with both "influenc[ing a] department chair in hiring [and] influenc[ing his own] interactions with female graduate students."

8.1 Limitations

The only demographic information we collected on participants related to their academic position as related to participation in the WAGES session. Thus, we could not track factors that may be related to attrition nor to variables that may be related to the degree of WAGES' effectiveness. Specifically, we do not know the gender, the racial ethnic makeup of the sample as a whole or of individual groups, the age range or years in career of participants, or which participants may also identify with underrepresented groups (e.g., individuals with disabilities). This is relevant because it speaks to how well participants may be able to relate WAGES items to their own experience. WAGES is designed to speak to multiple marginalization and to be relevant for people in various stages of their careers.

Given the small sample size, we were unable to compare all-female and mixed-gender groups. That said, none of the sessions in the present study were comprised of only male participants, and it may be that the dynamics of all-male groups may affect aspects of WAGES' effectiveness. Future research should examine gender composition, because many STEM workplaces have a higher proportion of men than women, and these environments may foster sexism, in part

by fostering a view of women colleagues as an outgroup. Perceptions of others as belonging to an outgroup both exacerbates the activation of stereotypes by perceivers and undermines women's sense of belonging in that environment (Nosek et al., 2002; Shapiro and Williams, 2012; Stout et al., 2011).

With one exception (an item pertaining to masked evaluation), ratings of the effectiveness of four different candidate evaluation procedures did not yield change in the predicted direction in immediate follow-up. Evaluation procedures are not directly addressed in the game-play portion but are among points to be covered in postgame discussion. Postgame discussion varies somewhat in content depending on the facilitator and the course of group discussion. It may be that some groups emphasized these issues while others did not. As evaluation of WAGES' effectiveness continues in the future, additional information should be gathered concerning participants' identification of specific strategies for reducing the impact of unconscious bias on hiring and promotion decisions.

9. CONCLUSION

Although women have made strong advances in some STEM fields, most notably, the life sciences, in other fields, such as physics, the proportion of women remains low. Finding ways to counteract the occurrence of unconscious bias that cumulatively hinders women's advancement in academic STEM fields is a challenging and ongoing task. Research has demonstrated the benefits of interventions designed to reveal the processes of unconscious bias and identify practices that minimize bias (Engberg, 2004; Sheridan et al., 2010). Understanding the nature, power, and persistence of stereotypes and unconscious biases is an important foundation for fostering effective implementation of policies and procedures that can, in turn, transform climate and promote successful efforts to increase diversity of faculty and administrators in STEM disciplines (Agerstrom and Rooth, 2011; Greenwald et al., 2004; McConnell and Leibold, 2001). Training about unconscious biases by itself cannot correct organizational sexism (Kalev et al., 2006). Yet, without an understanding of how these biases operate and exert their influence, interventions will have difficulty in getting the buy-in needed to succeed in changing the work environment. Overall, WAGES appears to be effective in initiating discussion of unconscious bias which, in turn, can influence broader institutional practices over time.

ACKNOWLEDGMENTS

This research was supported in part through a grant awarded by the National Science Foundation (Award No. 0820212). We thank the many colleagues who generously agreed to complete questionnaires and follow-up contact in conjunction with their participation in a WAGES session.

REFERENCES

- Agerstrom, J., & Rooth, D. O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology*, 96(4), 790–805.
- Bagilhole, B., & Goode, J. (2001). The contradiction of the myth of individual merit, and the reality of a patriarchal support system in academic careers: A feminist investigation. *European Journal of Women Studies*, 8(2), 161–180.
- Banerjee, D., & Pawley, A. L. (2013). Gender and promotion: How do science, technology, engineering and mathematics (STEM) faculty members survive a foggy climate? *Journal of Women and Minorities in*

- Science and Engineering*, 19(4), 329–347.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: advances in social cognition* (pp. 1–61). Mahwah, NJ: Lawrence Erlbaum Associates.
- Becker, J. C., Zawadzki, M. J., & Shields, S. A. (2014). Confronting and reducing sexism: A call for research on intervention. *Journal of Social Issues*, 70(4), 603–614.
- Biernat, M., & Fuegen, K. (2001). Shifting standards and the evaluation of competence: Complexity in gender-based judgment and decision making. *Journal of Social Issues*, 57(4), 707–724.
- Blackwell, L. V., Snyder, L. A., & Mavriplis, C. (2009). Diverse faculty in STEM fields: Attitudes, performance, and fair treatment. *Journal of Diversity in Higher Education*, 2(4), 195–205.
- Bonnot, V., & Croizet, J. C. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology*, 43(6), 857–866.
- Bornmann, L., Mutz, R., & Daniel, H. D. (2007). Gender differences in grant peer review: A meta-analysis. *Journal of Informetrics*, 1(3), 226–238.
- Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., & Lortie, C. J. (2008). Double-blind review favours increased representation of female authors. *Trends in Ecology and Evolution*, 23(1), 4–6.
- Burgess, D., & Borgida, E. (1999). Who women are, who women should be: Descriptive and prescriptive gender stereotyping in sex discrimination. *Psychology, Public Policy, and Law*, 5(3), 665–692.
- Burke, M. J., Salvador, R. O., Smith-Crowe, K., Chan-Serafin, S., Smith, A., & Sonesh, S. (2011). The dread factor: How hazards and safety training influence learning and performance. *Journal of Applied Psychology*, 96(1), 46–70.
- Burkley, M., & Blanton, H. (2008). Endorsing a negative in-group stereotype as a self-protective strategy: Sacrificing the group to save the self. *Journal of Experimental Social Psychology*, 44(1), 37–49.
- Cantor, J. A. (1997). *Experiential learning in higher education: Linking classroom and community*. Hoboken, NJ: Wiley.
- Cundiff, J. L., Zawadzki, M. J., Danube, C. L., & Shields, S. A. (2014). Using experiential learning to increase the recognition of everyday sexism as harmful: The WAGES intervention. *Journal of Social Issues*, 70(4).
- Danube, C. L., Cundiff, J. C., Zawadzki, M. J., & Shields, S. A. (in press). *Using WAGES as an intervention to increase recognition and reporting of gender bias*.
- De Welde, K., & Stepnick, A. (Eds.). (2014). *Disrupting the culture of silence: confronting gender inequality and making change in higher education*. Sterling, VA: Stylus.
- Delisle, M., Guay, F., Senecal, C., & Larose, S. (2009). Predicting stereotype endorsement and academic motivation in women in science programs: A longitudinal model. *Learning and Individual Differences*, 19(4), 468–475.
- Duch J., Zeng X. H. T., Sales-Pardo, M., Radicchi F., Otis S., & Woodruff, T. K. (2012). The possible role of resource requirements and academic career-choice risk on gender differences in publication rate and impact. *PLoS One*, 8(5).
- Eagly, A. H. (2005). Achieving relational authenticity in leadership: Does gender matter? *Leadership Quarterly*, 16(3), 459–474.
- Eagly, A. H., Karau, S. J., & Makhijani, M. G. (1995). Gender and the effectiveness of leaders: A meta-analysis. *Psychology Bulletin*, 117(1), 125–145.
- Engberg, M. E. (2004). Improving intergroup relations in higher education: A critical examination of the influence of educational interventions on racial bias. *Review of Education Research*, 74(4), 473–524.
- Eubank, D., Orzano, J., Geffken, D., & Ricci, R. (2011). Teaching team membership to family medicine residents: What does it take? *Families, Systems, and Health*, 29(1), 29–43.
- Ferber, M. A., (2003). Women's uneven progress in academia: Problems and solutions. In L.S. Hornig (Ed.), *Equal rights, unequal outcomes: women in American research universities* (pp. 281–309). New York:

- Kluwer Academic/Plenum Publishers.
- Fletcher, J. K. (1999). *Disappearing acts: Gender, power, and relational practice at work*. Cambridge, MA: MIT Press.
- Garvin, M. R., & Ramsier, R. D. (2003). Experiential learning at the university level: A US case study. *Education Training*, 4(5) 280–285.
- Glick, P., & Fiske, S. T. (1996). The ambivalent sexism inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, 70(3), 491–512.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2004). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41.
- Gutiérrez y Muhs, G., Flores Niemann, Y., González, C. G., & Harris, A. P. (Eds.). (2012). *Presumed incompetent: The intersections of race and class for women in academia*. Logan, UT: Utah State University Press.
- Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology*, 1(4), 393–398.
- Heilman, M. E., & Okimoto, T. G. (2008). Motherhood: A potential source of bias in employment decisions. *Journal of Applied Psychology*, 93(1), 189–198.
- Hollenshead, C. (2003). Women in the academy: confronting barriers to equality. In L. S. Hornig (Ed.), *Equal rights, unequal outcomes: women in American research universities* (pp. 211–225). New York: Kluwer Academic/Plenum Publishers.
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best practices or best guesses? Assessing the efficacy of corporate affirmative action and diversity policies. *American Sociological Review*, 71(4), 589–617.
- Keren, G., & Teigen, K. H. (2004). Yet another look at the heuristics and biases approach. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (pp. 89–109). Malden, MA: Blackwell Publishing.
- King, E. B. (2008). The effect of bias on the advancement of working mothers: Disentangling legitimate concerns from inaccurate stereotypes as predictors of advancement in academe. *Human Relations*, 6(12), 1677–1711.
- King, E. B., Hebl, M. R., George, J. M., & Matusik, S. F. (2010). Understanding tokenism: Antecedents and consequences of a psychological climate of gender inequity. *Journal of Management*, 36(2), 482–510.
- Klonoff, E. A., & Landrine, H. (1995). The schedule of sexist events: A measure of lifetime and recent sexist discrimination in women's lives. *Psychology Women Quarterly*, 19(4), 439–472.
- Knobloch-Westerwick, S., Glynn, C. J., & Huge, M. (2013). The Matilda Effect in science communication: An experiment on gender bias in publication quality perceptions and collaboration interest. *Science Communication*, 35(5), 603–625.
- Kogan, M., & Laursen, S. L. (2011). Obstacles in advancement of young female geoscientists: Research results from the earth science women's network (ESWN). *American Geophysical Union Fall Meeting*, 2011.
- Kolb, A. Y., & Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning and Education*, 4(2), 193–212.
- Kolb, D.A. (1984). *Experiential learning: Experience as the source of learning and development*. Upper Saddle River, NJ: Prentice Hall.
- Lincoln, A. E., Pincus, S., Bandows Koster, J., & Leboy, P. S. (2012). The Matilda effect in science: Awards and prizes in the US, 1990s and 2000s. *Social Studies of Science*, 42(2), 307–320.
- Lips, H., & Lawson, K. (2009). Work values, gender, and expectations about work commitment and pay:

- Laying the groundwork for the “motherhood penalty”? *Sex Roles*, 61(9), 667–676.
- Logel, C., Walton, G. M., Spencer, S. J., Iserman, E. C., von Hippel, W., & Bell, A. E. (2009). Interacting with sexist men triggers social identity threat among female engineers. *Journal of Personality and Social Psychology*, 96(6), 1089–1103.
- London, B., Downey, G., Romero-Canyas, R., Rattan, A., & Tyson, D. (2012). Gender-based rejection sensitivity and academic self-silencing in women. *Journal of Personality and Social Psychology*, 102(5), 961–979.
- Long, J. S. (2003). The presence and participation of women in academic science and engineering: 1973–1995. In L. S. Hornig (Ed.), *Equal rites, unequal outcomes: Women in American research universities* (pp. 145–174). New York, NY: Kluwer Academic/Plenum Publishers.
- Martell, R. F., Lane, D. M., & Emrich, C. (1996). Male-female differences: A computer simulation. *American Psychologist*, 51(2), 157–158.
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37(5), 435–442.
- National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: The National Academy Press.
- National Science Foundation. (2014). *Women, minorities, and persons with disabilities in science and engineering: 2013* (Special Report NSF 13-304). Arlington, VA: National Science Foundation.
- Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology*, 83(1), 44–59.
- Rosser, S. V. (2004). *The science glass ceiling: Academic women scientist and the struggle to succeed*. New York: Routledge.
- Sayers, R. C. (2012). The cost of being female: Critical comment on block. *Journal of Business Ethics*, 106(4), 519–524.
- Scott, K. A., & Brown, D. J. (2006). Female first, leader second? Gender bias in the encoding of leadership behavior. *Organizational Behavior and Human Decision Process*, 101(2), 230–242.
- Settles, I. H., Cortina, L. M., Malley, J., & Stewart, A. J. (2006). The climate for women in academic science: The good, the bad, and the changeable. *Psychology Women Quarterly*, 30(1), 47–58.
- 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.
- Sheridan, J. T., Fine, E., Pribbenow, C. M., Handelsman, J., & Carnes, M. (2010). Searching for excellence and diversity: Increasing the hiring of women faculty at one academic medical center. *Academic Medicine*, 85(6), 999–1007.
- Shields, S.A. (2013). Gender and emotion: What we think we know, what we need to know, and why it matters. *Psychology Women Quarterly*, 37(4), 423–435.
- Shields, S. A., Zawadzki, M. J., & Johnson, R. N. (2011). The impact of a workshop activity for gender equity simulation in the academy (WAGES-Academic) in demonstrating cumulative effects of gender bias. *Journal of Diversity in Higher Education*, 4(2), 120–129.
- Shih, T. H., & Fan, X. (2009). Comparing response rates in e-mail and paper surveys: A meta-analysis. *Educational Research Review*, 4(1), 26–40.
- Sinclair, S., Pappas, J., & Lun, J. (2009). The interpersonal basis of stereotype-relevant self-views. *Journal of Personality*, 77(5), 1343–1364.
- Spalter-Roth, R., & Erskine, W. (2005). Beyond the fear factor: Work/family policies in academia-resources or rewards? *Change*, 37(6), 19–25.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1997). *Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis*. Madison, WI: University of

Wisconsin-Madison.

- Stout, J. G., Dasgupta, N., Hunsinger, M., & McManus, M. A. (2011). STEMing the tide: Using in-group experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). *Journal of Personality and Social Psychology*, 100(2), 255–270.
- Swim, J. K., Hyers, L. L., Cohen, L. L., & Ferguson, M. J. (2001). Everyday sexism: Evidence for its incidence, nature, and psychological impact from three daily diary studies. *Journal of Social Issues*, 57(1), 31–53.
- Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and Society*, 14(2), 191–220.
- Turner, C. S. V., Gonzalez, J. C., & Wong, K. (2011). Faculty women of color: The critical nexus of race and gender. *Journal of Diversity in Higher Education*, 4(4), 199–211.
- Valian, V. (1998). *Why so slow? The advancement of women*. Cambridge, MA: The MIT Press.
- Wennerås, C., & Wold, A. (1997). Nepotism and sexism in peer review. *Nature*, 387, 341–343.
- Wessel, J. L., & Ryan, A. M. (2008). Past the first encounter: The role of stereotypes. *Industrial and Organizational Psychology*, 1(4), 409–411.
- West, M. S., & Curtis, J. W. (2006). *AAUP faculty gender equity indicators 2006*. Washington, DC: American Association of University Professors.
- Williams, J. (2005). *Unbending gender: Why family and work conflict and what to do about it*. New York, NY: Oxford University Press.
- Xu, Y. J. (2008). Gender disparity in STEM disciplines: A study of faculty attrition and turnover intentions. *Research in Higher Education*, 49(7), 607–624.
- Yost, E., Winstead, V., Cotten, S. R., & Handley, D. M. (2013). The recruitment and retention of emerging women scholars in STEM: Results from a national web-based survey of graduate students, postdoctoral fellows, and junior faculty. *Journal of Women and Minorities in Science and Education*, 19(2), 143–163.
- Zawadzki, M. J., Danube, C. L., & Shields, S. A. (2012). How to talk about gender inequity in the workplace: Using WAGES as an experiential learning tool to reduce reactance and promote self-efficacy. *Sex Roles*, 67(11), 605–616.
- Zawadzki, M. J., Shields, S. A., Danube, C. L., & Swim, J. K. (2013). Reducing the endorsement of sexism using experiential learning: The workshop activity for gender equity simulation (WAGES). *Psychology Women Quarterly*, 38(1), 75–92.

APPENDIX

INSTRUCTIONS: Using the scale below, rate the extent to which you agree with the following statements.

1	2	3	4	5	6	7
Do Not Agree at All			Somewhat Agree			Agree Strongly

_____ Discrepancy between women and men in career success can be explained by the cumulative effect of many small incidents of gender inequity.

_____ Case-by-case comparisons of individual applicants are difficult to do objectively.

_____ Masked evaluations – evaluations where we don’t know a person’s gender, race, etc. – tend to be more biased than those where we know the individual’s identity. (R)*

_____ Even when we try to be fair, we can be influenced by stereotypes and cognitive shortcuts.

INSTRUCTIONS: Please rate how effective you believe the following evaluation procedures are.

1	2	3	4	5	6	7
Not at All Effective			Somewhat Effective			Very Much Effective

_____ Standardized evaluation forms.

_____ Masked evaluation (i.e., where the person’s identity is not known to the evaluator).

_____ Directly comparing the CVs or merits of one candidate to another. (R)

_____ Systems where people are held accountable for the fairness of their decisions.

*(R) indicates reverse-coded item

