2016

The Influence of Implicit and Explicit Gender Bias on Grading, and the Effectiveness of Rubrics for Reducing Bias

Sarah Marie Jackson
Wright State University

THE INFLUENCE OF IMPLICIT AND EXPLICIT GENDER BIAS ON GRADING, AND THE EFFECTIVENESS OF RUBRICS FOR REDUCING BIAS

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

By

SARAH MARIE JACKSON
B.A., The Ohio State University, 2002
M.S., Wright State University, 2011

2016
Wright State University

WRIGHT STATE UNIVERSITY

GRADUATE SCHOOL

April 29, 2016

I HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER MY SUPERVISION BY Sarah Marie Jackson ENTITLED The Influence of Implicit and Explicit Gender Bias on Grading, and the Effectiveness of Rubrics for Reducing Bias BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Tamera Schneider, Ph.D.
Dissertation Director

Scott Watamaniuk, Ph.D.
Graduate Program Director

Debra Steele-Johnson, Ph.D.
Chair, Department of Psychology

Robert E. W. Fyffe, Ph.D.
Vice President for Research and
Dean of the Graduate School

Committee on
Final Examination

Tamera Schneider, Ph.D.

Gary Burns, Ph.D.

Martin Gooden, Ph.D.

Kevin Bennett, Ph.D.

ABSTRACT

Jackson, Sarah M. Ph.D., Department of Psychology, Wright State University, 2016.
The Influence of Implicit and Explicit Gender Bias on Grading, and the Effectiveness of Rubrics for Reducing Bias.

The effect of implicit bias on discriminatory grading in education has received considerable attention but, to date, no study has examined the effectiveness of using a rubric to reduce biased grading. Current research has demonstrated that the presence of a gender-normative name is sufficient to activate implicit gender bias, which can result in disparate treatment. The purpose of this study was to examine the effects of implicit and explicit gender bias on grading decisions for written assignments. When grading identical essays on the topic of computers (stereotypically-male), participants assigned significantly lower grades when the essay was supposedly written by a female author, compared to a male author. This difference was more pronounced in participants who had a stronger implicit association of men with science (high implicit bias). Male and female author grades did not differ when assigned by participants who were low in implicit bias. Further, participants who were high in implicit bias, but reported low explicit prejudice toward women in STEM graded the female author more harshly than the male author. This study also investigated the effectiveness of using a rubric to decrease bias effects on grading. Unexpectedly, use of the rubric enhanced the effect of implicit bias on grading when the author gender and essay topic were stereotype-inconsistent (i.e. female computer author). It is possible that rubric use further depleted cognitive resources already limited by dissonant implicit and explicit attitudes. While rubrics might increase the perception of objectivity, they might also inadvertently serve to amplify the effect of implicit gender bias when the topic being graded is strongly-gender normative.

TABLE OF CONTENTS

I. INTRODUCTION AND PURPOSE 1

Stereotypes, Discrimination, and Implicit Bias 4

Gender Discrimination..6

Bias in Online Education 12

Educational Performance Assessments and Rubrics 14

Purpose..17

Hypotheses..18

II. METHOD...23

Design 23

Participants..23

Power Analysis 25

Task Apparatus 25

Materials 25

Measures 27

Procedure 31

III. RESULTS 32

Manipulation check...32

Evidence of Implicit Bias Activation 32

Effects of Implicit and Explicit Attitudes on Essay Grades..........................37

Effect of Author Gender on Essay Grades 41

Effect of Implicit Bias on Grades, by Author Gender 43

Effects of Implicit & Explicit Attitudes on Grades, by Author Gender 48

Effectiveness of Rubrics to Increase Consistency in Grading 61

Effectiveness of Rubrics to Decrease Bias Impact on Grading 62

Gender Differences in Implicit and Explicit Attitude Scores 75

Interaction of Author Gender, Rubric, & Participant Gender 77

IV. DISCUSSION..80

Summary of Results..81

Theoretical Implications..92

Practical Implications..93

Limitations 96

Future Research 98

Conclusion 99

V. REFERENCES 101

LIST OF FIGURES

Figure Page

1. Computer essay grades by gender-topic & implicit bias group.........................46

2. Computer essay grades by implicit bias & explicit sexism 49

3. Exercise essay grades by implicit bias & explicit sexism 50

4. Computer essay grades by implicit bias & explicit prejudice 51

5. Exercise essay grades by implicit bias & explicit prejudice.............................52

6a. Computer grades by implicit bias & explicit sexism, anonymous 55

6b. Computer grades by implicit bias & explicit sexism, female author 56

6c. Computer grades by implicit bias & explicit sexism, male author 56

7a. Computer grades by implicit bias & explicit prejudice, anonymous...............59

7b. Computer grades by implicit bias & explicit prejudice, female author 60

7c. Computer grades by implicit bias & explicit prejudice, male author...............60

8. Computer grades by gender-topic & rubric condition 68

9a. Computer grades by author gender & implicit bias, rubric condition..............74

9b. Computer grades by author gender & implicit bias, no rubric condition 74

LIST OF TABLES

Table Page

1. Number of Participants Per Study Condition, by Participant Gender...............24

2. Reading Statistics for the Computer and Exercise Essays...............................26

3. Pronoun Used in Participants' Descriptions of Essay Author 33

4. Frequency of Computer Author Descriptors by Author Gender 36

5. Correlation Matrix for Key Study Variables 38

6a. Multiple Regression: Implicit & Explicit Attitudes on Computer Grades......39

6b. Multiple Regression: Implicit & Explicit Attitudes on Exercise Grades........40

7a. MANOVA: Computer & Exercise Grades (%) by Gender-Topic................. 42

7b. Planned Contrast: Computer & Exercise Grades by Gender-Topic 43

8a. MANOVA: Computer & Exercise Grades by Gender-Topic
& Implicit Bias 45

8b. Planned Contrasts: Computer Grades by Gender-Topic & Implicit Bias.......47

8c. Planned Contrasts: Exercise Grades by Gender-Topic & Implicit Bias 48

9. Implicit Bias & Explicit Sexism on Computer Grades (%) by
Author Gender...54

10. Implicit Bias & Explicit Prejudice on Computer Grades (%) by
Author Gender...58

11. Hierarchical Regression: Computer Grades by Implicit Bias & Rubric..........63

12. Hierarchical Regression: Computer Grades by Implicit Bias &
Rubric, Split by Author Gender 65

13a. MANOVA: Computer & Exercise Grades (%) by Gender-Topic
& Rubric 67

13b. Planned Contrasts: Computer & Exercise Grades (%) by
Gender-Topic & Rubric 69

14a. MANOVA: Computer & Exercise Grades (%) by Gender-Topic, Implicit Bias & Rubric 70

14b. Planned Contrasts: Computer Grades by Gender-Topic, Implicit Bias, & Rubric 72

14c. Computer Grades (%) by Gender-Topic, Implicit Bias, & Rubric 73

15. Computer & Exercise Grades (%) by Implicit Bias, Explicit Attitudes, & Participant Gender 76

16. Computer Grades (%) by Gender-Topic, Rubric, & Participant Gender.........78

17. Exercise Grades (%) by Gender-Topic, Rubric, & Participant Gender...........79

LIST OF APPENDICES

Appendix Page

A: Online Learning Management System (LMS) 113

B: Computer and Exercise Essays...116

C: Rubric (grade sheet used in the Rubric condition)..118

D: Grade Sheet (used in the No Rubric condition)..119

E: Instructions and Stimuli Used for Implicit Association Test...........................120

F: Modern Sexism Scale (Swim et al., 1995) 121

G: Women in STEM Stereotype Scale (Jackson et al., 2014) 122

H: Semantic Differential Scale Items 123

I: Feeling Thermometers...124

J: Experimental Laboratory Layout 125

K: Procedural Flowchart...126

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Tamera Schneider, for her advice and guidance. I would also like to thank my committee members, Dr. Kevin Bennett, Dr. Gary Burns and Dr. Martin Gooden, for their feedback and contributions.

I couldn’t have made it through this journey without the love and support of my amazing family. Thank you especially to my husband and best friend, Mark, who never gave up on me, and always had my back. Thank you to my parents, Russ and Jan, whose passions for learning and teaching inspired me from an early age. I am, and will always be, a life-long learner thanks to you. Thank you to my daughter, Olivia, for your patience and love and joyous spirit, even on those long days and nights when I couldn’t play or read or cuddle as much as I wanted to. I hope you always know how much I love you, and that one day you’ll understand that all these sacrifices were made for you.

Finally, I would like to thank my friends who have been there for me along the way. Your own dedication and perseverance inspired and encouraged me. Thank you for providing a listening ear, a shoulder to cry on, or a good laugh when needed.

I. INTRODUCTION AND PURPOSE

> If [people] (male or female) conclude that women are inferior, [their] perceptions of women – their personalities, behavior, abilities, and accomplishments – will tend to be colored by [their] low expectations of women. ...whatever the facts about sex differences, anti-feminism – like any other prejudice – *distorts perception and experience*. What defines anti-feminism is not so much the belief that women are inferior, as allowing that belief to distort one's perceptions of women. More generally, it is not the partiality itself, but the distortion born of that partiality, that defines prejudice (italics in original, Goldberg, 1968, p. 29).

Goldberg viewed prejudicial action as a conscious decision based on distorted perceptions about a target group, in this case, women. This further implies that biased ratings of written work in favor of male authors (and against female authors) was the result of an explicit belief that women were inferior to men: "Women seem to think that men are better at *everything* (italics in original, Goldberg, 1968, p. 30). The participants in his study were described as unwilling to concede that women's competence could be comparable to the competence of men. The idea that one might possess conflicting attitudes, one explicit and one implicit, would not enter the scientific dialogue for years to come. There was no consideration given to the possibility that unconscious associations could affect behavior, resulting in prejudicial outcomes, despite favorable explicit attitudes. Goldberg concludes his study by answering his own question, "Is the intellectual double-standard really dead? Not at all..." (p. 30). Nearly 50 years later, this question is still relevant. Explicit attitudes regarding women have become increasingly more favorable (Buchmann, 2004; Mladnic & Eagly, 1994), yet disparate treatment of women still occurs across a variety of professional fields, especially those areas that are traditionally considered predominantly male (Devine, 1989; Heilman & Eagly, 2008; Fuchs, Tamkins, Heilman, & Wallen, 2004; National Research Council [NRC], 2007; Nosek et al., 2009).

Stereotypes can be damaging to women in STEM through several modes, as discrimination resulting from prejudice can impact education, hiring, promotion, retention, and availability of resources (National Academy of Science, 2006). The implicit biases held by both men and women can significantly hinder the success of women who choose to enter STEM fields, and gender stereotypes can prevent women from initially entering STEM fields in the first place. Women who choose to major in fields related to computers, technology, engineering, and math report increased overt and covert hostility, and are frequently one of only a few (if not the only one) in these courses (Morganson, Jones, & Major, 2010).

Online education programs have grown increasingly commonplace throughout the United States in recent decades. In 2013, approximately 7.1 million students in the United States took at least one online course, and the vast majority of all institutions of higher education offer online learning options (Allen & Seaman, 2014). Even among traditionally “brick-and-mortar” institutions, online-only programs are becoming acceptable routes for degree completion, and web-enhanced courses have now become the norm. Although some authors have argued that the online learning format should reduce discrimination toward disadvantaged groups by creating a more even playing field (Koenig, 2015), stereotypes and bias can continue to result in disparate treatment toward disadvantaged groups, even in programs that are entirely online (Postmes & Spears, 2002).

The social structures in face-to-face classrooms continue to persist in virtual environments, and therefore continue to result in discriminatory behavior (Gunn et al., 2002). Grades are among the most important methods used to assess learning outcomes; when discrimination impacts grading, the effects are wide-reaching, affecting social, emotional, and academic outcomes (Tierney & Simon, 2004). When discrimination takes more subtle forms, it can be more difficult to address. People are often unaware of the ways in which implicit bias can affect their behaviors (Carnes et al., 2012; Chen & Bargh, 1997; Devine, 1989). Mere exposure to a normative name is sufficient to activate bias resulting in inaccurate or unfair assessments, and this effect can manifest even without direct exposure to targets (Budden et al., 2007; Easterly & Ricard, 2011; Spelke & Grace, 2007; Towers, 2008; Trix & Psenka, 2003). The greatest risk of biased grading occurs when grading expectations are more subjective in nature. Rubrics are designed not only to communicate expectations to students, but also to increase reliability and objectivity in grading.

The purpose of this study was to examine the effects of implicit and explicit gender bias on grading decisions for written assignments. This study also investigated the moderating effect of a rubric on grading bias, to determine whether the use of a rubric would reduce the effect of bias on grading. Finally, this study evaluated whether the use of a measure of implicit associations could predict grading outcomes above and beyond explicit attitude measures.

Stereotypes, Discrimination, and Implicit Bias

Explicit attitudes have been defined as psychological tendencies to evaluate a target with favor or disfavor (Eagly & Chaiken, 1993); in contrast, implicit attitudes reflect automatic psychological tendencies or social cognitions that are purported to be outside the control of the individual (Aberson & Haag, 2007; Greenwald & Banaji, 1995; Greenwald et al., 1998). Implicit and explicit attitudes reflect beliefs, feelings, and associations that originate from a number of sources. It is necessary to first examine some of those sources, including stereotypes, prejudice, and discrimination.

Stereotypes are automatic, oversimplified attitudes toward a target group, and may be favorable or unfavorable (Allport, 1954; Eagly & Chaiken, 1993). To a degree, stereotyping is a natural part of cognitive processing in the same way that schematic heuristics are; they can improve information processing efficiency by allowing an individual to create organizing categories and make generalizations based on selective attention to specific identifying features (Allport, 1954; MacCrae, Milne, & Bodenhausen, 1994). Stereotypes can help people efficiently make decisions about how to interact with others and they help maintain self-image, group esteem, and in-group identification (Maccrae, Milne, & Bodenhausen, 1994). Individuals who are low in prejudice possess the same knowledge about the stereotypes that exist toward target groups (Devine, 1989). Individuals who are high in explicit prejudice are more likely to discriminate, and overt prejudice can take the form of hostility and intentional discrimination.

Stereotypes and prejudice reflect some of the cognitive components of attitudes, while discrimination reflects a behavioral component (Devine, 1989; Hackney, 2005); as a result, stereotyping may or may not result in discriminatory behavior. Knowledge of a stereotype does not always equate to high explicit prejudice (Olson & Fazio, 2004). Effortful cognitive processing can be employed to control or change behaviors to reduce discrimination (Petty, Wegener, & White, 1998). However, because stereotypes are activated automatically when one is exposed to a target, people might not be consciously aware of how these unconscious associations can affect their behaviors (Carnes et al., 2012; Chen & Bargh, 1997; Devine, 1989; Greenwald & Banaji, 1995; Rudman, Ashmore, & Gary, 2001). Individuals who report low levels of explicit prejudice can also engage in discriminatory behaviors, whether they are aware of these disparate outcomes or not. One source of this unintentional discrimination is implicit bias.

The theory of implicit social cognition holds that past experiences and exposures to target groups affect behaviors, even when the experiences are not consciously recalled or available to introspection, and therefore not consciously available for self-report (Greenwald & Banaji, 1995; Olson & Fazio, 2004; Rudman, 2004). Implicit bias can conflict with an individual's explicit attitudes and affect behaviors and decision-making. Implicit and explicit measures often correlate weakly at best (Hofmann, Gawronski, Gschwendner, & Schmitt, 2005; Karpinski & Hilton, 2001; Olson & Fazio, 2004). Relying on self-reports alone, one might mistakenly believe that stereotypes toward women and minorities have been reduced to the point of no longer being a concern

(Eagly & Mladnic, 1994). Despite these explicit reports, discrimination and disparate treatment persist (Christopher & Wojda, 2008). Differences in accessibility, activation, and awareness between implicit and explicit attitudes explain this dissociation. Implicit associations can result in discrimination, even when people see themselves as egalitarian, and have no explicit intention to discriminate. People are often unaware that their unconscious associations can influence their behavior. As a result, despite the fact that they disagree with overt prejudice, prejudicial outcomes can occur if they do not consciously engage their egalitarian beliefs (Devine, 1989).

Gender Discrimination

Stereotypes regarding gender can be activated by subtle cues, even in the absence of direct contact with a target. Something as simple as a stereotypically-normative name (i.e. male versus female), can be sufficient. To test this assumption, Goldberg (1968) instructed female undergraduate students to evaluate the quality of six articles. The articles were identical for all participants apart from the author names, which were either male or female. Goldberg reported that women rated the male essays higher than female essays, whether the articles were from traditionally masculine or feminine fields. However, significant differences were only found in the three fields that were considered masculine: city planning, linguistics, and law. Further, despite Goldberg's explicit conclusion that "[w]omen seem to think that men are better at *everything*" (1968, p. 30, emphasis in original), there were no significant differences in ratings on the feminine topics (art history, dietetics, and education). Regardless of topic, participants rated the author as more competent (one of the 9 dimensions rated) when they thought it was a male (Goldberg, 1968).

Subsequent replications of this study have mixed results. Pheterson, Kiesler, and Goldberg (1971) found that when the competency of the female author was unambiguous, evaluation differences were negligible. Women devalued female authors only when the authors' achievements were not made clear. Both of these studies were limited, however, in that they each used only female participants. Levenson et al. (1975) included both male and female undergraduates in a series of studies. In their first study, they attempted to replicate Goldberg's original findings. They found no significant differences by author name or participant gender, and no significant interactions. In their second study, they recruited male and female undergraduates in a political science class to evaluate an essay supposedly written for that course in order to control for participant-level knowledge of the topics being rated. They found no significant difference in evaluations made by male participants, but female participants rated female-authored essays higher than male-authored essays. A meta-analysis of 123 studies using the Goldberg paradigm found a main effect of gender: female authors received lower ratings than male authors, although the effect sizes were small (Swim et al., 1989). They further found that male authors were rated more favorably than female authors when the topics were masculine rather than feminine.

Goldberg's original hypothesis was that women explicitly devalued the work of other women, and their evaluations reflected conscious beliefs. There is a considerable body of research identifying explicit stereotypes regarding women. Most people are aware of the stereotypes that exist in society regarding the types of roles men and women should occupy (Heilman & Eagly, 2008; Fuchs et al., 2004). Prescriptive gender stereotypes are beliefs that members of a society possess about the kinds of characteristics that men and women should exhibit (Prentice & Carranza, 2002). When women violate these prescriptive norms, they are met with discrimination through disparate impact or disparate treatment. Discrimination can also take the form of either hostile sexism or benevolent sexism (Christopher & Wojda, 2008). Hostile sexism occurs when discrimination is overt, resulting from negative beliefs about women. In contrast, benevolent sexism occurs when discrimination is covert, resulting from positive beliefs about women (e.g. they are agreeable, supportive, and nurturing), but also from stereotypes that demean women (e.g. they are weak, overly emotional, passive, and in need of protection). The result of either type of sexism is maintenance of the status hierarchy. As a result, women are less likely to be hired, promoted, or offered leadership positions, particularly in fields that are viewed as traditionally masculine.

Women in STEM fields are often the target of each of these forms of discrimination, affecting education, hiring, promotion, retention, availability of resources, and even the likelihood of entering to STEM fields of study in the first place (National Academy of Science, 2006). A number of specific stereotypes are widely held about women, such as beliefs that they are not good at math, are not competitive or assertive, and that women faculty are less productive in their research and more interested in family than in careers (National Academy of Science, 2006). The belief that men are more inclined to participate and excel in math and science is widely held, even among women (National Academy of Science, 2006; Nosek, Banaji, & Greenwald, 2002).

As noted above, women are generally evaluated favorably, and most people see themselves as egalitarian. Although most people report positive attitudes toward women (Buchmann, DiPrete, & McDaniel, 2008; Eagly & Mladinic, 1994), research continues to support the existence and impact of unconscious gender bias. For example, when presented with a male or female computer avatar, participants were less likely to trust the advice of female avatars over male avatars (Webb, 2001). This main effect of avatar gender occurred regardless of participant gender, and in the absence of explicit favoritism toward advice from males or females. In another study, participants were randomly assigned to a “tutor” computer that was programmed with either a male or female voice (Nass, Moon, & Green, 2007). The tutor computer provided information on either love and relationships or computers and technology. After the tutoring session, participants completed a test to evaluate what they learned. Then, an evaluator computer that also had either a male or female voice, gave feedback to participants about their test performance. They were told upfront that the tutoring and evaluation programs could have been written by either a man or woman and that the voice they were hearing did not necessarily reflect the gender of the programmer. Despite participants’ self-reported beliefs that gender stereotyping a computer is illogical, the researchers found that the male-voiced evaluator was rated as more competent and friendlier than the female-voiced evaluator computer across all conditions. When the evaluator computer was male, subjects reported that the female tutor was more informative on feminine topics such as love and relationships, while the male tutored was reported to be more informative on masculine topics like computers and technology.

Implicit gender bias extends beyond the laboratory, with significant disadvantages occurring in both academic and professional settings. Resumé studies have shown that identical resumes labeled with male versus female names tend to result in a number of advantages for men: more positive evaluations (Uhlmann & Cohen, 2005), greater chance of selection (Gill, 2003; Koch et al., 2015), and higher starting salary (Lips, 2013). Given identical application packages, both male and female university psychology professors preferred the name Brian over the name Karen twice as often (Steinpreis, Anders, & Ritzke, 1999). In hiring decisions, men not only have an advantage over women, women who reveal that they are mothers are further penalized in terms of perceived competence and commitment, performance and punctuality standards, starting salary, and recommendations for hiring (Benard, Palik, & Correll, 2007). Men who reveal that they are fathers not only escape penalization, but in some cases they benefit further as a result of their parental status.

Long before candidates seek out employment, they are subjected to implicit bias in educational contexts. In early school years, research finds that girls and boys perform similarly, yet as children age a gender gap appears with girls scoring higher on verbal skills and boys scoring higher on math skills (Buchmann et al., 2008). While some earlier researchers suggested that the gap in performance was a result of biological differences rather than environmental differences (e.g. Pearson, 1987), more recent work has consistently shown that many of the differences can be attributed to environmental factors, including implicit bias held by the students themselves (Nosek, Banaji, & Greenwald, 2002; Steele & Aronson, 1995). One study comparing automatic bias among women in a coeducational college and a women's college found that automatic gender stereotypes increased for students after only one year of college (Dasgupta & Asgari, 2004). Even current researchers who suggest a biological component tend to concede that boys and girls share an equal aptitude for math and science (e.g. Spelke, 2005). In addition to the effect of self-selection and self-fulfilling prophecy that can result from implicit biases, it has been shown that teachers, as early as kindergarten level, demonstrate biased evaluations of male and female students' math performance (Robinson-Cimpian et al., 2014). Underrating girls' performance from an early age is likely to account partly for the gaps in ability that appear in later educational contexts, despite a lack of differences in early elementary school.

These differences continue throughout the schooling experience, and follow students into college. Milkman, Akinola, & Chugh (2014) found that faculty were more likely to respond to an email request for a meeting when they believed the message came from a man rather than a woman. This occurred across all fields, including business, education, human services, engineering, science, and math. Subtle gender biases have resulted in less support for female students in science fields, and science faculty preferred male applicants over female applicants when hiring for a laboratory manager position (Moss-Racusin et al., 2012). Another study found that when faculty members wrote recommendation letters for medical school applicants, the letters were longer for men compared to women and they contained more references to the male student's curriculum vita and accomplishments (Trix & Psenka, 2003). Letters for women were shorter, contained more references to the student's personal life, and included more irrelevant or "doubt-raising" comments. Similarly, performance evaluations of medical students included adjectives reflecting gender bias; women were more likely to be described as "compassionate", "sensitive", and "enthusiastic", whereas men were more likely to be described as "quick learners" (Axelson et al., 2010). This gender difference increased as student proficiency increased; at higher rates of performance, the biased differences between men's and women's evaluations became even more pronounced.

There is a widely-held implicit belief that women are better in school and better writers. While women may be favorably assessed for general academic ability and for writing ability (or penalized more harshly for poor writing), stereotypes regarding women's competence, intelligence, emotional stability, and others remain (Buchmann et al., 2008; Prentice & Carranza, 2002). Some stereotypes are shifting to a more equitable level. For example, a once large divergence between descriptions of men and women as being "nerdy" or "geeky" has now diminished such that there are no longer differences (Buchmann et al., 2008). On the other hand, people are still significantly more likely to refer to video games and computers when referencing males, compared to females (Buchmann et al., 2008). Knowledge of these stereotypes strengthens unconscious gender associations. Most people implicitly associate men with science more than women with science (Nosek et al., 2009). Weak implicit associations of women being linked to STEM fields may partly help explain why women faculty are paid less, promoted more slowly, receive fewer honors, and are given fewer leadership positions than men, despite there being no significant gender differences in knowledge, ability, or productivity (NRC, 2007).

Bias in Online Education

Gender is frequently mentioned in the literature on web-based learning, but gender bias in online education is rarely (Garland & Martin, 2005). Most empirical studies suggest that the perception of the online environment as being democratic and equalizing is naturally flawed, because the complex sociocultural relationships and resulting imbalances remain despite the use of computer communication (Gunn et al., 2002; Wolfe, 1999).

Universities first started supplementing courses by email and computer conferences in the mid-1970s (Harasim, 2000). Online courses were made available in adult non-credit education and executive training programs as early as 1981, and the first online undergraduate courses were introduced in 1984. These developments occurred even before the official launch of the Internet in 1989 and before the invention of the World Wide Web in 1992. Since then, the use of computer technology in classrooms has grown significantly. Today, a majority of degree programs employ a web-enhanced modality (Allen & Seaman, 2014; Gunn et al., 2002; Harasim, 2000), and virtually all public institutions have at least some online course offerings (Allen & Seaman, 2014). A web-enhanced course, also known as a computer-supported learning environment (CSL), is an educational setting where computer networking complements the traditional classroom environment, providing a platform for communication, learning, and administrative tasks (Gunn et al., 2002; Harasim, 2000). Most universities now recognize that online education provides an efficient and effective way to meet student needs and many researchers have found that web-based courses are as effective as traditional classroom formats (Allen & Seaman, 2014: Hamann, Pollock, & Wilson, 2008). Given the widespread use of computer-based interaction in education, it is necessary to study how biases can affect behaviors in the online classroom. This is particularly true regarding grades and performance evaluations, which predict student success in the form of course completion, degree completion, credit transfer, GPA, and admission into graduate schools, to name a few.

More women than men enroll in online courses (Garland & Martin, 2005). Female nontraditional students have reported that online classrooms reduce their feelings of discomfort and alienation compared to face-to-face environments (American Association of University Women, 2001). However, once computer access and computer literacy are controlled for, gender-based interactions and inequities found in face-to-face classrooms continue to persist online, dispelling the myth that technology provides a gender-neutral and equitable learning environment (Gunn et al., 2002; Postmes & Spears, 2002; Wolfe, 1999). Some researchers have found that women thrive in online environments, whereas younger male students achieve at a lower level (Gunn et al., 2002; Siann & Callaghan, 2001; Kleinfeld, 1998). This difference has been attributed to beliefs that women are more motivated, have greater ability to work independently, and can more effectively multi-task (Gunn et al., 2002). However, while women tend to fare better academically overall, differences in grading outcomes still exist in areas that are more strongly associated with men, such as science, technology, engineering, and math (Ackerman, Kanfer, & Beier, 2013; Buchmann et al., 2007).

Educational Performance Assessments and Rubrics

Most educators are aware of the possibility that subjective evaluations can unintentionally be influenced by personal bias. Objective criteria and assessment tools, such as rubrics, are often employed in an attempt to reduce this possibility, while also increasing consistency and transparency in grading. A rubric is typically defined as an assessment tool that describes expectations for performance quality (rating score) across different dimensions (criteria) on a particular task (Hafner & Hafner, 2003; Jonsson & Svingby, 2007; Reddy & Andrade, 2010). The three primary features of a rubric are a set of evaluation criteria, definitions of quality for each criterion, and a scoring guide (Popham, 1997). The criteria identify what is most important in the assignment, and the scoring guide describes what the grader should look for when determining the quality of a particular criterion, typically represented on a numeric scale ranging from 0 (poor) to excellent (4 or 5).

Rubrics are used across a wide range of disciplines in higher education, and can be used for several reasons (Jonsson & Svingby, 2007).   Rubrics improve efficiency in grading, quantify and clarify expectations, increase objectivity, and promote fairness and satisfaction (Jonsson & Svingby, 2007; Rippé, 2008). Rubrics are also used to provide feedback to students and to enhance learning and teaching (Reddy & Andrade, 2010). Students generally express positive perceptions of rubric use, citing the benefits of clear expectations and increased perceived fairness (Reddy & Andrade, 2010). Instructors, on the other hand, are at times resistant to using rubrics. Their reluctance is in part because most higher education instructors have little or no pedagogical preparation as teachers and because there is a commonly held belief that rubrics require a great deal of time and effort (Hafner & Hafner, 2003; Reddy & Andrade, 2010). Despite the reluctance found among some educators, rubrics are generally highly regarded due to the perceptions that they increase reliability and validity. A number of researchers have reported increased reliability in the presence of a rubric (Jonsson & Svingby, 2007; Renzai & Lovorn, 2010; Silvestri & Oescher, 2006), whereas no research has revealed any negative effects resulting from rubric use (Renzai & Lovorn, 2010).

There are two primary ways of measuring the effectiveness of a rubric: consensus and consistency (Reddy & Andrade, 2010). Consensus typically involves examining the proportion of ratings that match an expert evaluation (either identical in scoring, or falling within a certain acceptable scoring range). Consistency is often evaluated using inter-rater reliability. Not all researchers have found that rubrics result in consistent grading outcomes. For example, within medical training programs, validated scoring instruments are commonplace, yet there remains significant variability among faculty assessments of student performance (Ottolini et al., 2007). Oakleaf (2006) examined consistency and consensus in a group of raters on a literacy skill assessment. They found that consistency was adequate, but consensus (complete agreement) was far below acceptable levels.

On the other hand, several studies have shown rubrics to be effective in reliably assessing performance. Hafner and Hafner (2003) compared peer-grading and instructor grading in an undergraduate course and found significant consensus and consistency. Simon and Forgette-Giroux (2001) compared instructor grades with undergraduate self-assessments using a rubric and found that instructors and students reached consensus 75% of the time. Researchers examining essay grading without the use of a rubric have shown significant variability in grades, further supporting the use of a rubric to decrease grading variance. Gage and Berliner (1992) recruited experienced teachers to grade an identical essay without a rubric, and they found a great deal of variability in scores between teachers. On a scale from 0 to 100, the teacher grades ranged from 60 to the upper 90s, and teachers' evaluation of the essay writer's grade level also varied considerably. Most of the research on rubric use refers to the increase in consistency as a primary way of evaluating the effectiveness on rubrics.

When rubrics are used to increase consistency and decrease variability resulting from bias, grading with a rubric is likely more reliable than grading without a rubric (Jonsson & Svingby, 2007). There are a number of elements that can be employed to enhance consistency when designing a rubric. Rubrics that are analytic, topic-specific, include exemplars, and are complemented with rater training tend to be more reliable (Jonsson & Svingby, 2007). Rubrics are most effective when the number of criteria assessed is kept to a minimum (i.e. less than 10; Rhodes, 2010). The language in the rubric must be clear and consistent because ambiguity cannot be interpreted accurately by graders or students (Reddy & Andrade, 2010; Payne, 2003). One short-coming of some rubrics is a lack of narrative anchor, which is more likely to result in disparate scoring and less inter-rater reliability (Ottolini et al., 2007). Thus, the addition of narrative descriptions or the practice of encouraging raters to reflect upon their grading decisions in a narrative fashion is preferred in the design and implementation of rubrics.

Purpose

The purpose of this study was to examine the effects of implicit and explicit gender bias on grading decisions for written assignments. This study also investigated whether the use of a rubric would reduce the effect of bias on grading. Finally, this study evaluated whether implicit association and explicit attitude measures could explain a significant amount of variance in grading outcomes. Participants graded identical essays with manipulated author gender names (anonymous, female, or male) using either a rubric or no rubric. This grading task was followed by a series of implicit and explicit measures of gender bias, and a set of questions regarding participants’ impressions of the authors whose work they ostensibly read. This design was intended to elicit biased responses depending on author gender, which in turn would provide the opportunity to study the use of a rubric to reduce discrimination in grading.

Hypotheses

Evidence of Activation of Implicit Bias

Most people possess a stronger implicit association of men with STEM rather than women with STEM (Nosek et al., 2009). Although previously common explicit gender stereotypes are less common today (e.g. women are less intelligent or competent; men can be nerds, but women cannot), prescriptive gender norms continue to influence evaluations of women and men (Rudman & Glick, 2001; Prentice & Carranza, 2002). Faculty are more likely to respond to an email request from a male student (Milkman, Akinola, & Chugh, 2014), letters for female medical school applicants are shorter and contain more doubt raisers (Trix & Psenka, 2003), and performances evaluations for medical students contain more descriptions of men as quick learners and women as compassionate (Axelson et al., 2010). Based on these findings, in the current study, participants' descriptions of the author of the computer essay (a STEM topic) were expected to reflect implicit associations between author gender and essay topic.

Hypothesis 1. A: In the anonymous condition, participants will use a male pronoun ('he') to describe the author of the computer essay more often than a female pronoun ('she'), and participants will ascribe male and female pronouns equally to the anonymous exercise essay.

Hypothesis 1. B: Participants will describe female authors using fewer descriptive words and fewer words overall compared to when they describe male authors.

Hypothesis 1. C: Participants will use descriptors to describe the male and female authors differently, revealing implicit gender norms.

Implicit and Explicit Bias and Their Effects on Essay Grades

A number of attitude researchers have found that measures of implicit bias are better predictors of discriminatory behaviors than explicit attitudes (Lane et al., 2012; Nosek & Smyth, 2011; Nosek et al., 2002; Steffens et al., 2010). Even in studies where explicit attitudes are found to significantly predict behaviors, once implicit bias is added to the statistical model, explicit measures become non-significant (Nosek et al., 2009). Similar findings were expected in the current research.

Hypothesis 2. A: The implicit association measure will correlate significantly with computer essay grades. Explicit attitude scores will correlate weakly with computer essay grades and weakly with implicit association scores.

Hypothesis 2. B: None of the implicit or explicit measures are expected to correlate significantly with exercise essay grades.

Hypothesis 2. C: Implicit gender-science association scores will explain a significant amount of variance in computer essay grades, and the IAT will explain a significant amount of variance above and beyond explicit measures.

Effect of Author Gender on Essay Grades

Several studies have found that female authors (e.g. of essays, articles, and blogs) receive lower ratings and rated less competent or credible that male authors (Armstrong & McAdams, 2009; Goldberg, 1969; Levenson et al., 1975; Swim et al. 1989). Because most people have a stronger implicit association of men with science (Nosek et al., 2009), and because both men and women can face backlash if they violate prescriptive gender norms (Prentice & Carranza, 2002), author gender was expected to affect essay grades positively when author gender and essay topic were stereotype-consistent and negatively when author gender and essay topic were stereotype-inconsistent.

*Hypothesis 3. A***:** Computer essay grades with the male author name (stereotype-consistent) will be higher than computer essay grades with the female author name (stereotype-inconsistent). The computer essay with no author name will receive grades that are equal to the grades assigned to the stereotype-consistent (male-computer) author gender-essay topic pairing.

Hypothesis 3. B: Exercise essay grades with the female author name (stereotype-consistent) will be higher than exercise essay grades with the male author name (stereotype-inconsistent). The exercise essay with no author name will receive grades that are equal to the grades assigned to the stereotype-consistent (female-exercise) author gender-essay topic pairing.

Interaction of Implicit and Explicit Attitudes on Grades, by Author Gender

High prejudice individuals are those who endorse stereotypes toward a target group, whereas low prejudice individuals do not endorse stereotypes. People who report low explicit prejudice, yet harbor high implicit bias might report positive attitudes as an intentional method of replacing stereotypes with egalitarian views, or they might do so because they are unaware of their personal implicit biases (Devine, 1989). When people report low explicit prejudice and high implicit bias, they tend to be more vigilant and more scrutinizing toward members of the target group (Devine et al., 1991; Monteith et al., 1993; Petty et al., 1999; Johnson et al., 2011). Therefore, implicit bias was expected to interact with explicit attitudes to affect essay grades.

Hypothesis 4. A: High-prejudice individuals will exhibit greater bias in grading than low-prejudice individuals.

Hypothesis 4. B: Participants with low explicit prejudice and high implicit prejudice will assign lower grades to female authors but not to male authors compared to the no name condition.

Effectiveness of Rubrics to Increase Consistency in Grading

In addition to improving efficiency and clarifying expectations, effective rubrics are increase reliability among raters (Jonsson & Svingby, 2007; Renzai & Lovorn, 2010; Rippé, 2008). The rubric used in the current research was expected to result in greater consistency within essay grades.

Hypothesis 5: Essay grades will have greater consistency (less variability) within the rubric condition compared to grades in the no-rubric condition.

Effectiveness of Rubrics to Decrease Bias Impact on Grading

By increasing consistency and reliability, rubrics are believed to increase objectivity and fairness in grading (Jonsson & Svingby, 2007; Renzai & Lovorn, 2010; Rippé, 2008). The rubric was expected to interact with author gender to affect essay grades, resulting in greater parity between grades assigned to different authors.

Hypothesis 6. A: There will be a significant difference in essay grades between author gender in the no-rubric condition, but not in the rubric condition. Differences observed among author gender grade assignments in the no rubric condition will become non-significant in the rubric condition.

Hypothesis 6. B: The rubric will interact with the IAT to reduce the effect of bias on grades, resulting in more equal grades among author genders.

Gender Differences in Implicit and Explicit Attitude Scores

Some attitude research has found little to no differences in implicit or explicit attitudes between male and female respondents (Nosek et al., 2009). Others, however, have found significant attitude differences by participant gender (Buchmann et al., 2007; Jackson, Hillard, & Schneider, 2014). Male participants' responses were expected to differ from those of female participants.

Hypothesis 7: Implicit bias and explicit stereotype endorsements were expected to be higher within male participants, compared to female participants.

Interaction of Author Gender, Rubric, and Participant Gender on Grading

In-group gender bias emerges early in childhood, with boys and girls evaluating members of their gender group more favorably. Girls tend to have stronger implicit own-gender preferences, but as men age, their implicit preferences begins to lean toward women as well (Dunham, Baron, & Banaji, 2015). Given the relatively young age of the subject pool from which this research drew its participants, it was expected that there would be evidence of in-group gender bias, but that rubric use would reduce this effect.

Hypothesis 8: There will be a three-way interaction between author gender, rubric, and participant gender. Male participants are expected to grade the male author more favorably, female participants are expected to grade the female more favorably, and rubric is expected to moderate this relationship.

II. METHOD

Design

The current study employed a between-subjects, factorial, experimental design. Manipulated independent variables were name of the essay author (anonymous, female, or male) and grade instructions (rubric or no rubric). Attitude variables included a measure of implicit gender-science bias and four measures of explicit gender attitudes. The dependent variables were the final grades (out of 20 points, converted to percentages) that participants assigned to each essay (computers, exercise). Participants were randomly assigned to receive a rubric or not. Order of essay topic presentation and author gender-essay topic combination (hereafter referred to as ‘gender-topic condition’) were completely counterbalanced. Stereotype-consistent gender-topic conditions were: (a) male author with computer essay and (b) female author with exercise essay. Stereotype-inconsistent gender-topic conditions were: (a) female author with computer essay, and (b) male author with exercise essay. In the anonymous condition, participants graded the same computer and exercise essays, but neither essay had an author name.

Participants

Participants were 216 undergraduate students (70% female, *n* = 151) taking an introductory psychology course at a midwestern university. Distribution of participants by study condition and participant gender were statistically equivalent across all cells (see Table 1). The average age was 21 years (ages ranged from 18 to 54). Of those participants who reported their ethnicity, 71% were White, 13% were African-American, 4% were Middle Eastern, 3% were Hispanic, 4% were mixed race, and 4% were other.

Table 1

Number of Participants Per Study Condition, by Participant Gender

				Participant Gender		
Condition	Grading Instructions	Stereotype Congruence	Gender-topic condition	Male	Female	*n* Per Condition
1	Rubric	Stereotype consistent	Male computer, female exercise	11	24	35
2	Rubric	Stereotype inconsistent	Female computer, male exercise	13	25	38
3	Rubric	Anonymous	Computer, exercise (no author name)	7	28	35
4	No Rubric	Stereotype consistent	Male computer, female exercise	15	21	36
5	No Rubric	Stereotype inconsistent	Female computer, male exercise	10	27	37
6	No Rubric	Anonymous	Computer, exercise (no author name)	9	26	35

Note. Total sample size = 216.

Power analysis.

Previous studies examining grading differences given for male and female authors have found mean effect sizes ranging from -0.08 (small) to -0.38 (medium; Swim et al., 1989). Using power tables from Cohen (1992), a sample size of 35 per condition will yield power of 0.80 with 6 groups (total *N* required = 210). This effect size estimation was based on planned contrasts and was expected to yield a small to moderate effect size.

Task Apparatus

Online course environment. Participants viewed the essays in the university's online Learning Management System (LMS; see Appendix A). All students enrolled in introductory psychology courses (currently or within the past 5 years) had prior experience with this system and were familiar with the way assignments are uploaded and reviewed. Essays were pre-loaded in the "Dropbox" folder under "Assessments", a feature in the LMS wherein students electronically upload assignments for their courses and receive grades and comments from their instructors. There were six folders ("writing sections") within Dropbox, each containing six files. The first two contained the experimental essays, and the remaining four were files that were deliberately manipulated so they did not open when clicked (an error would appear and further attempts to open the file would result in a warning stating that the file was corrupt). Participants believed they would be grading 6 essays, but only graded the first two.

Materials

Essays. Participants read two essays, ostensibly written by other students, on the topics of computers and exercise (see Appendix B). The essays were adapted from existing essays available at a free essay writing website. The resulting composite essays were reviewed by the researchers to ensure that no explicit information remained that might be suggestive of author gender. Reading statistics were comparable for both essays (see Table 2).

Table 2

Reading Statistics for the Computer and Exercise Essays

	Essay Topic	
Statistic	Computer	Exercise
Counts		
Words	421	456
Characters	2027	2101
Paragraphs	3	3
Sentences	30	34
Averages		
Sentences per paragraph	10.0	11.3
Words per sentence	14.0	13.4
Characters per word	4.7	4.5
Readability		
Passive Sentences	6%	5%
Flesch Reading Ease	58.5	66.1
Flesch-Kincaid Grade Level	8.5	7.3

Rubric/grading instructions. A blank rubric was provided in paper format to participants in the experimental condition (see Appendix C). The rubric contained the writing prompt, (described as the prompt given to the authors of the essays), participant instructions, and a list of 4 main objectives (e.g. content, writing mechanics, etc.). Each objective was evaluated on a scale from 0 (not acceptable or objective not present) to 5 (excellent). For each point value, a short narrative anchor was provided, describing in more detail what would constitute a rating at each level. Participants were instructed to write the point value assigned to each objective, then total all ratings for a combined grade out of a maximum of 20 points. The rubric also included a key providing percentage equivalents for each point value range. In the male and female author conditions, a blank was provided for "author name" (no blank was included in the no-name author condition).

Participants in the control condition (no rubric) received a grading sheet that contained the same writing prompt and instructions, blank for author name (in the male and female author conditions), and the percentage equivalent key (see Appendix D). Participants were instructed to grade the essay to the best of their ability, and then record the total score they assigned out of a maximum of 20 points in the space provided on the instruction sheet.

Measures

Essay grades. The dependent variables were the final grades (out of a maximum of 20 points and subsequently converted to percentages) assigned to each of the two essays.

Follow-up interview. Prior to the debrief, the researcher asked a series of follow-up questions regarding the essay grading task. These questions served both as a manipulation check, and as a measure of implicit gender bias. Participants were asked to describe their impressions of the two authors whose essays they read. In addition to recording the participants' descriptions, the researcher made note of the pronoun used to describe each author.

Gender-Science Implicit Association Test. The Implicit Association Test (IAT) assesses implicit attitudes and other automatic associations based on reaction times (Greenwald, McGhee, & Schwartz, 1998). Compared to self-report explicit attitude measures, the IAT is purported to be more resistant to validity threats such as social desirability. The IAT measures how quickly a participant classifies stimuli into categories. The target category contains the dichotomous aspects of the object attitude the researcher is interested in studying. The attribute category contains the valence of the attitudes. The traditional IAT measures how quickly participants associate dichotomous target groups (e.g. women or men) with favorable or unfavorable attributes (e.g. good or bad). The response time indicates the relative strength of association by assessing how quickly a participant can pair a target category with the attribute dimension. If a target category is associated with an attribute dimension that reflects the participant's implicit association, he or she should respond more quickly (Greenwald et al., 1998). Participants are instructed to correctly sort stimuli items as quickly as possible, to elicit responses that are instant, uncontrollable, and automatic.

Stereotype IATs, rather than traditional attitude (good-bad, pleasant-unpleasant) IATs were used in this study because the hypotheses are directly related to stereotypes regarding women with STEM, rather than a general positive or negative attribution. Furthermore, stereotype IATs have been shown to higher predictive validity than traditional attitude IATs (Rudman & Ashmore, 2007). The Gender-Science IAT was used in this study (Nosek et al., 2009). This IAT is intended to reveal the relative association between liberal arts or science and females or males. The Gender-Science IAT uses the target categories of "Male" and "Female", and the attribute categories of "Science" and "Liberal Arts". Stimuli used in this IAT, along with testing procedure, can be found in Appendix E.

Explicit attitude measures.

The Modern Sexism Scale (Swim et al., 1995) presents eight statements reflecting beliefs about women across three dimensions: (1) denial of continuing discrimination (e.g. "Discrimination against women is no longer a problem in the United States."), (2) antagonism toward women's demands (e.g. "It is easy to understand the anger of women's groups in America," reverse-scored) and (3) resentment about special favors for women (e.g. "The government and media have shown more concern about the treatment of women than is warranted by women's actual experiences.") Participants rate their agreement with each statement using a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree). A Principal Axis Factor analysis (PAF) with a Varimax rotation was conducted to ensure that the items loaded on the same factor. The 8-item scale was reliable (Cronbach's alpha = .80; see Appendix F).

The Women in STEM Stereotype Scale (Jackson, Hillard, & Schneider, 2014) presents 14 stereotype-derived statements (e.g. "Women are worse at math than men;" NAS, 2006). Participants rated their agreement with each statement using a 5-point

Likert scale. A Principal Axis Factor analysis (PAF) with a Varimax rotation was conducted for this relatively new scale. Items were retained when their loading was greater than .40 on that factor and less than .30 on any other factor. The final 10-item scale was reliable, with a Cronbach's alpha of .76 (see Appendix G).

A semantic differential scale contains a pair of dichotomous words of opposite meaning anchored on opposite ends of a numeric ratings scale containing number ratings from 1 to 5 spaced equally between the words. Participants are asked to rate a target group by circling where on the scale their beliefs about the target group falls. In the current study, participants were asked to rate "women" on each of 12 semantic differential scale items (adapted from Blair, Ma, & Lenton, 2001; Jackson, Hillard, & Schneider, 2013; Olson & Fazio, 2004). Items assessed general favorability toward women (e.g. good versus bad; favorable versus unfavorable) and stereotypes regarding women (e.g. analytical versus emotional; passive versus assertive). A Principal Axis Factor analysis (PAF) with a Varimax rotation was conducted for the 12 semantic differential items. Of the 12 semantic differential items, 8 were included in the final attitude scale, resulting in a Cronbach's alpha of .81 (see Appendix H). Because of the strong correlations and high factor loadings for these 8 items, scores for the semantic differential scales were collapsed to produce a single semantic differential average.

The final explicit attitude measure consisted of 6 feeling thermometer items. Participants were asked to rate on a scale from 0 (very cold/unfavorable) to 100 (very warm/favorable) their feelings toward each item. Of these 6 items, three were specifically about women (e.g. "female scientists"), three were about men (e.g. "male faculty”). Cronbach’s alpha for the combined feeling thermometer items was .80 (see Appendix I).

Procedure

Participants were told that they would be reading and evaluating a series of short written assignments. The researcher provided each participant with a grade sheet containing either a rubric with a space for the final grade, or a blank grade sheet with a space for the final grade. After explaining the task to the participant, the researcher opened the first and second essay in sequence, providing the same rubric or grade sheet each time. When the researcher attempted to open any additional files, the error message appeared on the screen, indicating that the files were corrupt and could not be opened.

At this point, the researcher apologized, explaining that there had been technical problems with this writing section in the past, and asked the participant if they would be willing to take part in a “second study” being conducted in the lab. They were informed that it was voluntary and that they would receive their full credit for participation either way. If the participant did not agree to do the second study, the researcher thanked them again for their time, provided a demographic survey, and read the debrief statement.

Participants who agreed to participate were then escorted to a different computer in the laboratory (see Appendix J). The participant then completed the Gender-Science IAT on the computer, followed by a pencil-and-paper copy of the explicit measures and demographics. At the end of the survey, the researcher asked a series of follow-up questions regarding both studies, read the debrief statement, and thanked them for their participation (see Appendix K for procedural flowchart).

III. RESULTS

Manipulation Check.

To ensure that the author name manipulation was effective, participants were asked to describe the author, and the pronoun used was tallied. For the computer essay, 80% of participants correctly described the male author as “he”. The remaining 20% used either the gender-neutral, singular “they” or did not use any pronoun. When the computer essay had a female author, 75% of participants correctly described the author as “she”, and 20% used “they” or no pronoun. For the exercise essay, there were no significant differences in pronoun use for either of the gendered author name conditions: 77% of participants correctly described the male exercise author as “he”, and 77% of participants correctly described the female exercise author as “she”; participants were equally likely to use the singular “they” or to use no pronoun, in both author gender conditions. These results indicate that the manipulation was effective.

Evidence of Activation of Implicit Bias

Pronouns used to describe the writer of the computer essay in the anonymous condition (no author name) were examined to investigate the activation of implicit bias. In the no author name condition, participants were expected to use pronouns that would reflect implicit associations between gender and essay topic. For the anonymous computer essay, participants were expected to use a male pronoun (‘he’) to describe the author. For the anonymous exercise essay, participants were expected to use male or female pronouns. In the computer condition, 56% of participants described the anonymous author as male, whereas only one participant described the author as female, χ^2 (5) = 42.82, *p* < .01. The remaining 42% used either a gender-neutral pronoun, or no pronoun. For the no name exercise essay, pronoun use was approximately evenly distributed across male and female pronouns; none of the observed pronoun categories differed significantly from expected values, all *p*s > .47 (see Table 3).

Table 3

Pronoun Used in Participants' Descriptions of Essay Author

	Pronoun									
	he		she		they		none			
Condition	*n*	%	*n*	%	*n*	%	*n*	%	total	χ^2
Computer essay										
Anonymous	39	56%	1	1%	22	31%	8	11%	70	42.82**
Female author	4	5%	56	75%	9	12%	6	8%	75	14.51**
Male author	57	80%	0	0%	10	14%	4	6%	71	0.28
Exercise essay										
Anonymous	18	26%	19	27%	26	37%	7	10%	70	1.40
Female author	2	3%	55	77%	10	14%	4	6%	71	2.54
Male author	58	77%	1	1%	11	15%	5	7%	75	0.12

Note: ***p* < .01.

Participants' descriptions of authors were also expected to reflect implicit gender norms. Descriptions given for the female author were expected to be shorter, contain more references to appearance than intelligence, have fewer references to video games and to interest in STEM majors or careers, and be described as more extraverted, agreeable, and emotional than males. Descriptions of the male author were expected to include more references to intelligence, conscientiousness, and introversion, and the male was expected to be described as nerdy and anti-social. An independent-samples *t*-test was conducted to compare word count for the male author and female author descriptions. To compare descriptions of the male and female author, comments were coding in three stages (based on Moni, Beswick, & Moni, 2005): (1) open coding: each concept from each description was recorded on a separate line, (2) axial coding: key words were tallied, and words with similar stems or definitions were grouped together, and (3) selective coding: descriptors were then collapsed into categories based on similar or related meanings. Chi-square analyses were then conducted to compare frequencies of descriptive words within each category.

As shown in Table 4, there were no differences in word count for male author or the female author of the computer essay. Number of descriptive words also did not differ between the male and female authors. The type of descriptive adjectives used to describe the computer essay differed depending on author gender, $\chi^2(9) = 23.92$, $p < .001$. Predictions were partially supported. There were no significant differences between the male and female author in references to intelligence or likelihood of majoring/seeking a career in STEM, yet the female author received significantly more criticisms regarding English and writing ability. There were no differences in descriptions of the male author of female author being “nerdy” or “geeky”, yet the female author received significantly fewer references to video games.

There were no significant differences in frequency of personality descriptors applied to the male and female authors, all *p*s > .11, although there appeared to be a trend toward females receiving more descriptions referencing low conscientiousness. A surprising marginal difference occurred in the number of descriptions referencing physical appearance. The male author received marginally more comments regarding the way participants imagined he looked, compared to the female author. The category labeled “doubt-raisers” included four references to the female author and zero references to the male author. These references implied that the author was disingenuous or not serious about the topic. Expected frequencies were less than five, precluding the possibility of conducting a chi-square analysis, but the fact that these comments were only recorded in reference to the female author is note-worthy. Doubt-raising statements regarding the female author were:

“Real people in computer science don’t talk...like that.”

“She is trying to appear smart, lacks interest.”

“She has no idea what she is talking about.”

“She’s apparently not interested in this hobby.”

Table 4

Frequency of Computer Author Descriptors by Author Gender

	Author Gender			
	Female	Male	χ^2 (1)	*t* (111)
Intelligent	26	28	0.07	-
Major or Career in STEM	17	24	1.20	-
Nerdy or Geeky	5	6	0.09	-
Hard Worker, Motivated	13	9	0.73	-
Poor English / Writing	26	12	5.16*	-
Interest in Video Games	4	20	10.67**	-
Low Conscientiousness	21	12	2.46	-
Extraverted	5	4	0.11	-
Introverted	18	15	0.27	-
Physical Appearance	6	14	3.20†	-
Doubt-raisers	4	0	-	-
Counts [*M (SD)*]:				
Total Word Count	17.82 (12.54)	18.51 (12.35)	-	0.43
Number of Descriptors	2.87 (1.43)	3.11 (1.36)	-	0.39

Note. **p* < .05; ***p* < .01; †*p* < .10.

Effects of Implicit and Explicit Attitudes on Essay Grades

Implicit gender bias (i.e., IAT) was expected to correlate significantly with computer grades, but explicit attitudes were not expected to correlate significantly with the IAT or with computer grades. Exercise essay grades were not expected to correlate significantly with either the IAT or explicit attitudes. Table 5 shows Pearson's correlations among IAT scores, explicit measures, and both computer and exercise essay grades. The IAT correlated significantly with computer grades, but not with any of the explicit measures, all *p*s > .31. None of the explicit scales correlated with computer grades, all *p*s > .17. The IAT was not correlated with exercise grades, but the semantic differential scale correlated with exercise grades, such that more favorable attitudes toward women were associated with better grades. None of the remaining explicit measures correlated with exercise grades, all *p*s > .35.

To test differences in strength of relationships among the IAT, explicit measures, and essay grades, the correlation coefficients were converted into *z*-scores and compared using a Steiger's *z*-test (Steiger, 1980). For the computer essay, the correlation between the IAT and computer grades was significantly different from the correlation between the Modern Sexism Scale and computer grades, $z(207) = 2.73$, $p < .001$. The correlation between IAT and computer grades was also significantly different from the correlation between the semantic differential scale and computer grades, $z(206) = 1.47$, $p = .01$. There were no significant correlation differences between the IAT and any other explicit measure, or between computer grades and any other explicit measures, all *p*s > .23.

Table 5

Correlation Matrix for Key Study Variables

	1	**2**	**3**	**4**	**5**	**6**	**7**
1. Computer Essay Grades							
2. Exercise Essay Grades	.38**						
3. Gender-Science IAT	-.16*	.06					
4. Modern Sexism Scale	.07	.06	.06				
5. Women in STEM Stereotypes	.04	-.05	.03	.38**			
6. Semantic Differential	.02	.15*	.04	-.22**	-.11		
7. Feeling Thermometer-Women	.05	.07	-.03	-.03	-.15*	.21*	
8. Feeling Thermometer-Men	.10	.03	-.08	.01	-.06	.08	.74**

Note. Gender-Science IAT: higher scores denote greater association of men with science and women with liberal arts; Modern Sexism Scale: higher score = more sexist; Women in STEM Stereotype Scale: higher score = more strongly endorses stereotypes; Semantic Differential Scale (average of all semantic differential scale items): higher score = more favorable toward women; Feeling Thermometer-Women: higher score = more favorable toward women; Feeling Thermometer-Men: higher score = more favorable toward men.

**p* < .05; ***p* < .01.

The IAT was expected to explain a significant amount of incremental variance in computer grades beyond that of explicit scores. Hierarchical regression analysis revealed that the explicit scales did not explain a significant amount of variance in computer essay grades, *p*s > .14 (step 1; see Table 6a). When the IAT was entered in step 2, the model was significant, and the IAT explained a significant amount of incremental variance in computer essay grades, above the explicit scales, $\Delta R^2 = .02$, $\Delta F(1, 201) = 4.19$, $p < .05$, as predicted.

Table 6a

Multiple Regression Analyses of Explicit Attitude and Implicit Bias Predicting Computer Grades

	β	*t*	R^2	ΔR^2	ΔF
Step 1			0.02	0.02	0.61
Modern Sexism Scale	0.06	0.88			
Women in STEM Stereotype Scale	0.03	0.33			
Semantic Differential	0.04	0.52			
Feeling Thermometer - Women	-0.07	-0.62			
Feeling Thermometer - Men	0.12	1.17			
Step 2			0.04*	0.02	4.19
IAT Score	-0.14*	-2.00			

Note. R^2 value is cumulative for all variables entered in each step. ΔR^2 represents incremental variance explained by IAT over and above all other variables added in step 1. Beta values are final standardized regression coefficients from the full model. *$p < .05$.

For exercise grades, only the β for the semantic differential was significant, β. = 0.15, *t*(201) = 2.09, *p* = .04, (see Table 6b). As favorability toward women increased, exercise grades also increased. Apart from this unexpected finding, the results for the exercise grades partially supported the prediction. Neither the IAT, nor any of the other explicit measures explained a significant amount of variance in exercise grades.

Table 6b

Multiple Regression Analyses of Explicit Attitude and Implicit Bias Predicting Exercise Grades

	β	*t*	R^2	ΔR^2	ΔF
Step 1			0.04	0.04	1.52
Modern Sexism Scale	0.11†	1.48			
Women in STEM Stereotype Scale	-0.07	-0.96			
Semantic Differential	0.15*	2.09			
Feeling Thermometer - Women	0.07	0.64			
Feeling Thermometer - Men	-0.06	-0.61			
Step 2			0.04	0	0.74
IAT Score	0.06	0.89			

Note. R^2 value is cumulative for all variables entered in each step. ΔR^2 represents incremental variance explained by IAT over and above all other variables added in step 1. Beta values are final standardized regression coefficients from the full model. **p* < .05; †*p* < .10.

Effect of Author Gender on Essay Grades

It was expected that author gender would significantly affect essay grades, depending on author gender-topic condition. Computer essay grades for the male author (stereotype-consistent) were expected to be higher than computer essay grades for the female author (stereotype-inconsistent). Computer essay grades in the anonymous condition were expected to be greater than or equal to grades assigned to the male author. Exercise essay grades for females (stereotype-consistent) were expected to be higher than exercise essay grades for males (stereotype-inconsistent). The anonymous exercise essay was expected to receive grades that were greater than or equal to those given to the female author.

To test these predictions, a one-way Multivariate Analysis of Variance (MANOVA) was conducted, with gender-topic (anonymous, stereotype inconsistent, stereotype consistent) as the independent variable and computer grades and essay grades were the dependent variables. Stereotype-consistent (male computer, female exercise) gender-topic conditions and the no author name essays received grades that were on average 4% higher than stereotype-inconsistent (female computer, male exercise) gender-topic conditions. However, the result of the MANOVA was not significant, Wilks' λ = 0.98, *F*(4, 424) = 1.06, *p* = .38 (see Table 7a).

Table 7a

MANOVA of Computer and Exercise Essay Grades (%) by Gender-Topic Condition

	Anonymous	Stereotype Consistent	Stereotype Inconsistent		
	(*n* = 70)	(*n* = 75)	(*n* = 71)		
Essay	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)	*F*(2, 213)	*p*
Computer	75.15 (13.60)$_a$	74.30 (15.25)$_a$	70.80 (15.25)$_b$	1.40	0.25
Exercise	84.15 (12.25)$_a$	84.45 (11.10)$_a$	81.20 (11.95)$_b$	1.48	0.23

Note. Stereotype-consistent condition: male computer, female exercise. Stereotype-inconsistent condition: female computer, male exercise. Anonymous condition: no author names for computer and exercise.

Means with differing subscripts within rows are marginally different, *p* < .10.

As shown in Table 7b, planned contrasts revealed that, for both computer and exercise essays, the anonymous authors did not differ significantly from the stereotype-consistent authors (anonymous computer = male computer; anonymous exercise = female exercise), but the anonymous authors received marginally higher grades than the stereotype-inconsistent authors (anonymous computer > female computer; anonymous exercise > male exercise). For both essays, the combined weighted mean of the anonymous authors and the stereotype-consistent authors was significantly higher than the stereotype-inconsistent authors (anonymous computer & male computer > female computer; anonymous exercise & female exercise > male exercise).

Table 7b

Planned Contrast Results for Computer and Exercise Essay Grades by Gender-Topic Condition

Contrast	Ψ	*t*(213)	*p*
Computer			
Anonymous – Male	0.31	0.12	.45
Anonymous – Female	3.72†	1.50	.07
Female – Male	-3.41†	-1.38	.08
Anonymous & Male – Female	7.13*	1.67	<.05
Exercise			
Anonymous – Female	-0.63	-0.31	.38
Anonymous – Male	2.60	1.30	.10
Female – Male	3.23†	1.62	.05
Anonymous & Female – Male	5.53*	1.69	<.05

Note. Ψ = mean weighted difference.

**p* < .05; †*p* < .10.

Effects of Implicit Bias on Grades, by Author Gender

Participants with high implicit bias were expected to grade the female computer essay more harshly than the male computer essay. To test this hypothesis, a 3 (gender-topic: anonymous, stereotype inconsistent, stereotype consistent) x 2 (implicit gender-science bias: low, high) MANOVA was computed, with computer grades and exercise grades as the dependent variables.

IAT scores were split at the mean, creating two groups, designated high bias or low bias. Mean scores for the variables were slightly above the mid-point, so conducting a mean split (rather than a median split) ensured that participants who did not have strongly biased attitudes and associations (neither for nor against females) were included in the low-bias group. The high bias group included participants whose scores indicated a stronger association of men with science than women with science. The low bias group included participants whose scores indicated a stronger association of women with science, and those whose implicit associations did not reflect a stronger association one way or the other (i.e. neutral). Table 8a shows the results of the MANOVA. There was no main effect of gender-topic condition for either computer or exercise grades, but there was a significant main effect of implicit bias on computer grades. Compared to participants in the low bias group, participants in the high bias group assigned significantly lower grades to the anonymous computer essay, and marginally lower grades to the female author computer essay (see Figure 1). Exercise grades did not differ by implicit bias, and the interaction term was not significant for computer or exercise grades.

Table 8a

MANOVA of Computer and Exercise Grades (%) in Each Gender-Topic Condition, by Low or High Implicit Gender-Science Bias

	Low Bias	High Bias	Gender-Topic			Implicit Bias			Gender-Topic x Implicit Bias		
	M (*SD*)	*M* (*SD*)	λ	*F*	*df*	λ	*F*	*df*	λ	*F*	*df*
			0.98	1.17	4, 408	0.96	3.98*	2, 204	0.98	1.18	4, 408
Computer Grades				1.84	2, 205		3.89*	1, 205		1.15	2, 205
Anonymous	78.08 (12.03)$_a$	71.84 (14.69)$_b$									
Stereotype Inconsistent (female author)	73.55 (14.03)†	67.39 (16.16)†									
Stereotype Consistent (male author)	73.78 (14.07)	74.17 (17.18)									
Exercise Grades				1.35	2, 205		1.32	1, 205		0.40	2, 205
Anonymous	82.64 (12.11)	85.84 (12.38)									
Stereotype Consistent (male author)	84.21 (11.93)	84.03 (10.36)									
Stereotype Inconsistent (female author)	80.06 (13.19)	82.67 (10.22)									

Note. Means with differing subscripts within the same row are significantly different, $p < .05$. Means marked with † within the same row are marginally different, $p < .10$.
*$p < .05$.

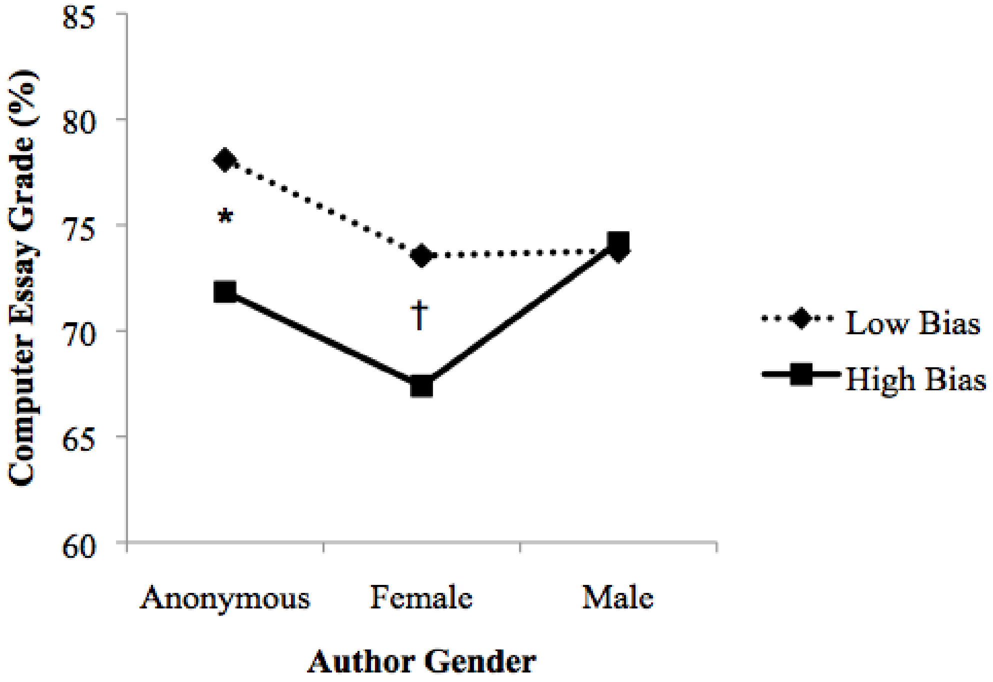


Figure 1. Computer essay grades by gender-topic and implicit bias group.

$^{\dagger}p < .10$; $*p < .05$.

As shown in Table 8b, planned contrasts revealed that participants in the low bias group graded the anonymous computer essay significantly higher than the male computer author and the female computer author, respectively. Participants in the low implicit bias group did not grade the male author or female author differently. In contrast, participants in the high implicit bias group graded the female computer author significantly lower than the male computer author. High implicit bias participants also rated the anonymous computer author higher than the female author, but there was no significant difference between the anonymous and male computer authors in this group.

Table 8b

Planned Contrasts for Computer Essay Grades by Gender-Topic Condition and Implicit Bias Group

Contrast (Computer Grades)	Ψ	*t*	*df*	*p*
Low Implicit Bias				
Anonymous – Male	4.30	2.50*	113	.01
Anonymous – Female	4.53	2.67**	113	<.01
Female – Male	-0.23	-0.14	113	.89
High Implicit Bias				
Anonymous – Male	-2.33	-1.23	92	.22
Anonymous – Female	4.45	2.41*	92	.02
Female – Male	-6.78	-3.61***	92	<.001

Note. Ψ = mean weighted difference.

**p* < .05; ***p* < .01; ****p* < .001.

As shown in Table 8c, planned contrasts revealed the participants in the low bias group graded the anonymous exercise author no differently than the male or female exercise authors. Participants with low implicit bias graded the female exercise author significantly higher than the male exercise author. Participants in the high implicit bias group graded the anonymous exercise author marginally higher than the male exercise author, but high implicit bias participants did not grade the anonymous exercise author or the male author differently than the female author.

Table 8c

Planned Contrasts for Exercise Essay Grades by Gender-Topic Condition and Implicit Bias Group

Contrast (Exercise Grades)	Ψ	*t*	*df*	*p*
Low Implicit Bias				
Anonymous – Male	2.58	1.52	113	.13
Anonymous – Female	-1.57	-0.91	113	.36
Female – Male	4.15	2.49*	113	.01
High Implicit Bias				
Anonymous – Male	3.17	1.72†	92	.09
Anonymous – Female	1.81	0.96	92	.34
Female – Male	1.36	0.72	92	.47

Note. Ψ = mean weighted difference.

†$p < .10$; *$p < .05$.

Effects of Implicit and Explicit Attitudes on Grades, by Author Gender

Participants with the combination of high implicit bias and low explicit attitudes were expected to grade the computer essay more harshly than those with low implicit bias and low explicit attitudes. Differences by implicit bias and explicit attitudes in the exercise essay were expected to be negligible. To create explicit sexism groups and stereotyping groups, respectively, mean dichotomous splits were conducted on the Modern Sexism Scale and the Women in STEM Stereotype scale. As with the IAT, mean split ensured that participants who did not have strongly biased attitudes and associations (neither for nor against females) were included in the low bias group. Two MANOVAs were computed to examine the interaction of implicit bias and the two explicit variables, with computer grades and exercise grades entered as the dependent variables. Pairwise comparisons were conducted using the Bonferrroni correction.

Figure 2 shows the effect of implicit bias and explicit sexism (Modern Sexism Scale) on computer grades. The main effects for implicit bias (*F*1, 206) = 1.27, *p* = .46) and sexism (*F*(1, 206) = 0.00, *p* = .99) were both non-significant. The interaction effect was marginally significant, *F*(1, 206) = 2.93, *p* = .09. Pairwise comparisons revealed that computer grades assigned by participants in the high sexism group differed significantly by level of implicit bias, *F*(1, 206) = 6.19, *p* = .01. Participants high in sexism and implicit bias (*M* = 69.51, *SD* = 18.35), graded the computer essay significantly lower than participants who were high in sexism and low in implicit bias (*M* = 76.89, *SD* = 13.28). Computer grades assigned by participants in the low sexism group did not differ by implicit bias group (*M* = 73.36, *SD* = 13.59 and *M* = 76.89, *SD* = 13.28, respectively), *F*(1, 206) = 0.03, *p* = .87.

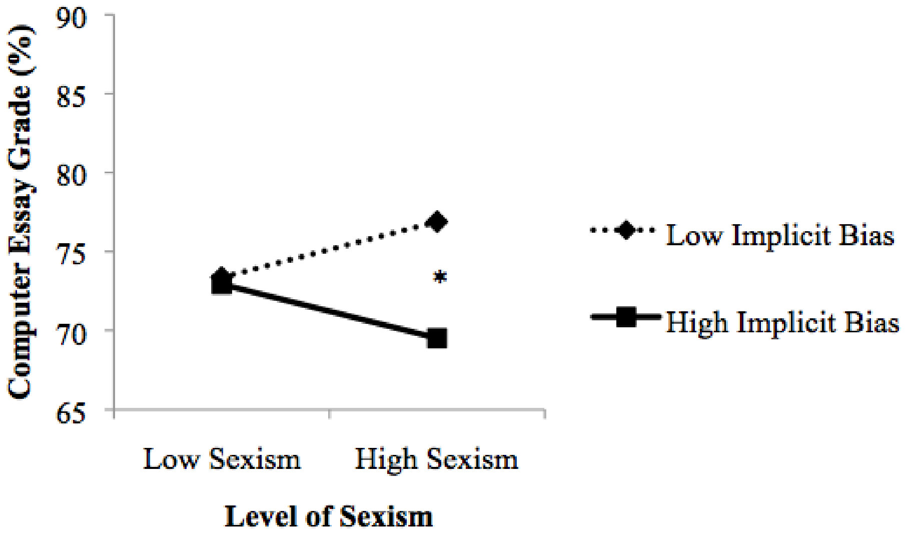


Figure 2. Computer essay grades by implicit bias & explicit sexism (Modern Sexism Scale). *$p < .05$.

Next, the effect of implicit bias and explicit sexism (Modern Sexism Scale) on exercise grades was examined (see Figure 3). The main effects for implicit bias (*F*1, 206) = 0.12, *p* = .79) and sexism (*F*(1, 206) = 0.05, *p* = .86) were both non-significant. However, the interaction was significant, *F*(1, 206) = 8.81, *p* < .01. Pairwise comparisons revealed that exercise grades assigned by participants who were high in sexism did not differ by implicit bias level (*M* = 85.36, *SD* = 11.31 and *M* = 82.21, *SD* = 13.06, respectively), *F*(1, 206) = 1.78, *p* = .18. In contrast, exercise grades assigned by participants who were low in sexism differed significantly by implicit bias level, *F*(1, 206) = 8.55, *p* < .01. Those who were low in sexism and high in implicit bias graded the exercise essay significantly higher (*M* = 85.90, *SD* = 8.79) than those who were low in sexism and implicit bias (*M* = 79.45, *SD* = 12.87).

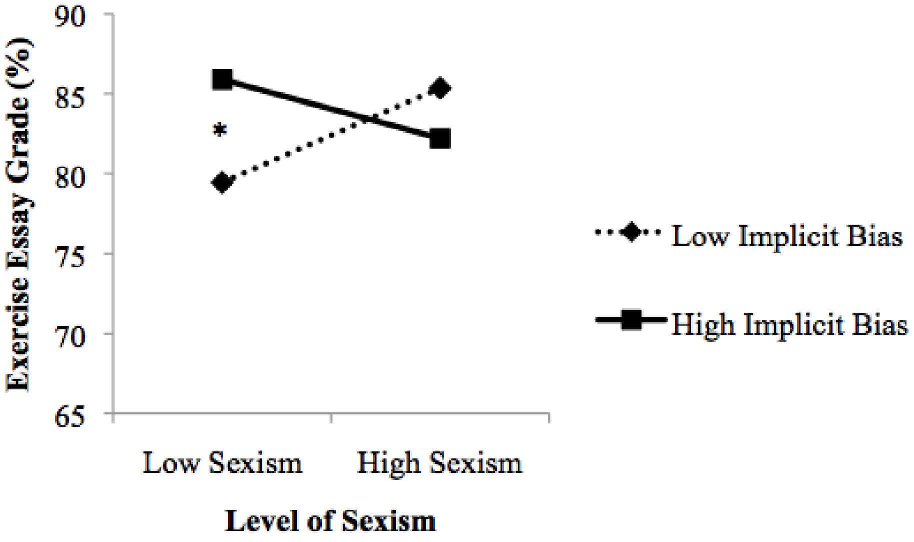


Figure 3. Exercise essay grades by implicit bias & explicit sexism (Modern Sexism Scale). **p* < .05.

Figure 4 shows the effect of implicit bias and explicit prejudice (Women in STEM Stereotype Scale) on computer grades. The main effects for implicit bias (*F*(1, 205) = 1.50, *p* = .44) and prejudice (*F*(1, 205) = 0.00, *p* = .99) were not significant, and the interaction was also not significant, *F*(1, 205) = 1.84, *p* = .18. Pairwise comparisons revealed that computer grades assigned by participants who were high in prejudice did not differ by implicit bias level (*M* = 72.83, *SD* = 16.83 and *M* = 73.44, *SD* = 13.68, respectively), *F*(1, 205) = 0.04, *p* = .83. However, computer grades assigned by participants low in prejudice differed significantly by implicit bias level, *F*(1, 205) = 4.90, *p* = .03. Those who were low in prejudice and high in implicit bias graded the computer essay significantly lower (*M* = 70.08, *SD* = 15.00) than participants who were low in prejudice and implicit bias (*M* = 76.23, *SD* = 13.37).

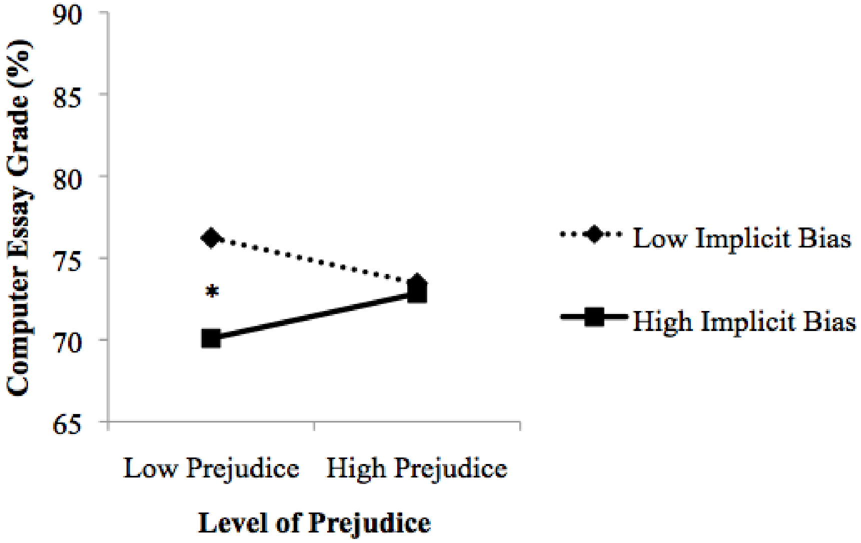


Figure 4. Computer essay grades by implicit bias and explicit prejudice (Women in STEM Stereotype Scale).

**p* < .05.

Next, the effect of implicit bias and explicit prejudice (Women in STEM Stereotype Scale) on exercise grades was examined (see Figure 5). The main effects for implicit bias [*F*(1, 205) = 0.89, *p* = .52] and prejudice [*F*(1, 205) = 0.11, *p* = .80] were not significant, and the interaction was also not significant, *F*(1, 205) = 1.34, *p* = .25. Pairwise comparisons revealed that exercise grades assigned by participants who were high in prejudice did not differ by implicit bias level (*M* = 82.86, *SD* = 11.01 and *M* = 82.97, *SD* = 11.71, respectively), *F*(1, 205) = 0.00, *p* = .97. Exercise grades assigned by participants who were low in prejudice differed marginally by implicit bias level, *F*(1, 205) = 2.71, *p* = .10. Those who were low in prejudice and high in implicit bias graded the exercise essay marginally higher (*M* = 85.40 *SD* = 11.05) than those who were low in prejudice and implicit bias (*M* = 81.68, *SD* = 13.23).

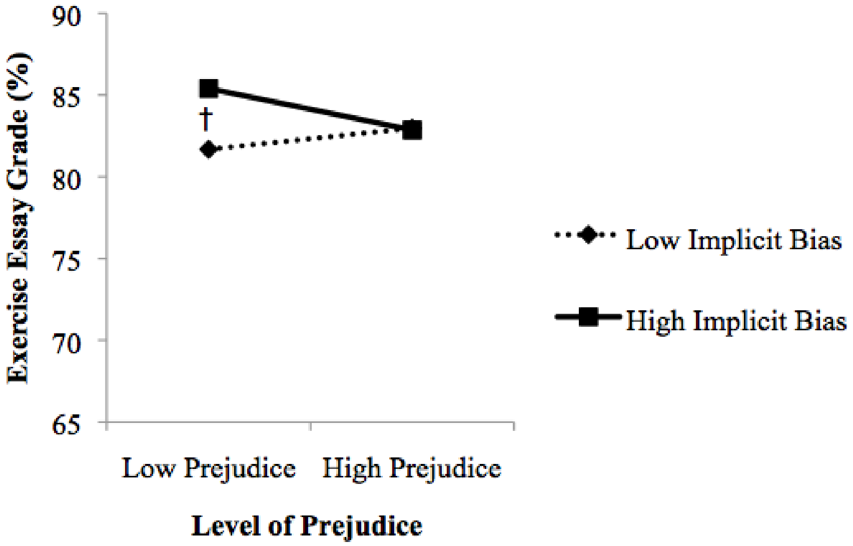


Figure 5. Exercise essay grades by implicit bias and explicit prejudice (Women in STEM Stereotype Scale).

$^{\dagger}p < .10$.

Next, the interaction of implicit bias and explicit attitudes on computer grades was compared by author gender. Because exercise grades did not differ significantly by implicit and explicit bias, only computer essay grades were examined at this level of analysis. Participants who were high in implicit bias and low in explicit bias were expected to grade the female author of the computer essay more harshly than the anonymous and male computer authors. Within female author grades, participants who were high in both implicit and explicit bias were expected to assign lower grades compared to participants who were low in both implicit and explicit bias. No group differences were expected in anonymous or male author grades. Two ANOVAs were computed to examine the interactions among the mean split implicit and explicit variables, by gender-topic condition, with computer grade entered as the dependent variable. Pairwise comparisons were conducted using the Bonferrroni correction.

The first ANOVA examined the effects of implicit bias and explicit sexism (Modern Sexism Scale) on computer grades by gender-topic condition. The individual main effects for gender-essay pair, implicit bias, and sexism were not significant, all *p*s > .39. The interaction term of implicit bias, explicit sexism, and gender-topic was not significant, *F*(2, 198) = 0.09, *p* = .91. Pairwise comparisons revealed that participants low in implicit bias graded the anonymous and female computer authors significantly higher than those who were high in implicit bias. Further, as shown in Table 9, computer grades assigned to the female author by participants who were high in sexism differed by implicit bias level, *F*(1, 198) = 4.23, *p* = .04. When participants were high in both sexism and implicit bias, they graded the female computer author marginally lower and the anonymous author significantly lower, compared to participants who were high in sexism and low in implicit bias. There were no significant differences in the male author condition.

Table 9

Effect of Implicit Bias and Explicit Sexism (Modern Sexism Scale) on Computer Essay Grades (%) by Author Gender

	Low Implicit Bias		High Implicit Bias	
	Low Sexism	High Sexism	Low Sexism	High Sexism
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Anonymous	75.91 (10.42)	80.05 (12.82)$_a$	72.90 (16.39)	67.86 (15.03)$_b$
Female Author	70.54 (14.62)	76.90 (12.27)†	67.14 (13.11)	67.00 (19.28)†
Male Author	74.17 (12.84)	71.76 (16.86)	76.88 (12.25)	76.25 (19.96)

Note. Means with differing subscripts within the same row are significantly different, $p < .05$. Means marked with † within the same row are marginally different, $p < .10$.

Figures 6a, 6b, and 6c display the interactions of implicit bias and explicit sexism on computer grades by author gender. Computer grades assigned to the anonymous author by participants who were high in both sexism and implicit bias were significantly lower than grades assigned by those who were high in sexism and low in implicit bias (Figure 6a). Grades assigned by those who were low in sexism did not differ by implicit bias level.

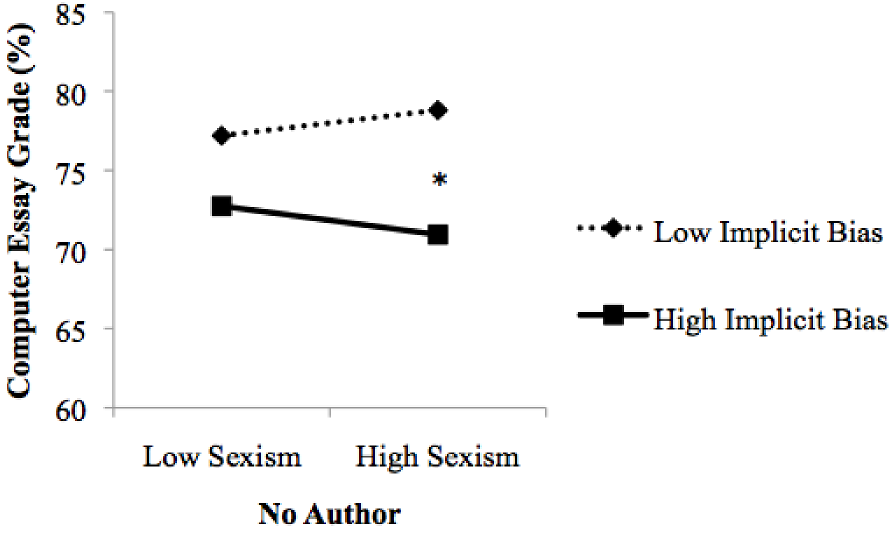


Figure 6a. Computer essay grades by implicit bias and explicit sexism in the anonymous condition.

*p < .05.

Computer grades assigned to the female author by participants who were high in both sexism and implicit bias were significantly lower than grades assigned by those who were high in sexism and low in implicit bias (Figure 6b). Grades assigned by those who were low in sexism did not differ by implicit bias level. Male computer grades did not differ significantly by implicit bias level or sexism level (see Figure 6c).

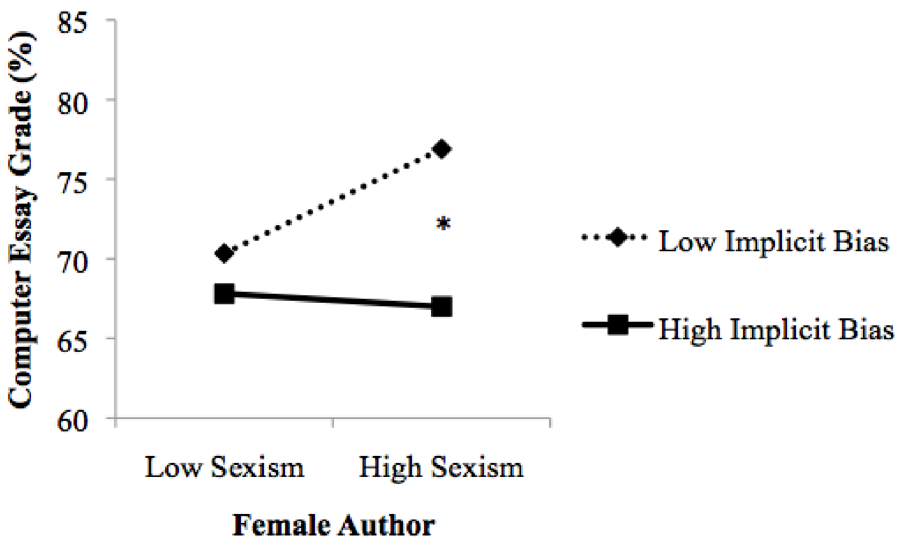


Figure 6b. Computer essay grades by implicit bias and explicit sexism in the female author condition.

*p < .05.

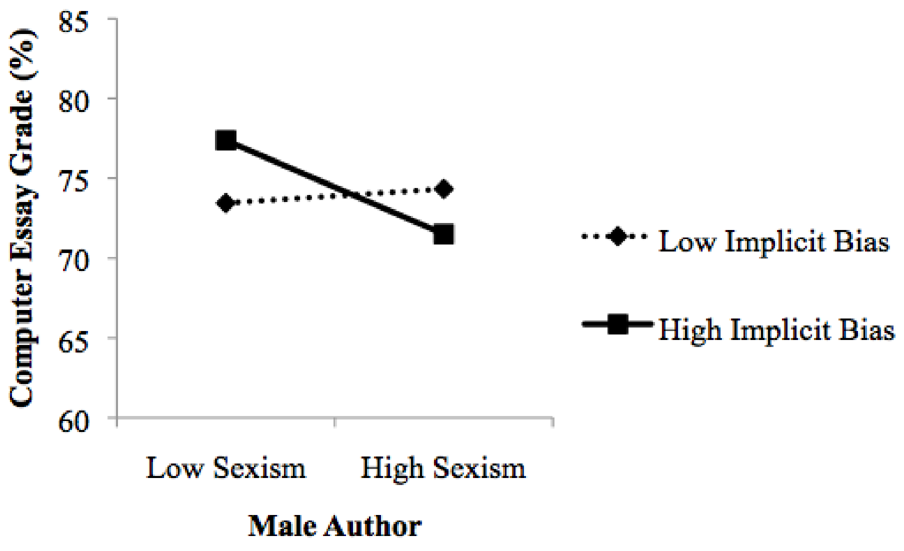


Figure 6c. Computer essay grades by implicit bias and explicit sexism in the male author condition.

In summary, the interactions of implicit bias and explicit attitudes resulted in differences in grading, depending on whether the author gender and essay topic pairing was stereotype-consistent (e.g. male computer) or stereotype-inconsistent (e.g. female computer). For the computer essays, low implicit gender-science bias appeared to result in higher grades across most conditions. For exercise essays, low implicit gender-science bias did not affect grades systematically. Implicit bias and sexism interacted marginally, but in an unexpected pattern. Grades in the low sexism group did not differ by level of implicit bias, but grades in the high sexism group were different depending on level of implicit bias.

The second ANOVA examined the effect of implicit bias and explicit prejudice (Women in STEM Stereotype Scale) on computer grades by author gender. None of the main effects were significant, all *p*s > .43. The interactions of implicit bias by explicit prejudice, *F*(12, 197) = 8.71, *p* = .10, and author gender by implicit bias, *F*(2, 197) = 7.09, *p* = .12, appeared to be trending toward significance. The 3-way interaction of implicit bias, explicit prejudice, and author gender was not significant, *F*(2, 197) = 0.23, *p* = .80.

Table 10 displays the results of pairwise comparisons, which revealed that participants who were low in prejudice but high in implicit bias graded the female computer author significantly lower than the male computer author, *F*(1, 197) = 4.54, *p* = .03. Participants who were low in prejudice and high in implicit bias graded the anonymous computer essay significantly lower than those who were low in both prejudice and implicit bias, *F*(1, 197) = 4.27, *p* = .04. Within participants who were high in prejudice, there were no differences implicit bias level for any author gender, *F*(1, 197) = 0.01, *p* = .94.

Table 10

Effect of Implicit Bias and Explicit Prejudice (Women in STEM Stereotype Scale) on Computer Essay Grades (%) by Author Gender

	Low Implicit Bias		High Implicit Bias	
	Low Prejudice	High Prejudice	Low Prejudice	High Prejudice
Anonymous	78.80 (11.35)$_{a}$	77.29 (12.25)	69.00 (16.15)$_{bc}$	71.88 (14.87)
Female Author	74.29 (12.06)$_{ac}$	71.81 (16.17)	64.06 (13.93)$_{b}$	70.27 (18.85)
Male Author	75.94 (14.98)	70.00 (13.28)	76.67 (15.42)	76.67 (17.11)

Note. Means with differing subscripts are significantly different, *p* < .05.

Figures 7a, 7b, and 7c display the interactions of implicit bias and explicit prejudice on computer grades by author gender. Computer grades assigned to the anonymous author by participants who were high in prejudice did not differ by implicit bias level (Figure 7a). In contrast, participants who were low in prejudice assigned significantly lower grades when they were high in implicit bias.

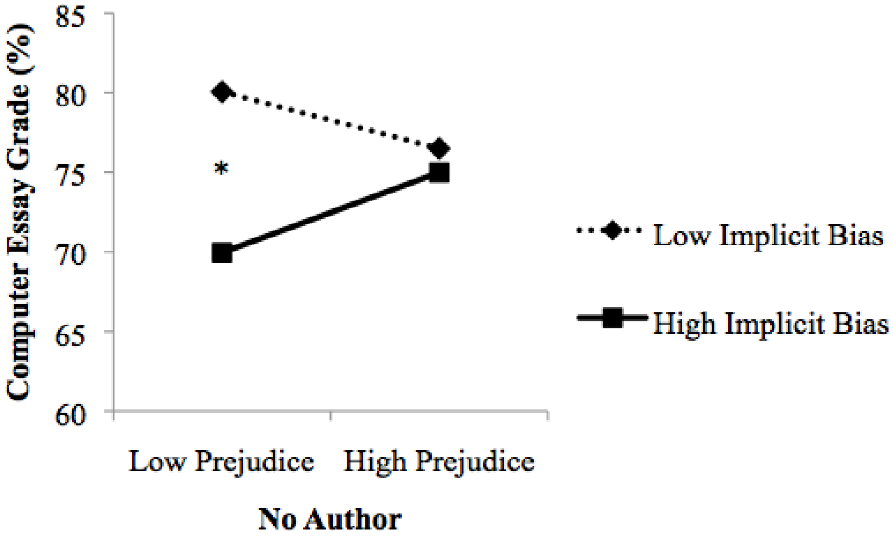


Figure 7a. Computer essay grades by implicit bias and explicit prejudice (Women in STEM Stereotype Scale) in the anonymous condition.

* *p* < .05.

Computer grades assigned to the female author by participants who were low in prejudice and high in implicit bias were significantly lower than grades assigned by those who were low in both prejudice and implicit bias (Figure 7b). Grades assigned by those who were high in prejudice did not differ by implicit bias level. Male computer grades did not differ significantly by implicit bias level or prejudice level (see Figure 7c).

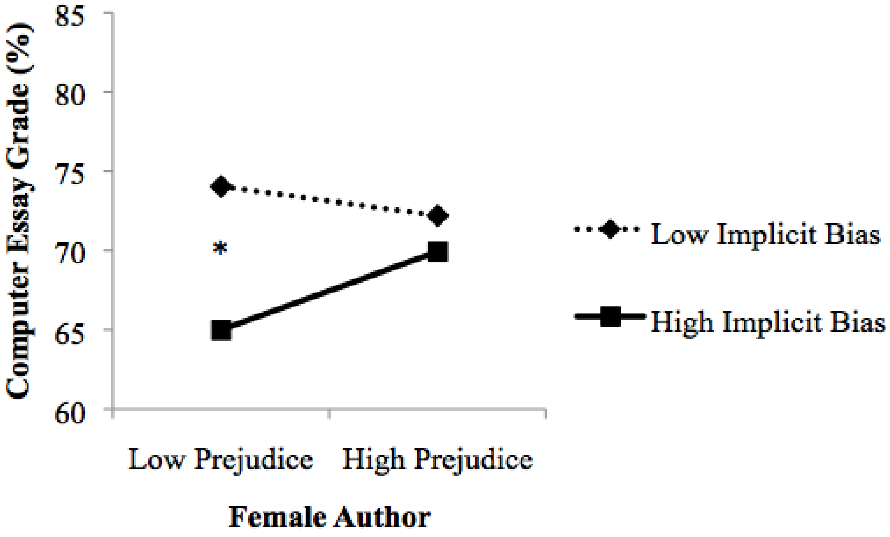


Figure 7b. Computer essay grades by implicit bias and explicit prejudice (Women in STEM Stereotype Scale) in the female author condition.

**p* < .05.

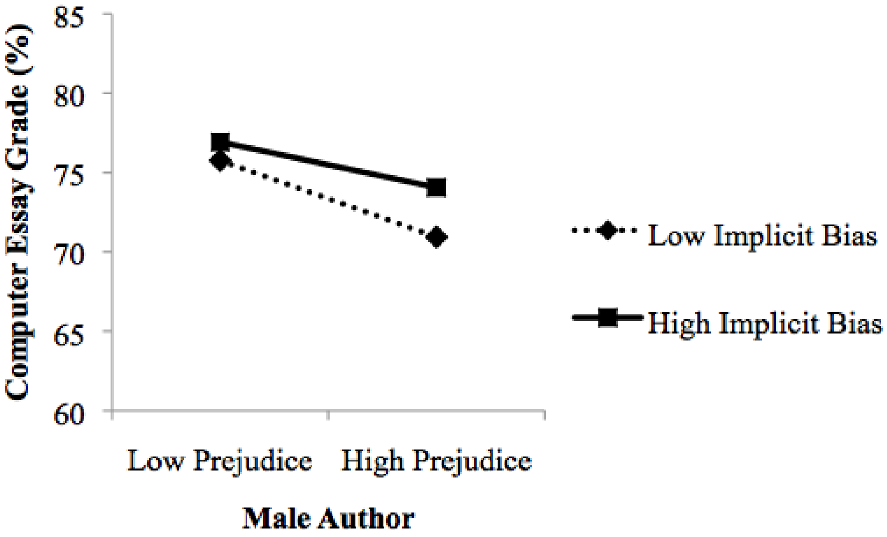


Figure 7c. Computer essay grades by implicit bias and explicit prejudice (Women in STEM Stereotype Scale) in the male author condition.

In summary, the interaction of implicit bias with explicit prejudice on computer grades followed the prediction that low explicit prejudice and high implicit bias would result in significantly lower grades, and this pattern was observed on grades for both the anonymous computer author and the female author. Grades for the male computer author did not differ depending implicit or explicit bias levels.

Effectiveness of Rubrics to Increase Consistency in Grading

Rubric use was expected to result in greater consistency and less variability in essay grades, compared to grades assigned in the no-rubric condition. To compare the variance of the rubric and no rubric conditions, ranges for computer and exercise grades were examined. Then, Levene's test for equality of variance was computed.

The percent grade range for the computer essay in the rubric condition (range = 35 – 100) appeared to be smaller than the grade range for the computer essay when no rubric was used (range = 45 – 100). The percent grade range for the exercise essay in the rubric condition (range = 55 – 100) was not different from the range in the no rubric condition (range = 55 – 100). The variance of the computer essay grades in the no rubric condition appeared to be greater (s^2 = 176.78) than that for the rubric condition (s^2 = 125.19), but Levene's test for equality of variance revealed no significant difference in variance between conditions (*F* (107, 106) = 1.19, *p* = .28). Similarly, there was no significant difference in exercise grade variance between the rubric conditions, (*F*(106, 107) = 0.09, *p* = .76).

To further examine consistency, correlations were examined. The correlation between computer grades for the rubric and no-rubric groups was not significant, *r* = -.04, *p* = .67. The correlation between exercise grades for the rubric and no-rubric groups was not significant, *r* = .05, *p* = .58. These findings suggest that the rubric did not increase consistency for either essay. A Fisher's *z*-test revealed no significant difference between these correlations, *z* = –0.93, *p* = .18, which indicates that the rubric was equally ineffective at increasing consistency for both essays.

Effectiveness of Rubrics to Decrease Bias Impact on Grading

Rubric use was expected to reduce the effect of implicit bias on essay grading. To test this hypothesis, hierarchical regression analyses were conducted. As shown in Table 11, IAT score was entered in step 1, rubric was entered in step 2, and the interaction term was entered in Step 3. In step 1, the IAT significantly predicted 2.4% of the variance in computer essay grades. As implicit bias level decreased, computer grades increased. Adding rubric condition in step 2 of the model explained an additional 8% of incremental variance in computer essay grades. Surprisingly, rubric use resulted in lower grades compared to no rubric condition. The interaction term of implicit bias and rubric condition did not explain any significant variance in computer grades. In step 3, the IAT became non-significant, leaving only the significant effect of rubric in the final model.

Table 11

Hierarchical Regression of Computer Grades by Implicit Bias and Rubric Condition

	β	*t*	R^2	ΔR^2	*F*	ΔF	*df*
Step 1			0.02*		5.15		1, 209
Gender-Science IAT	-0.16*	-2.27					
Step 2			0.11***	0.08***	12.16	18.73	1, 208
Gender-Science IAT	-0.15*	-2.23					
Rubric Condition	-0.28***	-4.33					
Step 3			0.11***	0.00	8.10	0.09	1, 207
Gender-Science IAT	-0.13	-1.39					
Rubric Condition	-0.28***	-4.32					
IAT x Rubric	-0.03	-0.30					

Note. R^2 value is cumulative for all variables entered in each step.

* $p < .05$; *** $p < .001$.

To evaluate the effect of implicit bias and rubric condition on computer grades, three hierarchical regression analyses were conducted, one for each author gender. As shown in Table 12, rubric was entered in Step 1, the IAT was entered in step 2, and the interaction term was entered in step 3. Because none of the steps in the hierarchical regression model explained any significant variance in exercise grades, only analyses examining these effects in computer grades were conducted by author gender.

For all 3 author conditions, the rubric entered in step 1 explained a significant amount of variance in computer grades, with rubric use resulting in lower grades. The addition of the IAT in step 2 did not explain a significant amount of variance in either the anonymous or the male author conditions, but the IAT did explain a marginal amount of incremental variance in grades assigned to the female author. Participants who were high in implicit bias assigned marginally lower grades to the female author, compared to participants who were low in implicit bias. The interaction term of implicit bias and rubric condition did not explain any significant variance in computer grades for any author gender condition.

Table 12

Hierarchical Regression Analyses of Computer Grades by Implicit Bias and Rubric, Split by Author Gender

	β	t	R^2	ΔR^2	F	ΔF	df
Anonymous Author							
Step 1			.14**		10.47		1, 66
Rubric	-0.37**	-3.24					
Step 2			.16**	.02	5.99	1.44	1, 65
Rubric	-0.33**	-2.75					
Implicit Bias	-0.14	-1.20					
Step 3			.16**	< .01	4.01	0.20	1, 64
Rubric	-0.33**	-2.71					
Implicit Bias	-0.20	-1.12					
Implicit Bias x Rubric	-0.08	-0.44					
Female Author							
Step 1			.07*		5.76		1, 72
Rubric	-0.27*	-2.40					
Step 2			.12*	.04†	4.72	3.48	1, 71
Rubric	-0.28*	-2.53					
Implicit Bias	-0.21†	-1.86					
Step 3			.14*	.02	3.64	1.43	1, 70
Rubric	-0.27*	-2.43					
Implicit Bias	-0.08	-0.49					
Implicit Bias x Rubric	-0.19	-1.20					
Male Author							
Step 1			.06*		4.22		1, 66
Rubric	-0.24*	-2.06					
Step 2			.06	< .01	2.21	1.44	1, 65
Rubric	-0.25*	-2.10					
Implicit Bias	-0.06	-0.50					
Step 3			.06	< .001	1.46	0.20	1, 64
Rubric	-0.25*	-2.05					
Implicit Bias	-0.07	-0.46					
Implicit Bias x Rubric	-0.02	-0.12					

Note. R^2 is cumulative for all variables entered in each step.
† $p < .10$, * $p < .05$, ** $p < .01$.

Rubric use was expected to moderate the relationship between author gender and essay grades. Compared to no rubric use, rubric use was expected to diminish any difference in grades by author gender. To test this hypothesis, a 3 (gender-topic: anonymous, stereotype inconsistent, stereotype consistent) x 2 (rubric, no rubric) MANOVA was computed, with computer grades and exercise grades as the dependent variables.

Table 13a shows the results of the MANOVA. The main effect of gender-topic was not significant. There was a main effect of rubric, with rubric use resulting in lower grades (see Figure 8). Exercise grades did not differ by implicit bias, and the interaction was not significant for either computer or exercise grades. Planned contrasts revealed no significant mean differences between any gender-topic conditions for either computer or exercise grades (see Table 13b).

Table 13a

MANOVA of Computer and Exercise Grades (%) in Each Gender-Topic Condition, by Rubric Condition

	Rubric	No Rubric	Gender-Topic			Rubric			Gender-Topic x Rubric		
	M (*SD*)	*M* (*SD*)	λ	*F*	*df*	λ	*F*	*df*	λ	*F*	*df*
			0.98	1.06	4, 418	0.91	10.24***	2, 209	0.99	0.12	4, 418
Computer				1.43	2, 210		17.09***	1, 210		0.02	2, 210
Anonymous	70.43 (13.90)	79.00 (13.07)									
Stereotype Inconsistent (female author)	67.17 (15.64)	74.92 (13.89)									
Stereotype Consistent (male author)	70.29 (15.62)	78.40 (14.46)									
Exercise				1.47	2, 210		0.01	1, 210		0.23	2, 210
Anonymous	83.14 (11.89)	84.34 (14.21)									
Stereotype Inconsistent (male author)	81.91 (12.43)	80.35 (11.42)									
Stereotype Consistent (female author)	84.43 (11.10)	84.31 (11.44)									

Note. Means with differing subscripts within the same row are significantly different, *p* < .05. Means marked with † within the same row are marginally different, *p* < .10.

**p* < .05.

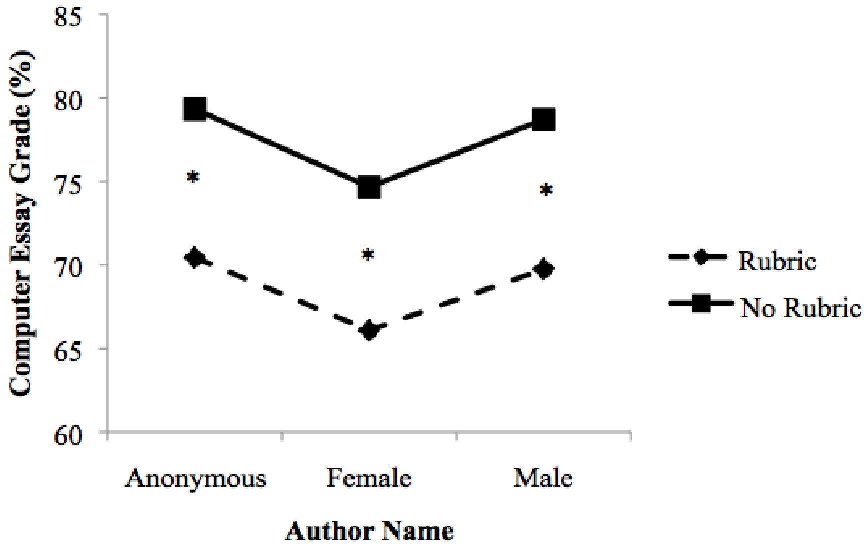


Figure 8. Computer essay grades by gender-topic and rubric condition.

*p < .05.

Table 13b

Planned Contrasts for Computer and Exercise Essay Grades by Gender-Topic and Rubric Condition

Contrast	Ψ	*t*	Contrast	Ψ	*t*
Rubric Computer			No Rubric Computer		
Anonymous – Female	3.26	0.96	Anonymous – Male	4.08	1.20
Anonymous – Male	0.14	0.04	Anonymous – Female	0.60	0.17
Female – Male	-3.12	-0.9	Female – Male	-3.48	-1.03
Rubric Exercise			No Rubric Exercise		
Anonymous – Female	1.23	0.43	Anonymous – Male	3.99	1.40
Anonymous – Male	-1.29	-0.45	Anonymous – Female	0.03	0.01
Female – Male	-2.52	-0.89	Female – Male	-3.96	1.40

Note. Ψ = mean weighted difference.

Next, a 3 (gender-topic condition) x 2 (implicit bias level) x 2 (rubric condition) MANOVA was conducted to examine their combined effect on essay grades. The results of the MANOVA are presented in Table 14a. The main effects of implicit bias and rubric were significant. The main effect of gender-topic was not significant. There were no significant 2-way interactions, and the 3-way interaction was not significant.

Table 14a

MANOVA of Computer and Exercise Grades (%) in Each Gender-Topic Condition, by Implicit Bias & Rubric Condition

	Rubric	No Rubric	Implicit Bias			Rubric			Gender-Topic x Implicit Bias x Rubric		
	M (*SD*)	*M* (*SD*)	λ	*F*	*df*	λ	*F*	*df*	λ	*F*	*df*
			0.96	3.88*	2, 198	0.91	9.74**	2, 198	0.99	0.61	4, 396
Computer Grades				2.82†	1, 199		17.32**	1, 199		0.08	2, 199
Anonymous	70.15 (14.01) $_{a}$	80.15 (11.34) $_{b}$									
Stereotype Inconsistent (female author)	66.69 (15.57)$_{a}$	74.92 (13.89)$_{b}$									
Stereotype Consistent (male author)	70.29 (15.62) $_{a}$	77.72 (14.39) $_{b}$									
Exercise Grades				1.96	1, 199		0.06	1,199		0.91	2, 199
Anonymous	82.79 (11.88)	85.50 (12.64)									
Stereotype Inconsistent (male author)	82.09 (12.55)	80.35 (11.42)									
Stereotype Consistent (female author)	84.43 (11.10)	83.82 (11.45)									

Note. Means with differing subscripts within the same row are significantly different, *p* < .05.

† *p* < .10; **p* < .05; ***p* < .01.

Planned contrasts for computer essay grades are displayed in Table 14b. These analyses revealed, when the rubric was used, participants who were high in implicit bias graded the female computer author significantly lower than the male computer author. There was no significantly difference between the male and female author by high or low implicit bias in the no rubric condition. Additionally, in the rubric condition, participants high in implicit bias graded the female author marginally lower, compared to participants who were low in implicit bias. This difference was not observed in the no rubric condition, and no other mean differences were significant.

Table 14b

Planned Contrasts for Computer Essay Grades by Gender-Topic Condition, Implicit Bias Level, & Rubric Condition

Contrast (Computer Grades)	Ψ	*t*(51)		Ψ	*t*(50)
Rubric - Low Bias			Rubric - High Bias		
Anonymous – Female	3.17	0.67	Anonymous – Female	4.88	1.03
Anonymous – Male	3.53	0.73	Anonymous – Male	-4.22	-0.87
Female – Male	0.36	0.08	Female – Male	-9.12	-1.92*

Contrast (Computer Grades)	Ψ	*t*(59)		Ψ	*t*(42)
No Rubric - Low Bias			No Rubric - High Bias		
Anonymous – Female	7.23	1.64	Anonymous – Female	10.32	1.65†
Anonymous – Male	7.00	1.62	Anonymous – Male	3.54	0.56
Female – Male	-0.23	-0.05	Female – Male	-6.78	-1.47†

Note. Ψ = mean weighted difference.

**p* < .05.

Table 14c shows the marginal effect of implicit bias level in the rubric condition on female computer grades. When the rubric was used, participants who were high in implicit bias graded the female author marginally lower, compared to those who were low in implicit bias. There was no difference by implicit bias level in the no rubric condition. The interaction of implicit bias and rubric condition on computer grades for the male and female authors are displayed in Figures 9a and 9b.

Table14c

Computer Essay Grades (%) by Gender-Topic Condition, Implicit Bias Level, & Rubric Condition

	Low Bias	High Bias			
	M (*SD*)	*M* (*SD*)	*t*	*df*	*p*
Rubric					
Anonymous	73.53 (12.22)	66.76 (15.20)	1.43	32	.16
Female author	70.36 (14.88)	61.88 (15.59)	1.68†	35	<.10
Male author	70.00 (14.65)	71.00 (18.68)	-0.17	33	.87
No Rubric					
Anonymous	80.78 (10.71)	77.71 (14.20)	0.63	32	.53
Female author	76.50 (12.27)	72.60 (16.15)	0.84	35	.41
Male author	77.50 (13.20)	78.00 (16.24)	-0.10	32	.92

Note. † $p < .10$.

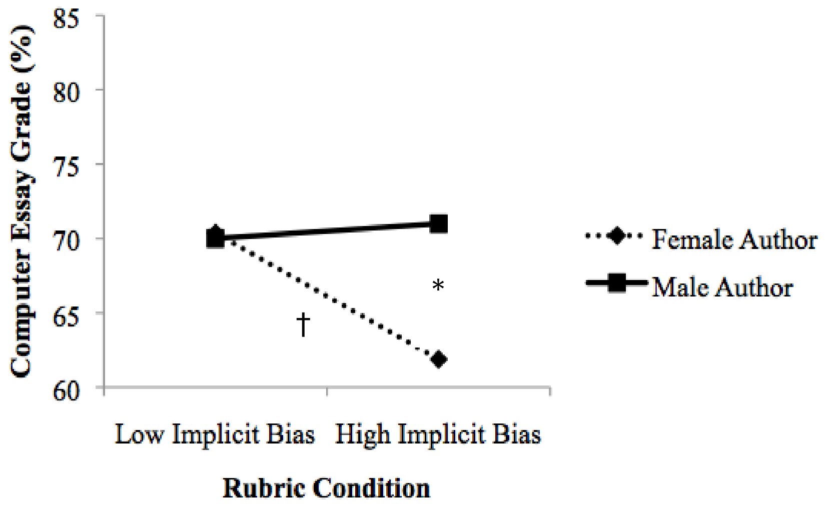


Figure 9a. Computer grades by author gender and implicit bias in the rubric condition.

*$p < .05$; $^{\dagger}p < .10$.

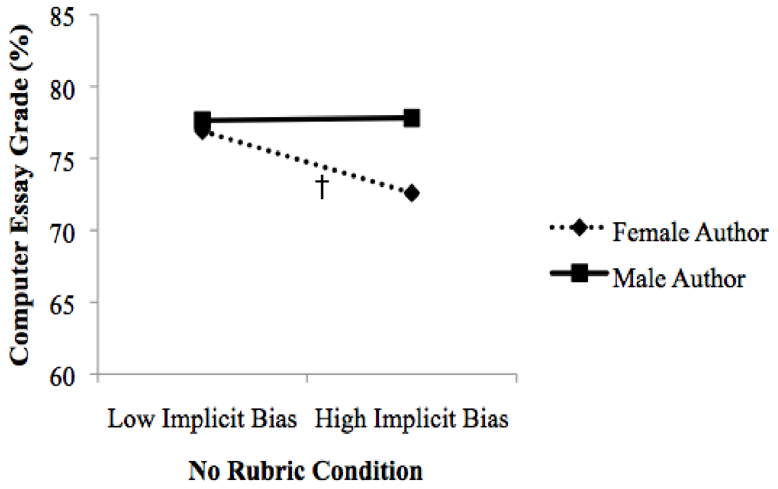


Figure 9b. Computer grades by author gender and implicit bias in the no rubric condition.

$^{\dagger}p < .10$.

In summary, computer grades differed significantly by implicit bias level, such that participants who were low in implicit bias assigned higher computer grades, and by rubric condition, with rubric use resulting in lower computer grades. The main effect of gender-topic was not significant, and the interaction of gender-topic, implicit bias, and rubric was not significant. Planned contrasts revealed that, when the rubric was used, high implicit bias resulted in lower grades for the female author, compared to low implicit bias. There was no difference by bias level for the male author. The differences between author gender and implicit bias in the no rubric condition were not significant.

Gender Differences in Implicit and Explicit Attitude Scores

Male participants were expected to have implicit and explicit bias levels that were more stereotypical and less favorable toward women, than female participants. Independent-samples *t*-tests were conducted to compare differences in implicit bias by participant gender. Table 15 shows that there were no gender differences in computer or essay grades, and there was no significant gender difference in implicit bias. Male participants had more sexist attitudes toward women, and more stereotypical beliefs regarding women in STEM, compared to female participants. There were no significant differences for the remaining explicit bias measures, all *p*s > .17.

There were no participant gender differences in feelings of warmth toward women or men. However, there was a significant difference between the feeling thermometer measures, with all participants reporting warmer, more favorable feelings toward women compared to men, $t(211) = 5.54$, $p = .00$. General favorability toward women (semantic differential scale) was not significantly different between male and female participants, but further analyses revealed gender differences in individual scale items. Female participants rated women as more wise (vs. foolish), compared to male participants. In contrast, male participants rated women as more good (vs. bad).

Table 15

Essay Grades (%), Implicit Bias, and Explicit Attitudes, by Participant Gender

	Male Participants	Female Participants		
Variable	*M* (*SD*)	*M* (*SD*)	*t*	*df*
Computer essay grade	72.96 (15.66)	73.47 (14.72)	-0.23	214
Exercise essay grade	81.85 (11.23)	83.56 (12.42)	-0.95	214
Gender-Science IAT	0.12 (0.43)	0.08 (0.38)	0.65	209
Modern Sexism Scale	2.71 (0.62)	2.34 (0.55)	4.26***	211
Women in STEM Stereotype Scale	2.45 (0.56)	2.27 (0.47)	2.37*	211
Feeling Thermometer: Women	70.54_{a} (14.51)	71.74_{a} (16.52)	-0.50	210
Feeling Thermometer: Men	64.54_{b} (15.18)	67.92_{b} (17.12)	-1.36	210
Semantic Differential (SD)	3.69 (0.55)	3.69 (0.54)	0.10	210
Individual SD Scale Items:				
Wise-Foolish	3.34 (0.84)	3.67 (0.78)	-2.73**	210
Good-Bad	3.83 (0.86)	3.53 (0.83)	2.40*	210
Logical-Irrational	3.05 (0.96)	3.29 (0.88)	1.65†	209
Analytical-Emotional	2.11 (0.78)	2.29 (0.73)	1.62†	210

Note. Means with differing subscripts are significantly different, *p* < .05.

† *p* < .10, * *p* < .05, ** *p* < .01, *** *p* < .001.

Interaction of Author Gender, Rubric, and Participant Gender on Grades

Author gender and rubric were expected to interact with participant gender to affect essay grades. In the no rubric condition, male participants were expected to grade male authors more favorably and female participants were expected to grade female authors more favorably. The rubric was expected to reduce this effect such that there would be no differences between male and female author grades by participant gender. This hypothesis was tested using a 3 (gender-topic condition) x 2 (rubric condition) x 2 (participant gender: male, female) MANOVA with computer and exercise grades as the dependent variables.

As shown in Table 16, there was a main effect of rubric on computer grades, $F(1, 204) = 15.38$, $p = .00$, such that computer essays graded with the rubric received significantly lower grades. This was the only significant effect on computer grades.

Table 16

Mean Computer Essay Grades (%) by Author Gender, Participant Gender, and Rubric

		Author Condition		
	Participant	No name	Female author	Male author
	Gender	*M (SD)*	*M (SD)*	*M (SD)*
Rubric	Male	66.43_a (12.15)	67.69 (14.38)	70.00 (15.97)
	Female	71.43 (14.33)	66.90_a (16.54)	70.42_a (15.81)
No Rubric	Male	82.78_b (6.67)	72.00 (20.44)	77.50 (16.29)
	Female	77.69 (14.54)	76.00_b (10.86)	79.05_b (13.38)

Note. Means with differing subscripts within columns are significantly different, $p < .05$.

For exercise grades, Table 17 shows a significant 3-way interaction of gender-topic, rubric, and participant gender for exercise essay scores, *F*(2, 204) = 3.60, *p* = .03. In the no rubric condition, male participants graded the male exercise author significantly lower than the anonymous exercise author, and marginally lower than the female author. However, when using the rubric, grades for male and female authors were not significantly different.

Table 17

Mean Exercise Essay Grades (%) by Gender-Topic, Rubric, & Participant Gender

		Gender-Topic Condition		
	Participant Gender	Anonymous *M (SD)*	Female author *M (SD)*	Male author *M (SD)*
Rubric	Male	75.00_a (11.55)	85.45_b (8.20)	82.69 (11.29)
	Female	85.18 (11.26)	83.96 (12.33)	81.50 (13.19)
No Rubric	Male	88.06_b (10.44)	$82.20^†$ (11.01)	$75.50_a^†$ (11.89)
	Female	83.06 (15.27)	85.81 (11.77)	82.15 (10.92)

Note. Means with differing subscripts are significantly different, *p* < .05. Means marked with † are marginally different from one another, *p* < .10.

IV. DISCUSSION

This study evaluated the effects of implicit and explicit gender bias on grading, and examined whether implicit stereotypes are predictive of discriminatory grading, whether implicit bias is a better predictor than explicit measures of sexism, stereotype endorsement, and favorability, and whether implicit and explicit measures interact, revealing effects of implicit bias even in the absence of explicitly prejudiced attitudes. The current study further examined whether rubrics help decrease the effects of bias on grading outcomes. Finally, grading differences among author gender-essay topic pairs were compared by participant gender.

The female author of the computer essay received lower grades than the male and anonymous computer authors, and the male author of the exercise essay received lower grades than the female and anonymous exercise authors. Further, whereas none of the explicit attitude or stereotype measures predicted computer essay grades, implicit bias was significantly related to computer grades. As expected, participants who were low in explicit prejudice toward women in STEM, but who were also high in implicit gender-science bias, graded the female computer essay significantly lower. Rather than reducing the effects of bias, rubric use enhanced the effect of bias on grading. One possible explanation for this surprising finding might be system justification bias (Jost, Banaji, & Nosek, 2004), whereby individuals unconsciously engage in behaviors that will bolster the status quo. Alternately, these results could indicate that rubric use increases demand on cognitive resources already limited in individuals whose implicit and explicit attitudes are dissonant (Park et al., 2008).

Summary of Results

Evidence of implicit gender bias. Participants' use of descriptive words and pronouns to describe authors were expected to reflect implicit associations between author gender and essay topic. Participants who read the computer essay with no author name were expected to use a male pronoun ('he') more than a female pronoun ('she') to describe the author. For the anonymous exercise essay, participants were expected to ascribe male and female pronouns equally. This hypothesis was supported. The anonymous author of the computer essay was described using a male pronoun significantly more often than any other pronoun, suggesting that when participants read the computer essay with no author name, they envisioned the author as male rather than female. For the exercise essay, there was no difference in frequency of male or female pronouns used.

It was expected that participants' descriptions of authors would reflect implicit gender norms and associations. Descriptions given for the female author were expected to be shorter than for the male. Compared to the male author, descriptions of the female author were expected to contain more references to appearance, agreeableness, neuroticism, and low intelligence. The female author was also expected to receive fewer descriptions indicating intelligence, interest in majoring in a STEM field, interest in video games, and a nerdy or geeky personality.

This hypothesis was partly supported. There was no significant difference in length of descriptions used between the male and female authors, but the content of the descriptions differed depending on author gender. As expected, the female author received significantly fewer references to interest in video games and more criticisms regarding poor writing ability, compared to the male author. Contrary to the hypothesis, there were no significant differences references to intelligence, interest in STEM fields of study, or descriptions of the authors as nerdy. Additionally, the male author received marginally more references to physical appearance, compared to the female author. Many of these descriptors were remarkably similar (e.g., ‘wears glasses, not athletic’) perhaps suggesting that participants possessed a mental schema about what a male computer user should look like. It can be surmised that participants did not possess a similar mental schema about what a female computer enthusiast should look like, presumably because they have been exposed to fewer exemplars.

Finally, it is worth noting the four instances of participants who described the female author in a way that implied dishonesty regarding the author’s interest in or knowledge of computers. No comments to this effect were made about the male author. This could be an example of punishment for violating prescriptive gender norms (Eagly, 1987; Prentice & Carranza, 2002; Rudman & Glick, 2001). Because participants had a more difficult time accessing a mental image of the female author, it is also possible that this translated to a feeling of distrust toward her.

These findings confirm recent research that has shown a decreasing gap in gender stereotypes within certain areas, but not others (Buchmann et al. 2008). For example, there is no longer a significant difference in the association of gender with the labels ‘nerdy’ or ‘geeky’, but there is still a strong difference in association of gender with involvement in video games (more strongly male) and expectations of good writing ability (more strongly female).

Implicit and explicit scores and their effects on essay grades. Implicit bias was expected to predict computer essay grades, but explicit attitudes (favorability, stereotype endorsement, warmth) were not expected to predict computer grades. This hypothesis was supported, with implicit bias predicting computer grades, and also contributing a significant amount incremental variance in computer grades beyond explicit attitudes. Implicit bias was not expected to correlate with explicit attitudes, which was also supported.

These findings indicate that implicit association of gender and science had a significant effect on grading decisions for the computer essay (a STEM field). Neither explicit evaluative attitudes (i.e. favorability and warmth toward women) nor explicit stereotype endorsement (i.e. sexism and prejudice against women in STEM) were significant predictors of computer grades. Participants with stronger associations of men with science (as opposed to women) graded the computer essay more harshly. This supports previous literature by contributing to the predictive validity of the Gender-Science IAT, and by demonstrating that implicit bias is a stronger predictor of grading compared to explicit attitudes (Greenwald et al., 2009; McConnell & Leibold, 2001; Rudman & Ashmore, 2007). Rudman and Ashmore (2007) found that evaluative stereotype IATs, such as the gender-science IAT used in this study, predicted overtly discriminatory behaviors, even after controlling for explicit attitudes. In contrast, traditional attitude IATs (good vs. bad; pleasant vs. unpleasant) did not predict behaviors. Mental schemas about roles and behaviors that are considered socially acceptable (Eagly, 1987; Heilman & Eagly, 2008; Fuchs et al., 2004; Olson & Fazio, 2004) also contribute to implicit associations, and are linked with prejudicial behaviors when these norms are violated (Prentice & Carranza, 2002; Rudman & Glick, 2001). This study represents a meaningful addition to research on implicit bias, by providing further evidence of a connection between implicit bias and discriminatory behaviors. Studies that test this connection between automatic associations and deliberate behaviors, though vital to assessing the validity of the IAT, are lacking in current literature (Greenwald et al., 2009; Rudman & Ashmore, 2007), making the current study even more relevant and meaningful.

Neither implicit bias nor explicit attitudes were expected to correlate significantly with exercise essay grades. This was partially supported. As expected, the IAT did not correlate significantly with exercise grades, nor did the Modern Sexism Scale or the Women in STEM Stereotype Scale. These findings make sense intuitively, as one would not expect gender-science association or attitudes to correlate with grades for an essay that is neither STEM-related nor strongly gendered. In particular, these findings provide divergent validity for the Gender-Science IAT by demonstrating that it is measuring a construct related to gender associations with science, rather than merely general favorability toward women.

Surprisingly, evaluative attitudes toward women correlated significantly with the exercise essay grades. More favorable evaluations of women (as measured by the semantic differential scale) resulted in higher exercise essay grades for the male and female authors. This could suggest that generally positive attitudes toward women can be predictive of decision-making when the subject matter does not break with gender norms.

This would be consistent with research on prescriptive gender norms, which finds that women and men avoid social punishment as long as they behave in ways that are congruent with their gender (Prentice & Carranza, 2002; Rudman & Glick, 2001). This is also consistent with theories about gender roles, which hold that a group's beliefs about the kind of roles that men and women should occupy inform and reinforce gender role stereotypes (Eagly, 1987; Fuchs et al., 2004; Heilman & Eagly, 2008; Koenig & Eagly, 2014). Taken together, these results also support the theory of benevolent sexism, whereby individuals report generally positive attitudes toward women, yet also endorse stereotypes about women that can result in prejudice overtly or covertly (Christopher & Wojda, 2008; Cuddy, Fiske, & Glick, 2008).

Effect of author gender on essay grades. Author gender was expected to significantly affect computer essay grades, such that higher grades would be given for stereotype-consistent author-topic essays (male author of computer essay) than for stereotype-inconsistent (female author of computer essay) pairings. This hypothesis was partially supported. The female author of the computer essay received marginally lower grades than the male and anonymous computer authors. For the exercise essay, the female author was expected to receive higher grades than the male author. This hypothesis was based on research showing that people view women as being better writers (Buchmann et al., 2008), and women tend to get higher grades in school in areas that are not strongly-gendered (Ackerman et al., 2013). This was also partially supported; the anonymous and female exercise authors received marginally higher grades than the male author.

The practical significance is worth noting, as the female computer author received lower letter grades than the male author or the anonymous author. Participants received a grading key as part of their grade sheet that provided point ranges corresponding to percentages that represent commonly-used letter grades. Although the female computer grade was only marginally different, if this same degree of effect occurred in a real classroom, the female author would receive a letter grade of ‘D’, compared to the ‘C’ assigned to the male and anonymous author. This can represent the difference between failure and success in academia.

Effect of implicit bias on grades, by author gender. Grading decisions were expected to differ depending on participants’ level of implicit bias. This was supported for computer grades. Participants who were high in implicit bias graded the female computer author significantly lower than the male computer author, and lower than the anonymous computer author, but the anonymous and male authors were not significantly different. Participants who were low in implicit bias graded the anonymous author higher than both the male and female authors, and there was no difference in grades assigned to the male and female computer authors. In the case of exercise grades, participants high in implicit bias graded the anonymous exercise author marginally higher than the male author, whereas participants low in implicit bias graded the female author significantly higher than the male author, suggesting a benefit to female authors when participants had a stronger association between women and STEM.

Interaction of implicit and explicit attitudes on grades, by author gender. Participants with low sexism coupled with high implicit bias were expected to assign lower grades to the female author but not to the male or anonymous authors. This hypothesis was partially supported. The interaction of implicit bias and sexism was marginal for the computer grades, but there were significant differences by level of implicit bias in the high sexism group. Participants who were high in sexism graded the computer essay lower when they were also high in implicit bias, compared to the combination of high sexism and high bias. Contrary to the hypothesis, computer grades assigned by participants who were low in sexism did not differ by level of implicit bias. This hypothesis was based on the finding that people can explicitly report positive attitudes, but also possess conflicting implicit bias. Whether an individual's self-report is the result of social desirability or a lack of awareness of personal bias, the unacknowledged implicit bias can result in discrimination (Greenwald et al., 2009; McConnell & Leibold, 2001; Rudman & Ashmore, 2007). Contrary to this, low implicit bias seems to have buffered the effect of sexism. There were no differences in the male computer grades by implicit bias or sexism. This further supports the need to consider implicit bias in addition to self-reports when predicting discriminatory behaviors.

The interaction of implicit bias and explicit prejudice toward women in STEM on computer grades was also examined. Participants with low prejudice coupled with high implicit bias were expected to assign lower grades to the female author but not to the male or anonymous authors. This hypothesis was supported. The interaction was not significant, but participants who were low in explicit prejudice and high in implicit bias graded the female computer author significantly lower than the male computer author, as predicted. Similarly, the anonymous author was graded lower by participants who were low in prejudice and high in implicit bias, compared to participants with low prejudice and low bias. Within participants with high prejudice, grades did not differ by level of implicit bias.

These findings support the hypothesis that participants who do not explicitly endorse stereotypes about women in STEM, but who have a stronger implicit association of men with science, would grade the stereotype-inconsistent essay (i.e. female computer) lower than the stereotype-consistent essay (male computer). One can view women favorably overall, yet still maintain implicit stereotypes, which can ultimately affect behaviors (Buchmann, 2004; Christopher & Wojda, 2008; Eagly & Mladnic, 1994; Rudman & Ashmore, 2007). These findings provide convergent validity for the Women in STEM Stereotype scale. Explicit stereotypes regarding women in STEM interacted with implicit bias regarding women in STEM, affecting grades by author gender for the STEM topic, but not for the exercise essay (i.e. non-STEM).

Effectiveness of rubrics to increase consistency in grading. Essay grades were expected to have greater consistency when a rubric was used to assign grades, compared to when a rubric was not used. This hypothesis was not supported. Although the ranges of grades assigned to essays in the rubric condition were smaller than the ranges of grades assigned to essays in the no rubric condition, there was no difference in variance between the rubric and no rubric groups. One of the mechanisms by which rubrics are thought to increase fairness is by increasing grading consistency between independent graders (Jonsson & Svingby, 2007; Renzai & Lovorn, 2010; Silvestri & Oescher, 2006). The lack of support for this in the current study may be due to the use of novice graders, rather than experienced graders or participants who have received rubric training. However, even among educators with extensive experience and training in rubric use, there can still be substantial differences among teachers, across academic fields, and even within a teacher's grading history (Tierney & Simon, 2004).

Effectiveness of rubrics to decrease bias effects on grading. Rubric use was expected to reduce differences in grades resulting from implicit and explicit bias, such that any biased grading in the no rubric condition, would no longer appear when the rubric was used. This hypothesis was not supported. High implicit bias resulted in lower female computer grades, as expected, but this effect was only significant in the rubric condition. Rather than reducing the effect of bias, the rubric appears to have enhanced it.

This alternative explanation was examined, and support was found for the hypothesis that the rubric contributed to biased grading outcomes. Even after controlling for implicit and explicit measures, and after controlling for grade rankings on the exercise essay, rubric use explained a significant amount of variance in computer. Cognitive dissonance, such as that seen in the current study wherein implicit and explicit attitudes are misaligned, can result in more extreme biased behaviors (Park et al., 2008). Rubric use for these individuals might have further depleted cognitive resources, enhancing the effect of bias. Alternately, rubric use might have provided the means to justify biased grading through unconscious bolstering of the status quo (Jost, Banaji, & Nosek, 2004; Jost, Pelham, & Carvallo, 2002). A final possibility is that, because a common stereotype about women is that they have better writing and communication abilities (Buchmann et al., 2008), and because rubric criteria emphasize the importance of written communication and clarity, the female author might have been held to a higher standard than the male author.

Gender differences in implicit and explicit attitudes. Male and female participants were expected to differ in both implicit and explicit bias levels. Male participants were expected to report more stereotype endorsement and less favorable attitudes toward women, compared to female participants (see Jackson, Hillard, & Schneider, 2014). This was supported for measures of sexism and prejudice. Male participants reported greater endorsement of sexist and stereotype statements. However, evaluative attitude measures (i.e. those that measure favorability or warmth) were not significantly different between men and women. These findings support previous research that has shown that men and women both generally report warm, favorable evaluative attitudes toward women, while still endorsing gender stereotypes. Benevolent sexism continues to result in disparate treatment toward women, despite self reports proclaiming that women are “good” (Buchmann, 2004; Mladnic & Eagly, 1994). Further supporting these findings was the emergence of significant differences between male and female participants on individual semantic differential scale items. Males rated women as “good” (more than “bad”), but also rated women as more “foolish” (rather than “wise”). These seemingly conflicting responses support the concept of benevolent sexism (Christopher & Wojda, 2008; Cuddy, Fiske, & Glick, 2008; Eagly & Mladnic, 1994), reflecting favorable evaluative attitudes while simultaneously maintaining gender stereotypes and implicit bias.

Male participants were expected to have stronger implicit associations of men with science, compared to female participants (Jackson, Hillard, & Schneider, 2014). This hypothesis was not supported. There was no significant difference between male and female participants’ implicit bias. The lack of difference in implicit associations by participant gender is consistent with the results of a large-scale, multi-country study of Implicit Association Task results (Nosek et al., 2002). Across more than 35 countries, female and male participants did not differ in average gender-science IAT scores, but there was a significant difference in self-reported attitudes, with males explicitly expressing more stereotypical associations of women in STEM. This latter finding is consistent with the results of the current study.

Interaction of author gender, rubric, and participant gender on grading. A three-way interaction between author gender, rubric, and participant gender was hypothesized. It was expected that male participants would grade the male author more favorably than the female author, and female participants would grade the female author more favorably. Rubric use was expected to moderate this relationship. This hypothesis was not supported for computer essay grades, but author gender, rubric, and participant gender did interact to affect exercise grades. Exercise grades assigned by female participants did not differ among author condition, or between rubric and no rubric condition. For male participants, average grades differed by author condition in both the rubric and no rubric conditions. Male authors were penalized on the exercise essay when there was no rubric, resulting in significantly lower grades for the male author compared to the anonymous author, and marginally lower grades for the male author compared to the female author. The rubric reduced this bias effect for exercise grades, resulting in no difference between the male and female author. Surprisingly, the anonymous author received significantly lower grades than the female author when the rubric was used. It is not clear from this data why that may have occurred.

Theoretical Implications

Explicit attitudes were not correlated with implicit bias, but implicit bias did predict computer grades, similar to findings from previous attitude research (Greenwald et al., 2009; Nosek, 2005). Whereas explicit attitudes did not significantly predict biased grading outcomes, the Gender-Science IAT was a significant predictor of grades for the computer essay, especially when author name (i.e. female computer) was stereotype-inconsistent (Greenwald et al, 2009; Rudman & Ashmore, 2007). As expected, the Gender-Science IAT did not predict grades for the exercise essay, a topic that was neither STEM-related, nor strongly gendered. These findings further validate the Gender-Science IAT for use in predicting discriminatory behaviors. Additionally, the interaction of prejudice toward women in STEM and implicit gender-science bias provides convergent validity for the newly developed Women in STEM Stereotype Scale (Jackson, Hillard, & Schneider, 2014). The fact that this measure of prejudice did not interact with any other explicit or implicit bias to affect grades on the exercise essay offers evidence of divergent validity for the scale.

Consistency and objectivity are considered two of the hallmarks of an effective rubric (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). There was no evidence to support an increase in consistency in the rubric condition, but rubric use did result in smaller grade ranges across author gender conditions. The rubric appears to have at least partially succeeded in this endeavor. Surprisingly, rather than reducing the effect of implicit gender bias on computer grades, the rubric appeared to enhance this relationship. Further, this occurred only when the author of the computer essay was female. The female author was expected to benefit most from the rubric. In contrast, the female author faired worst in the rubric group. This contradicts past claims that no negative effects of rubric use have been revealed (Renzai & Lovorn, 2010). As mentioned previously, it is possible that author gender was more salient to participants using rubrics, that the rubric provided the means to justify biased grading, or that the rubric further depleted cognitive resources already strained by dissonant implicit and explicit attitudes. Most importantly, these findings reveal a potential weakness in rubric use that has heretofore received no attention.

Practical Implications

Rubrics are generally regarded as useful assessment tools, purported to increase consistency, objectivity, and efficiency (Jonsson & Svingby, 2007; Rippé, 2008; Renzai & Lovorn, 2010; Silvestri & Oescher, 2006; Reddy & Andrade, 2010). They provide additional benefits with regard to clearly communicating expectations for students, increasing transparency (and therefore perceptions of fairness) in the grading process. Whereas some of the current research reveals no benefits in terms of consistency or objectivity (Tierney & Simon, 2004), at the very least, no research to date has revealed any negative effect of rubric use (Renzai & Lovorn, 2010). Yet questions still remain regarding the magnitude of benefit provided by rubric use, and whether they truly do increase fairness. As this study has shown, the mere presence of a rubric is clearly not sufficient to reduce bias effects.

When creating rubrics, designers should continue to assess outcomes beyond consistency when testing their effectiveness. This is especially true when considering situations or groups that are vulnerable to bias, which can result in discriminatory behaviors toward disadvantaged groups. This study raises the question of whether assessment tools are actually effective at decreasing bias. Clearly there is a need for additional research to confirm (or counter) the current findings and expand on our knowledge of how rubrics interact with conscious and unconscious cognitive processing. The belief that rubrics at worst do no harm is of particular concern. Using a rubric could provide a false sense of security, leading graders to believe that they are now immune to the negative impact of bias, and reluctant to explore personal bias as a potential contributor of systematic discrimination. This state could serve to further enhance the effects of implicit bias on grading. The finding that women are held to a higher standard of writing and communication should also be considered in rubric design. Wording could potentially be changed to reduce the emphasis on writing, where appropriate, and if the unequal weight assigned to women's writing can be quantified, perhaps a correction could be applied statistically to reduce the disparate impact of gender bias.

Until these concerns have been resolved, these findings support recommendations to keep raters blind to author or applicant names when grading or evaluating performance (Budden et al., 2008). Techniques to reduce bias should be considered essential to the rating process, as grades and performance ratings directly affect attrition, retention, and graduation in academics, and selection, retention, and promotion in professional settings. Special attention needs to be paid in traditionally white-male-dominated fields, as marginalized groups are already more vulnerable to bias in these settings.

Methods for reducing the effects of bias. A number of researchers maintain that, while they can be difficult to change, attitudes are malleable (Blair, 2002; Rudman, Ashmore, & Gary, 2001). Many institutions of higher learning are making a concerted effort to recruit, hire, and promote female STEM researchers and faculty. These initiatives arose, in part, as a result of research that has shown that exposure to target exemplars, especially those in leadership positions or who demonstrate attributes that are deemed desirable by society, can effectively reduce automatic bias (Dasgupta & Asgari, 2004). On the other hand, exposure to exemplars can result in backlash if a target group member violates prescriptive norms (Prentice & Carranza, 2002).

Initiatives that have been shown to work at least modestly well include those that appreciate differences rather than trying to eliminate or ignore them, and diversity education that focuses on bias education and fear reduction (Rudman, Ashmore, & Gary, 2001). Rather than trying to suppress thoughts about a target group, some researchers have shown that teaching people to be aware of their implicit and explicit bias, and encouraging *more* thinking about the underlying reasons for bias is effective in reducing stereotypes (MacCrae, Milne, & Bodenhausen, 1994; Richards & Hewstone, 2000). Petty, Wegener, and White (1998) describe a method that encourages effortful mental processing as a way of reducing or correcting for bias. This "correction process" is the process by which people consciously adjust their assessments of a target in order to correct for the effect of perceived bias. This method has been shown to reduce the impact of other variables (e.g. source likability or in-group identification) when people respond to persuasive messages.

These promising findings could be the result of increased effortful cognitive processing, but current research shows that increasing effortful processing alone does not always eliminate the effects of bias; in fact, increased processing itself can be very biased (Petty, Wegener, & White, 1998). It is, therefore, not enough to simply instruct people to think carefully about their actions to avoid bias. The ability to control explicit responses might have no effect on implicit associations or prejudicial behavior, and people can possess competing attitudes toward a target group (Christopher & Wojda, 2008; Cuddy, Fiske, & Glick, 2008). The current study supports these latter findings; consciously controlled explicit attitudes competing with high implicit stereotyping resulted in lower grades in the target group (i.e. female author of the computer essay). Using a rubric, which should have increased effortful cognitive processing during grading, not only failed to reduce the impact of bias, but actually appeared to enhance the negative effects of unconscious stereotyping.

Limitations

As is the case with any research, there were some limitations in the current study. First, participants were all undergraduate students in a midwestern university, so results might not be generalizable to the rest of the population. While the range of ages did include a subset of individuals who represented other age groups, their frequency was not large enough to allow for group comparisons. The diversity of ethnicity in the study is also limited to predominantly white students. Cultural differences could inform unconscious associations regarding prescriptive gender norms. Finally, as is often the case with undergraduate psychology student subject pools, there are significantly more female participants than male participants, limiting the use of some participant gender comparisons. Replicating this study with a sample that is more evenly representative of demographic groups could increase generalizability and external validity. It is likely, however, that the pattern of effects observed here would not change. Nosek, Banaji, and Greenwald (2002) found no significant relationships in implicit gender-science associations by participant gender, age, or ethnicity.

Another limitation was the lack of training provided on using the rubric. The participants in this study are novice graders, with little to no experience in using a rubric. A number of researchers (Hitt & Helms, 2009; Jonsson & Svingby, 2007; Reddy & Andrade, 2010; Renzai & Lovorn, 2010) strongly endorse the need for training before using rubrics. It is common for novice rubric users in training to grade an assignment and compare their grading decisions with someone who is considered an expert. Talking through the justification used for each objective helps fine-tune the process so that future ratings have greater inter-rater reliability. However, this technique might not have changed the current pattern of results. Rubric training is not guaranteed to result in greater consistency; even among experienced graders, training can have little to no effect on inter-rater reliability (Bloxham et al., 2015; Pufpaff, Clarke, & Jones, 2015; Tierney & Simon, 2004). As Bullough (2010) illustrated, even when this supposedly objective technique employed multiple raters with rubric-use experience, there was a considerable amount of subjective decision-making that went into reconciling rater differences.

It is possible that the use of the ostensible two-study design could have affected participant performance, particularly if they indicated suspicion about the deception or relatedness of the two tasks. It is unlikely that participants identified the deception. Only two participants indicated suspicion, but added that they only considered the possibility of deception after the study was done. They further did not identify the purpose of the grading tasks. This is also of little concern; if any participant were to have identified the deception and the true purpose of the study, such a realization would have occurred after they had finished the grading task. As a result, this knowledge would have had no effect on essay grades, which were completed prior to the deception and bias measures.

Finally, the current study did not control for prior exposure to or knowledge of the IAT. Nosek, Banaji, ask Greenwald (2002) asked participants to indicate how many times they had taken any IAT previously, in order to control for potential practice effects. It is unlikely that most participants in this sample would be familiar with the IAT, but there were approximately 5 participants near the end of data collection who had learned about the IAT in their introductory psychology course prior to participation. However, it is unlikely that this limitation had any negative effect on the current study, since results from the Harvard Implicit demonstration website (Nosek et al., 2002) revealed no significant practice effect on IAT scores, and responses of those participants who had been exposed did not differ systematically compared to other participants in the study.

Future Research

Future research should include a more diverse sample of participants in terms of age, gender, and ethnicity. This will allow researchers to parse out more sources of variance in grading outcomes and could inform more tailored methods for reducing bias effects. Future studies should also include participants with a range of prior experience in grading and using rubrics, including current instructors or teaching assistants to assess the degree of influence that implicit bias might have on “expert” graders. Comparisons between novice graders who receive or do not receive training on rubric would provide valuable insight into the effectiveness of rubrics. Finally, implicit bias training could be introduced and subsequent grading behavior between the rubric and no rubric techniques compared.

In the future, psycholinguistic analyses should be conducted using transcripts from actual recorded conversations or from participants’ written responses. This would allow for more accurate measures of word count and descriptor counts, and would reduce demand effects on participants' answers. It is reasonable to assume that such methods would confirm and strengthen the current findings.

Conclusion

Implicit gender-science bias predicted discriminatory grading for the computer essay, whereas self-report measures of attitudes toward women did not, as expected. Implicit bias explained additional variance in grades, over and above explicit measures. Essay grades differed depending on whether the author names were stereotype-consistent or not, resulting in greater penalties given to authors who violated gender-role expectations (i.e. female computer author; male exercise author). The practical significance of the grade differences is also worth noting, as the female author of the computer essay consistently received grades that were a letter grade below those given to the male author. This degree of difference would result in significant deficits in academic and professional success, further widening the gender gap in STEM fields. As predicted, implicit bias and explicit attitudes were not correlated, but explicit sexism and prejudice toward women in STEM interacted with implicit bias. Participants who reported low prejudice differed by implicit bias level, such that those with dissonant implicit and explicit attitudes graded stereotype-inconsistent authors more harshly. When sexism and implicit bias were examined, implicit bias appears to have buffered the impact of explicit sexism on grades. The rubric reduced the range of grades, but did not increase consistency. Surprisingly, the rubric not only failed to create more equitable grading, it appears to have enhanced the effect of implicit, resulting in more discriminatory grading. This suggests that rubric use alone is not sufficient to reduce bias effects, and the hallmark standard of rubric use to increase consistency should be only one measure of its effectiveness.

As academic and professional fields that were once highly segregated by gender are working to increase representation of women, methods for reducing the effects of implicit and explicit bias are essential. In addition to contributing to the body of research on the effects of implicit bias on behaviors and methods for assessing this relationship, this study also sheds light on a potential disadvantage of rubric use, a finding with wide-spread implications for a range of assessment evaluation tools.

V. REFERENCES

Aberson, C. L., & Haag, S. C. (2007). Contact, perspective taking, and anxiety as predictors of stereotype endorsement, explicit attitudes, and implicit attitudes. *Group Processes*, *10*(2), 179-201. doi: 10.1177/1368430207074726.

Ackerman, P. L., Kanfer, R., & Beier, M. E. (2013). Trait complex, cognitive ability, and domain knowledge predictors of baccalaureate success, STEM persistence, and gender differences. *Journal of Educational Psychology*, *105*(3), 911.

Allen I. E. & Seaman, J. (2014, January). Grade Change: Tracking Online Education in the United States. Babson Survey Research Group: The Sloan Consortium.

Allport, G. W. (1954). *The nature of prejudice*. Cambridge, MA: Addison-Wesley. American Association of University Women, 2001.

Axelson, R. D., Solow, C. M., Ferguson, K. J., & Cohen, M. B. (2010). Assessing implicit gender bias in medical student performance evaluations.*Evaluation & the health professions*, *33*(3), 365-385.

Blair, I.V., Ma, J. E., & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828-840.

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality & Social Psychology Review (Lawrence Erlbaum Associates)*, *6*(3), 242-261. doi: 10.1207/S15327957PSPR0603_8.

Buchmann, C., DiPrete, T. A., & McDaniel, A. (2008). Gender inequalities in education. *Annu. Rev. Sociol*, *34*, 319-337.

Budden, A. E., Tregenza, T., Aarssen, L. W., Koricheva, J., Leimu, R., Lortie, C. J. (2008, January). Double-blind review favours increased representation of female authors. *Trends in Ecology & Evolution, 23*(1), 4-6.

Bullough, R. V., Jr. (2010). Proceed with caution: Interactive rules and teacher work sample scoring strategies, an ethnomethodological study. *Teachers College Record, 112*(3), pp. 775-810.

Carnes, M., Devine, P. G., Isaac, C., Manwell, L. B., Ford, C. E., Byars-Winston, A., Fine, E. & Sheridan, J. (2012). Promoting institutional change through bias literacy. *Journal of Diversity in Higher Education, 5*(2), 63-77.

Chen, M. & Bargh, J.A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology, 33*, 541-560.

Christopher, A.N. & Wojda, M.R. (2008). Social dominance orientation, right-wing authoritarianism, sexism, and prejudice toward women in the workforce. *Psychology of Women Quarterly, 32*, 65-73.

Cohen (1992) - Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal and Social Psychology, 65*, 145-153.

Benard, S., Palik, I., & Correll, S. J. (2007). Cognitive bias and the motherhood penalty. *Hastings LJ*, *59*, 1359.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in experimental social psychology*, *40*, 61-149.

Dasgupta, N., & Asgari, S. (2004). Seeing is believing: Exposure to counterstereotypic women leaders and its effect on the malleability of automatic gender stereotyping. *Journal of Experimental Social Psychology, 40,* 642–658.

Devine, P.G. (1989). Stereotypes and prejudice: Their automatic and controlled components. Journal of Personality and Social Psychology, 56, 5–18.

Eagly, A. H. (1987). *Sex differences in social behavior: A social-role interpretation.* Hillsdale, NJ: Erlbaum.

Eagly, A. E., & Chaiken, S. (1993). *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

Eagly, A. H., & Mladinic, A. (1994). Are people prejudiced against women? Some answers from research on attitudes, gender stereotypes, and judgments of competence. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 5, pp. 1–35). New York: Wiley.

Easterly, D. M., & Ricard, C. S. (2011). Conscious Efforts to End Unconscious Bias: Why Women Leave Academic Research. *Journal of Research Administration*, *42*(1), 61-73.

Fuchs, D., Tamkins, M. M., Heilman, M. E., & Wallen, A. S. (2004). Penalties for Success: Reactions to Women Who Succeed at Male Gender-Typed Tasks. Journal of Applied Psychology, 89(3), 416-427.

Gage, N., & Berliner, D. (1992). *Educational psychology* (5th ed.), Princeton, New

Jersey: Houghton Mifflin Company.

Garland, D. & Martin, B. N. (2005). Do gender and learning style play a role in how online courses should be designed? *Journal of Interactive Online Learning.*

Goldberg, P. (1968). Are women prejudiced against women?. *Society*, *5*(5), 28-30.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4-27. doi: 10.1037/0033-295X.102.1.4.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology, 85*, 197–216.

Greenwald, A.G., McGhee, D.E., Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*(6), 1464-1480.

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of personality and social psychology, 97*(1), 17.

Gunn, C. (2002). Gender issues in computer-supported learning. *Research in Learning Technology, 10*(1), 32-44.

Gunn, C., McSporran, M., Macleod, H., & French, S. (2003). Dominant or different? Gender issues in computer supported learning. *Journal of Asynchronous Learning Networks 7*(1), 14-30.

Hackney, A. (2005). Teaching students about stereotypes, prejudice, and discrimination: An interview with Susan Fiske. *Teaching of Psychology*, *32*(3), 196-199. doi:10.1207/s15328023top3203_13

Hafner, J., & Hafner, P. (2003). Quantitative analysis of the rubric as an assessment tool: An empirical study of student peer-group rating. *Int. J. Sci. Educ.*, *25*(12), 1509-1528.

Hamann, K., Pollock, P. H., & Wilson, B. M. (2009). Who SoTLs where? Publishing the scholarship of teaching and learning in political science. *PS: Political Science & Politics*, *42*(04), 729-735.Harasim, L. (2000). Shift happens: Online education as a new paradigm in learning. *The Internet and higher education*, *3*(1), 41-61.

Heilman, E.H. & Eagly, A.H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. Industrial and Organizational Psychology, 1, 393-398.

Hitt, A. M., & Helms, E. C. (2009). Best in show: Teaching old dogs to use new rubrics. *The Professional Educator*, *33*(1), 1.

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369-1385.

Jackson, S. M., Hillard, A. L., & Schneider, T. R. (2014). Using implicit bias training to improve attitudes toward women in STEM. *Social Psychology of Education*, *17*(3), 419-438.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational research review*, *2*(2), 130-144.

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political psychology*, *25*(6), 881-919.

Jost, J. T., Pelham, B. W., & Carvallo, M. R. (2002). Non-conscious forms of system justification: Implicit and behavioral preferences for higher status groups. *Journal of Experimental and Social Psychology, 38*, 586–602.

Karpinski, A. & Hilton, J.L. (2001). Attitudes and the Implicit Association Test. *Journal of Personality and Social Psychology*, *81*(5), 774-787. doi:10.1037//0022-3514.81.5.774

Kleinfeld, J. (1998). The Myth That Schools Shortchange Girls: Social Science in the Service of Deception.

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371.

Levenson, H., Burford, B., Bonno, B., & Davis, L. (1975). Are women still prejudiced against women? A replication and extension of Goldberg's study. *The Journal of Psychology*, *89*(1), 67-71.

Macrae, C.M., Milne, A.B., & Bodenhausen, G.V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology, 66*, 37-47.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, explicit attitudes, and discriminatory behavior. *Journal of Experimental Social Psychology*, *37*, 435–442.

Milkman, K., Akinola, M., & Chugh, D. (2014). *Discrimination in the Academy: A Field Experiment*. Working Paper, Wharton.

Morganson, V. J., Jones, M. P., & Major, D. A. (2010). Understanding women's underrepresentation in science, technology, engineering, and mathematics: The role of social coping. *The Career Development Quarterly*, *59*(2), 169-179.

Moni, R. W., Beswick, E., & Moni, K. B. (2005). Using student feedback to construct an assessment rubric for a concept map in physiology. *Advances in physiology education*, *29*(4), 197-203.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. Proceedings of the National Academy of Sciences of the United States of America, 109, 16474-16479.

Nass, C., Moon, Y., & Green, N. (1997). Are Machines Gender Neutral? Gender-Stereotypic Responses to Computers With Voices. *Journal of applied social psychology*, *27*(10), 864-876.

National Academy of Science. (2006). *Beyond Bias and Barriers: Fulfilling the Potential of Women in Academic Science and Engineering*. Washington, DC: National Academies Press.

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*, 565–584.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics, 6*, 101–115.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., Bar-Anan, Y., Bergh, R., Cai, H., Gonsalkorale, K., Kesebir, S., Maliszewski, N., Neto, F., Olli, E., Park, J., Schnabel, K., Shiomura, K., Tulbure, B., Wiers, R. W., Somogyi, M., Akrami, N., Ekehammar, B., Vianello, M., Banaji, M. R., & Greenwald, A. G. (2009). National differences in gender-science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*, 10593-10597.

Oakleaf, M. J. (2006). *Assessing information literacy skills: A rubric approach* (Doctoral dissertation, University of North Carolina at Chapel Hill).

Olson, M. A., & Fazio, R. H. (2004). The influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology, 86*(5), 653-667. doi: 10.1037/0022-3514.86.5.653.

Ottolini, M. C., Cuzzi, S., Tender, J., Coddington, D. A., Focht, C., Patel, K. M., & Greenberg, L. (2007). Decreasing variability in faculty ratings of student case presentations: A faculty development intervention focusing on reflective practice. *Teaching and learning in medicine, 19*(3), 239-243.

Pearson Jr, W. (1987). The flow of black scientific talent: Leaks in the pipeline. *Humboldt Journal of Social Relations*, 44-61.

Petty, R. E., Wegener, D. T., & White, P. H. (1998). Flexible correction processes in social judgment: Implications for persuasion. *Social cognition*, *16*(1), 93.

Pheterson, G. I., Kiesler, S. B., & Goldberg, P. A. (1971). Evaluation of the performance of women as a function of their sex, achievement, and personal history. *Journal of Personality and Social Psychology*, *19*(1), 114.

Popham, W. J. (1997). What's wrong – and what's right – with rubrics. *Educational Leadership, 55,* 72-75.

Postmes, T. & Spears, R. (2002). Behavior online: Does anonymous computer communication reduce gender inequality? *Personality and Social Psychology Bulletin 28*(8), 1073-1083.

Prentice, D. A. & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26,* 269-281.

Pufpaff, L. A., Clarke, L., & Jones, R. E. (2015). The Effects of Rater Training on Inter-Rater Agreement. *Mid-Western Educational Researcher*, *27*(2), 117.

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*(4), 435-448.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing. *Assessing writing*, *15*(1), 18-39.

Richards, Z. & Hewstone, M. (2001). Subtyping and subgrouping: Processes for the prevention and promotion of stereotype change. *Personality and Social Psychology Review, 5,* 52-73.

Rippé, C. (2008). Using rubrics to improve teaching, learning, and retention in distance education. *Online Classroom: Ideas for Effective Online Instruction*, p. 3-4.

Robinson-Cimpian, J. P., Lubienski, S. T., Ganley, C. M., & Copur-Gencturk, Y. (2014). Teachers' perceptions of students' mathematics proficiency may exacerbate early gender gaps in achievement. *Developmental psychology*,*50*(4), 1262.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science, 13*(2), 79-82. doi 10.1111/j.0963-7214.2004.00279.x.

Rudman, L. A., Ashmore, R. D., Gary, M. L. (2001). “Unlearning” automatic biases: The malleability of implicit prejudice and stereotypes. *Journal of Personality and Social Psychology, 81*(5), 856-867.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 743–762.

Siann, G., & Callaghan, M. (2001). Choices and barriers: Factors influencing women's choice of higher education in science, engineering and technology.*Journal of Further and Higher Education*, *25*(1), 85-95.

Silvestri, L., & Oescher, J. (2006). Using Rubrics to Increase the Reliability of Assessment in Health Classes. *International Electronic Journal of Health Education*, *9*, 25-30.

Simon, M., & Forgette-Giroux, R. (2001). A rubric for scoring postsecondary academic skills. *Practical Assessment, Research & Evaluation*, *7*(3).

Spelke, E. S., & Grace, A. D. (2007). Sex, math, and science. In S. J. Ceci & W. M. Williams (Eds.), Why aren’t more women in science? (pp. 57-77). Washington, D.C.: American Psychological Association

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

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

Steinpreis, R. E., Anders, K. A., & Ritzke, D. (1999). The impact of gender on the review of the curricula vitae of job applicants and tenure candidates: A national empirical study. *Sex roles*, *41*(7-8), 509-528.

Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations?. *Psychological Bulletin*, *105*(3), 409.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology*, *68*(2), 199-214.

Tierney, R., & Simon, M. (2004). What's still wrong with rubrics: focusing on the consistency of performance criteria across scale levels. *Practical Assessment, Research & Evaluation*, *9*(2), 1-10.

Towers, S. (2008). A case study of gender bias at the postdoctoral level in physics, and its resulting impact on the academic career advancement of females. Retrieved from http://arxiv.org/abs/0804.2026.

Trix, F., & Psenka, C. (2003). Exploring the color of glass: Letters of recommendation for female and male medical faculty. *Discourse and society, 14,* 191-220.

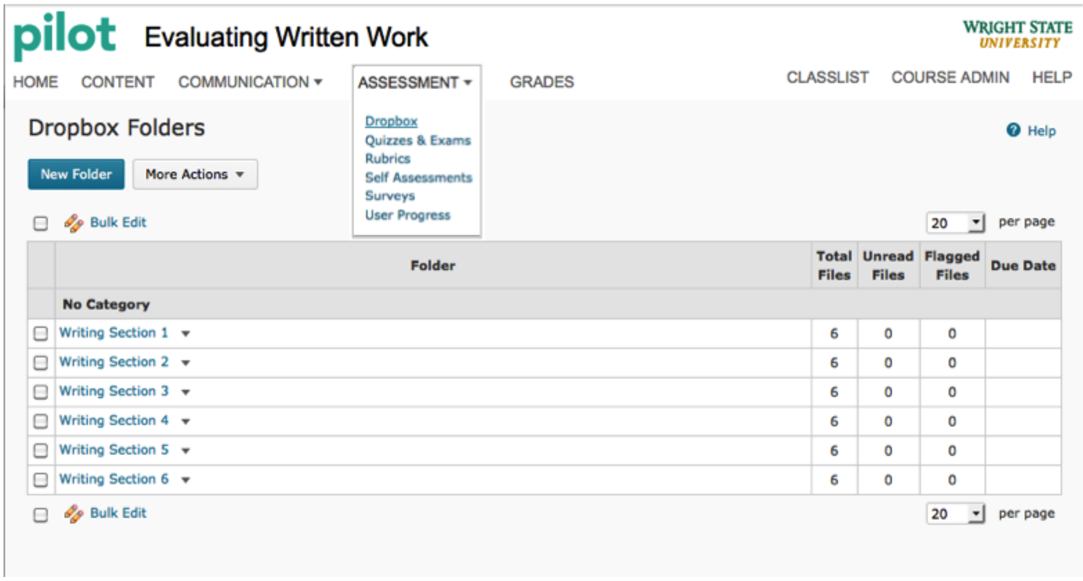
Turley, E. D., & Gallagher, C. W. (2008). On the" uses" of rubrics: reframing the great rubric debate. *English Journal*, 87-92.

Webb, S. (2001). Avatarculture: Narrative, power and identity in virtual world environments. *Information, Communication & Society*, *4*(4), 560-594.

Wilcox, C., Sigelman, L., & Cook, E. (1989). Some Like it Hot: Individual Differences in Responses to Group Feeling Thermometers. *Public Opinion Quarterly*, *53*(2), 246-257.

Wolfe, J. L. (1999). Why do women feel ignored? Gender differences in computer-mediated classroom interactions. *Computers and Composition, 16*, 153-166.

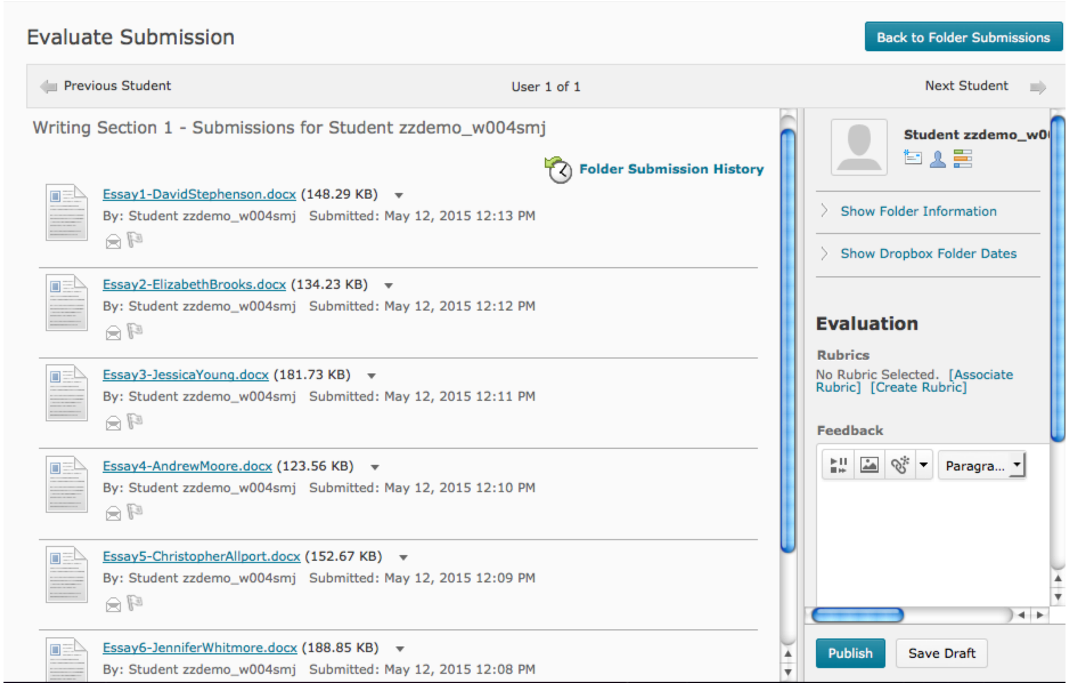
Appendix A: Online Learning Management System (LMS)



Dropbox Folders

	Folder	Total Files	Unread Files	Flagged Files	Due Date
	No Category				
□	Writing Section 1 ▾	6	0	0	
□	Writing Section 2 ▾	6	0	0	
□	Writing Section 3 ▾	6	0	0	
□	Writing Section 4 ▾	6	0	0	
□	Writing Section 5 ▾	6	0	0	
□	Writing Section 6 ▾	6	0	0	

Figure A-1. Participants accessed experimental essays through the assessment menu in the LMS.



Evaluate Submission

Writing Section 1 - Submissions for Student zzdemo_w004smj

Essay1-DavidStephenson.docx (148.29 KB)
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:13 PM

Essay2-ElizabethBrooks.docx (134.23 KB)
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:12 PM

Essay3-JessicaYoung.docx (181.73 KB)
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:11 PM

Essay4-AndrewMoore.docx (123.56 KB)
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:10 PM

Essay5-ChristopherAllport.docx (152.67 KB)
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:09 PM

Essay6-JenniferWhitmore.docx (188.85 KB)
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:08 PM

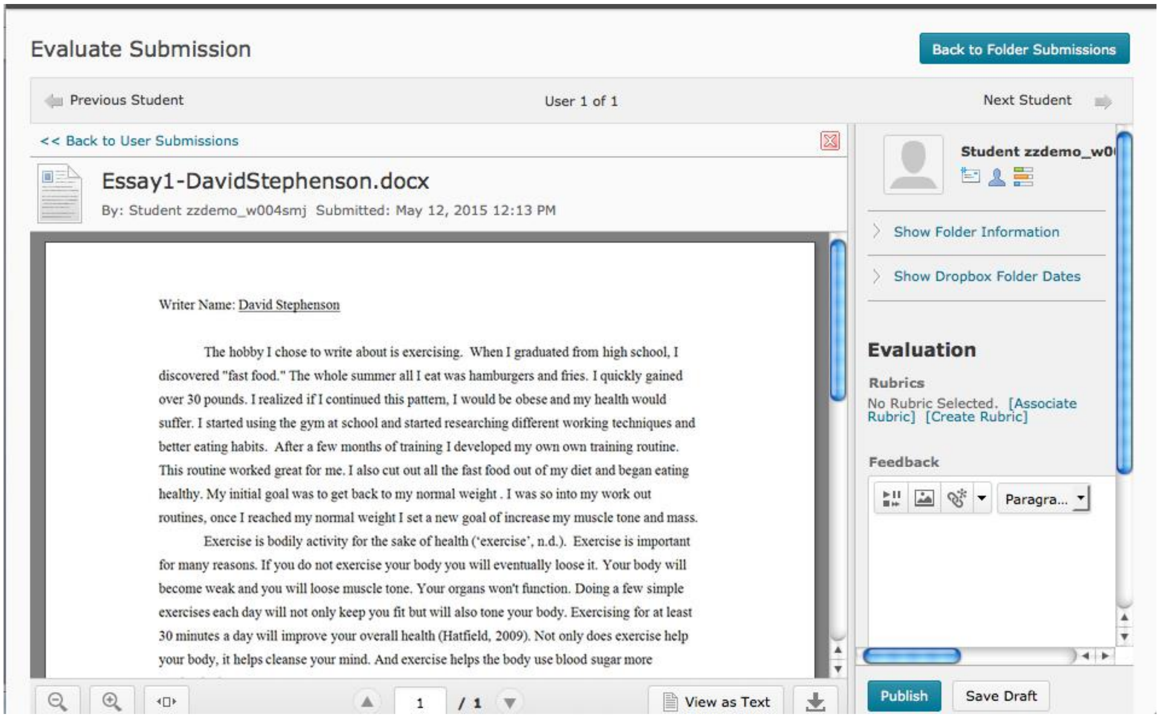
Evaluation

Rubrics

No Rubric Selected. [Associate Rubric] [Create Rubric]

Feedback

Figure A-2. Participants selected each essay, in order, from a list of six possible links (only the first two links were operational).



Essay1-DavidStephenson.docx

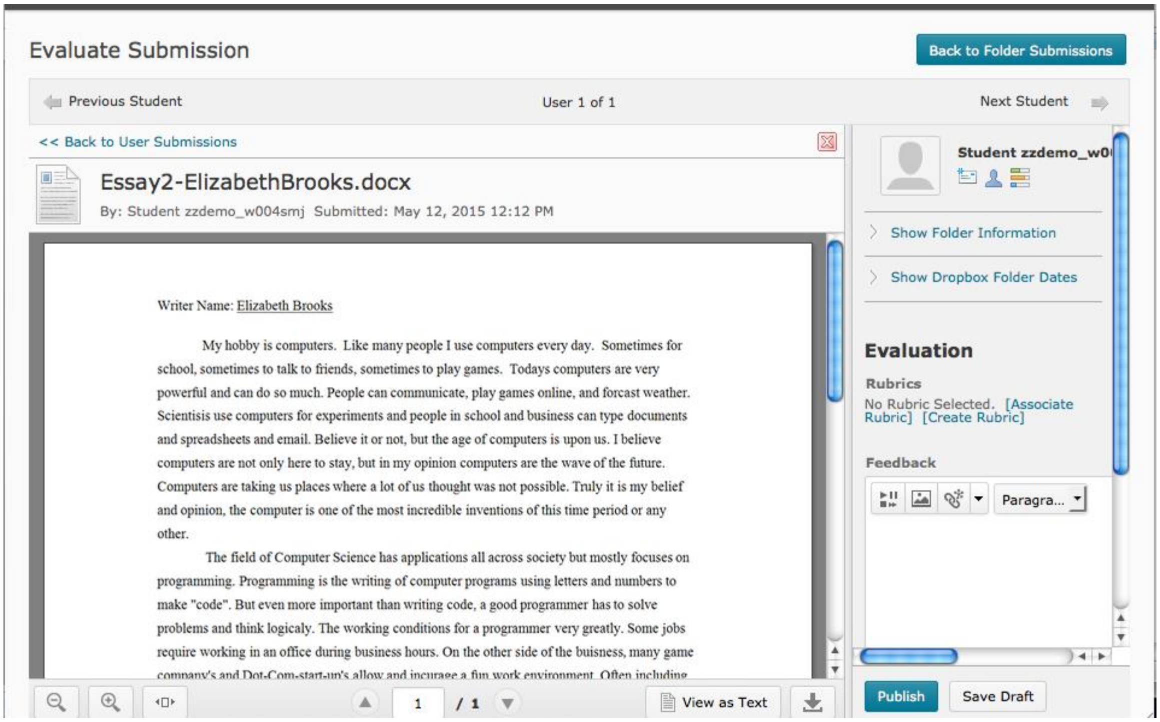
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:13 PM

Writer Name: David Stephenson

The hobby I chose to write about is exercising. When I graduated from high school, I discovered "fast food." The whole summer all I eat was hamburgers and fries. I quickly gained over 30 pounds. I realized if I continued this pattern, I would be obese and my health would suffer. I started using the gym at school and started researching different working techniques and better eating habits. After a few months of training I developed my own own training routine. This routine worked great for me. I also cut out all the fast food out of my diet and began eating healthy. My initial goal was to get back to my normal weight . I was so into my work out routines, once I reached my normal weight I set a new goal of increase my muscle tone and mass.

Exercise is bodily activity for the sake of health ('exercise', n.d.). Exercise is important for many reasons. If you do not exercise your body you will eventually loose it. Your body will become weak and you will loose muscle tone. Your organs won't function. Doing a few simple exercises each day will not only keep you fit but will also tone your body. Exercising for at least 30 minutes a day will improve your overall health (Hatfield, 2009). Not only does exercise help your body, it helps cleanse your mind. And exercise helps the body use blood sugar more

Figure A-3. Sample essay view (First essay: male author, exercise topic).



Essay2-ElizabethBrooks.docx

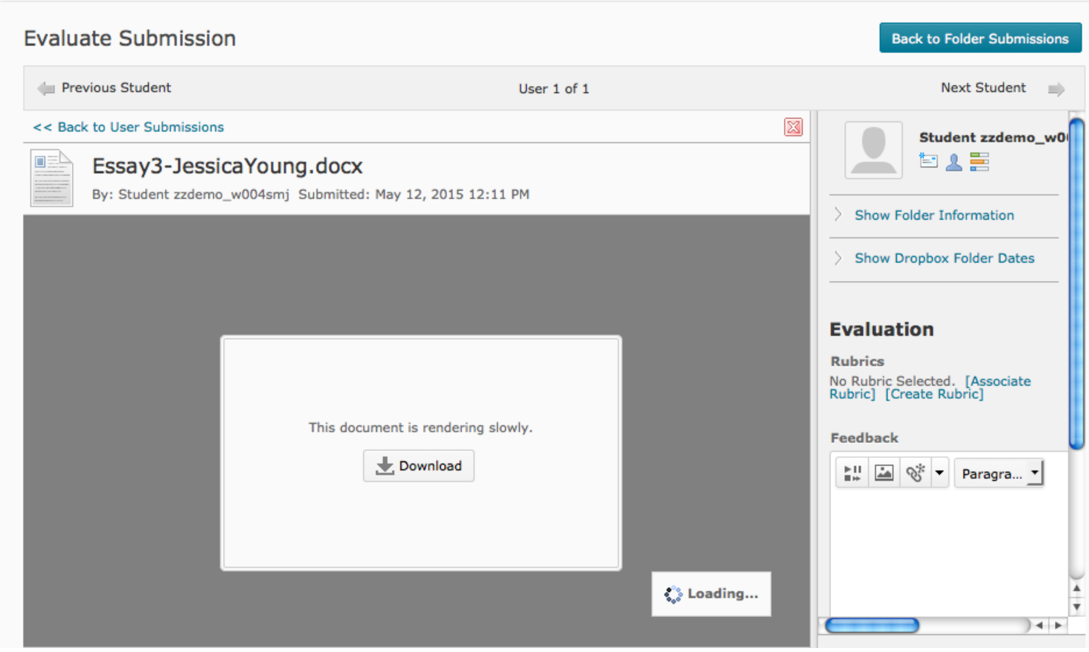
By: Student zzdemo_w004smj Submitted: May 12, 2015 12:12 PM

Writer Name: Elizabeth Brooks

My hobby is computers. Like many people I use computers every day. Sometimes for school, sometimes to talk to friends, sometimes to play games. Todays computers are very powerful and can do so much. People can communicate, play games online, and forcast weather. Scientisis use computers for experiments and people in school and business can type documents and spreadsheets and email. Believe it or not, but the age of computers is upon us. I believe computers are not only here to stay, but in my opinion computers are the wave of the future. Computers are taking us places where a lot of us thought was not possible. Truly it is my belief and opinion, the computer is one of the most incredible inventions of this time period or any other.

The field of Computer Science has applications all across society but mostly focuses on programming. Programming is the writing of computer programs using letters and numbers to make "code". But even more important than writing code, a good programmer has to solve problems and think logicaly. The working conditions for a programmer very greatly. Some jobs require working in an office during business hours. On the other side of the buisness, many game company's and Dot-Com-start-up's allow and incurage a fun work environment. Often including

Figure A-4. Sample essay view (Second essay: female author, computer topic).



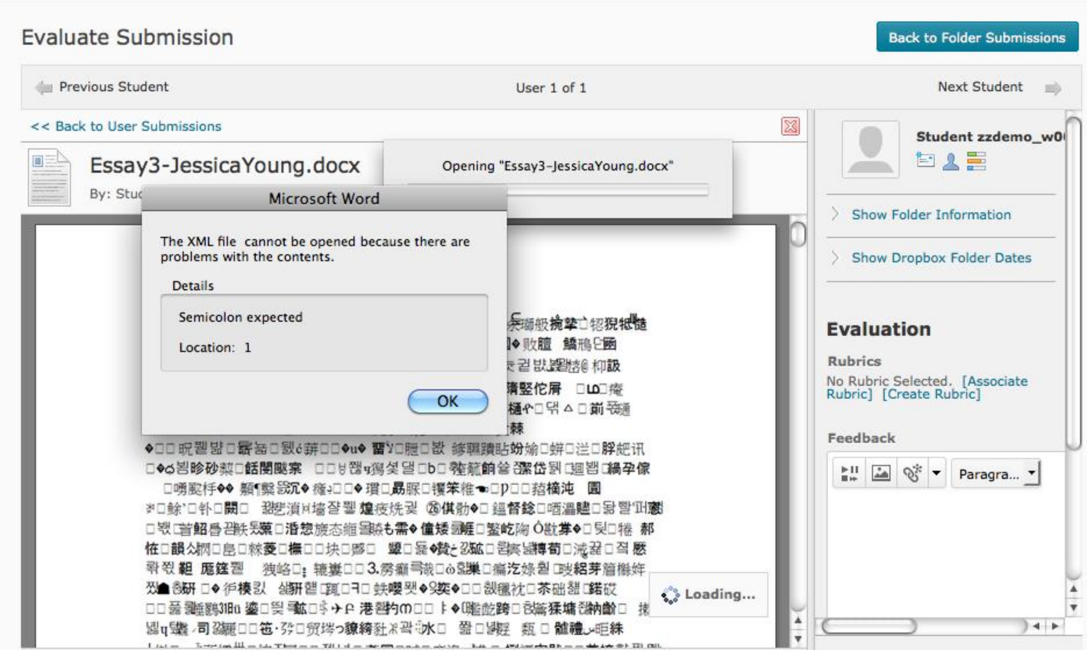
Essay3-JessicaYoung.docx

By: Student zzdemo_w004smj Submitted: May 12, 2015 12:11 PM

This document is rendering slowly.

No Rubric Selected. [Associate Rubric] [Create Rubric]

Figure A-5. Example of “technical error” when one of the non-operational essay links was clicked.



Essay3-JessicaYoung.docx

By: Stu

Opening "Essay3-JessicaYoung.docx"

The XML file cannot be opened because there are problems with the contents.

Details

Semicolon expected

Location: 1

No Rubric Selected. [Associate Rubric] [Create Rubric]

Figure A-6. Example of “technical error” when attempting to download a non-operational file.

Appendix B: Computer and Exercise Essays

Computer Essay:

My hobby is computers. Like many people I use computers every day. Sometimes for school, sometimes to talk to friends, sometimes to play games. Todays computers are very powerful and can do so much. People can chat, play games online, and check weather. Scientists use computers for studies and people in school and work can type documents and spreadsheets and email. Believe it or not, but the age of computers is upon us. I believe computers are not only here to stay, but in my opinion computers are the wave of the future. Computers are taking us places where a lot of us thought was not possible. Truly it is my belief and opinion, the computer is one of the most incredible inventions of this time period or any other.

The field of Computer Science has uses all across society but mostly focuses on programming. Programming is the writing of computer programs using letters and numbers to make "code". But even more important than writing code, a good programmer has to solve problems and think logicaly. The working conditions for a programmer very greatly. Some jobs require working in an office during business hours. On the other side of the buisness, many game company's and Dot-Com-start-up's allow and incurage a fun work environment. Often including toys, office sleep-in's and cold pizza laying across many a desk. Yet nomatter what the company they all involve the employee to stare at a monitor for endless hours and write the applications of tomorrow on a standard keyboard. There are disadvanges to working with computers. One being that you must risk eye damage with a computer screen every day (Wikipedia).

Opportunities in the computer field are very open to qualified personel (about.com). I have heard first-hand accounts of people being yanked out of collage for a programming position at $80,000 a year. With the job market for technology growing, comes the need for programmers of all backgrounds. Job-security is good as long as you dont kill somebody (witch recently happened at a dot-com-start-up). Comp. Sci. is becoming widely available in colleges and even High schools. Some things can not be taught and the person who wants to work with computers has to have some skills of your own. For example: the ability to solve problems and with logic. I know I can do well in a computer field because I really enjoy computers and I'm good at logic and math. I would love to be able to use my hobby in my career every day.

References

Computer vision syndrome . (2013). Wikipedia.
http://en.wikipedia.org/wiki/Computer_vision_syndrome

McKay, D.R. (2015). Computer Science Careers.
http://careerplanning.about.com/od/occupations/a/computercareers

Modified from existing essays available at http://www.essaypride.com/essays.php

Exercise Essay:

The hobby I chose to write about is exercising. When I graduated from high school, I discovered fast food. The whole summer all I ate was hamburgers and fries. I quickly gained over 30 pounds. I realized if I continued this pattern, I would be obese and my health would suffer. I started using the gym at school and started researching different working techniques and better eating habits. After a few months of training I developed my own own training routine. This routine worked fantastic for me. I also cut fast food out of my diet and began eating healthy. Once I reached my normal weight I set a new goal of increase my muscle tone and mass.

Exercise is bodily physical activity for the sake of health ('exercise', n.d.). Exercise is important because If you do not exercise your body you will eventually loose it. Your body will become weak and you will loose muscle tone. Your organs won't function. Doing a few simple exercises each day will not only keep you fit but will also tone your body. Overall health is improved by exercising at least 30 minutes a day (Hatfield, 2009). Not only does exercise help your body, but your mind is cleansed too. And exercise helps the body metabolize blood sugar more efficiently.

A career in physical therapy or personal training would be perfect for me. Many people in today's society are health conscientious. They know if you exercise you will be helping the body feel better and improve your health. Becoming and staying fit are very hard challenges that many people struggle with. If I were to find a job involving exercise, I could help people with their struggle. Bone density is lost when stay in bed at a hospital for long stretches, they can loose bone mass. As people age there bones become more frail, so they can break there hips and other bones easy. But bone mass can grow back with activity in addition to muscle. This is why physical therapy is so important. Physical therapists are in demand and pay a high salary well right off the bat (APTA, 2013). Training isn't as long as medical school, even though they also have to learn anatomical structures. This means I could start doing what I love even sooner. I could also consider becoming a personal trainer. I could work in a gym or health club or hospital or even at a university. Salary paid for personal trainers is lower, but education might be shorter which translates into financial savings. Both of these jobs can be physically demanding on my body, but I feel like, in my mind, I'm in good shape, so its still a good option for me.

References:

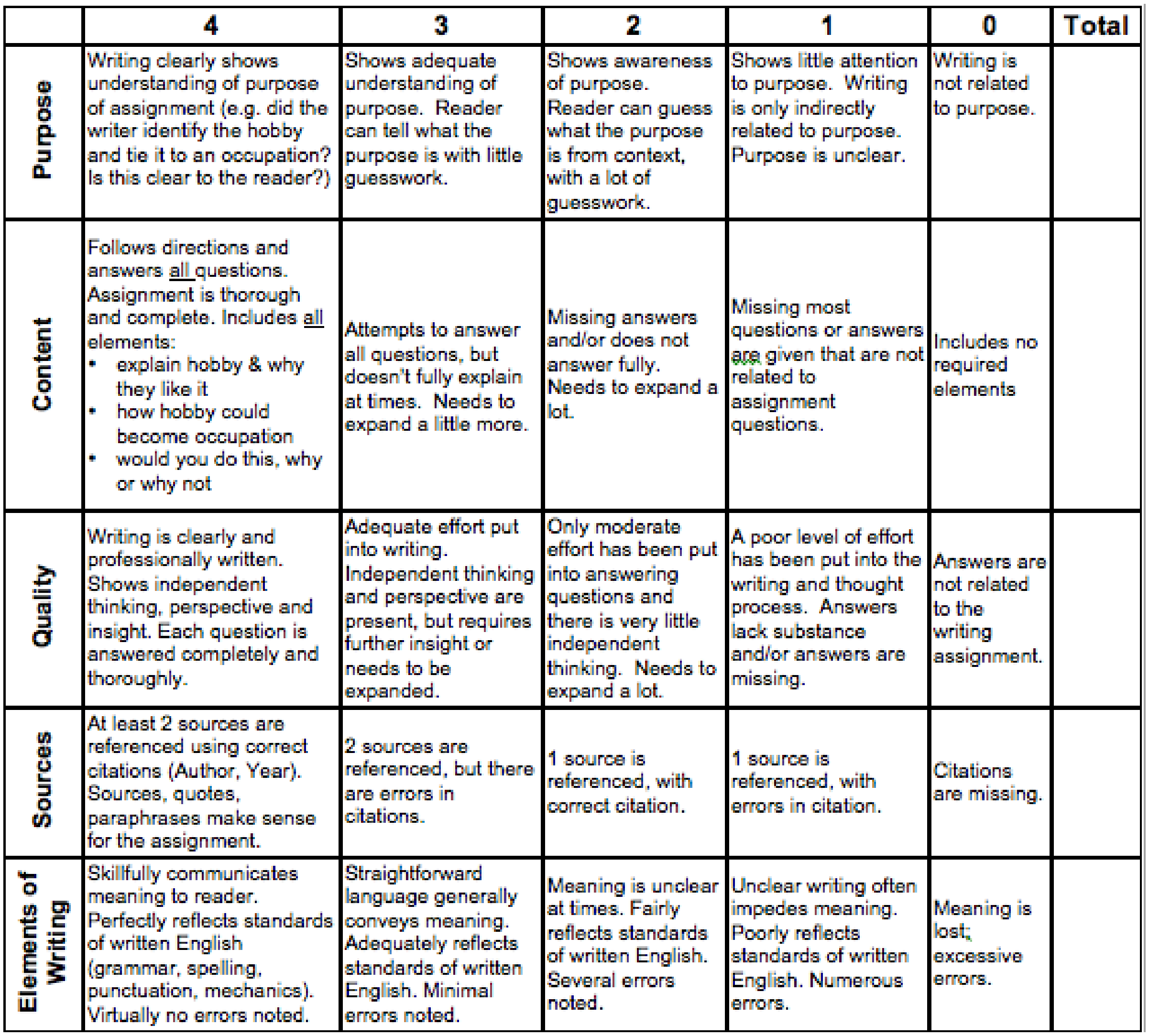
- APTA. (2013). Benefits of physical therapist career. *American Physical Therapy Association*. Retrieved from http://www.apta.org/PTCareers/Benefits/.
- exercise [Def. 3]. (n.d.). *Webster's Dictionary Online.* Retrieved from http://www.webster-dictionary.net/definition/exercise.
- Hatfield, H. (2009). Your exercise routine: How much is enough? *WebMD*. Retrieved from http://www.webmd.com/fitness-exercise/guide/your-exercise-routine-how-much-is-enough.

Modified from existing essays available at http://www.essaypride.com/essays.php

Appendix C: Rubric (grade sheet used in the Rubric condition)

After reviewing the writing prompt above, read and grade the writing activity you've been given. Below, fill in the name of the writer whose work you are grading, and the total points you are awarding to the assignment (out of 20 total possible points). Feel free to write anywhere on this grade sheet.

Writer's name: _______________________________(do NOT write your name here)



	4	3	2	1	0	Total
Purpose	Writing clearly shows understanding of purpose of assignment (e.g. did the writer identify the hobby and tie it to an occupation? Is this clear to the reader?)	Shows adequate understanding of purpose. Reader can tell what the purpose is with little guesswork.	Shows awareness of purpose. Reader can guess what the purpose is from context, with a lot of guesswork.	Shows little attention to purpose. Writing is only indirectly related to purpose. Purpose is unclear.	Writing is not related to purpose.	
Content	Follows directions and answers all questions. Assignment is thorough and complete. Includes all elements: • explain hobby & why they like it • how hobby could become occupation • would you do this, why or why not	Attempts to answer all questions, but doesn't fully explain at times. Needs to expand a little more.	Missing answers and/or does not answer fully. Needs to expand a lot.	Missing most questions or answers are given that are not related to assignment questions.	Includes no required elements	
Quality	Writing is clearly and professionally written. Shows independent thinking, perspective and insight. Each question is answered completely and thoroughly.	Adequate effort put into writing. Independent thinking and perspective are present, but requires further insight or needs to be expanded.	Only moderate effort has been put into answering questions and there is very little independent thinking. Needs to expand a lot.	A poor level of effort has been put into the writing and thought process. Answers lack substance and/or answers are missing.	Answers are not related to the writing assignment.	
Sources	At least 2 sources are referenced using correct citations (Author, Year). Sources, quotes, paraphrases make sense for the assignment.	2 sources are referenced, but there are errors in citations.	1 source is referenced, with correct citation.	1 source is referenced, with errors in citation.	Citations are missing.	
Elements of Writing	Skillfully communicates meaning to reader. Perfectly reflects standards of written English (grammar, spelling, punctuation, mechanics). Virtually no errors noted.	Straightforward language generally conveys meaning. Adequately reflects standards of written English. Minimal errors noted.	Meaning is unclear at times. Fairly reflects standards of written English. Several errors noted.	Unclear writing often impedes meaning. Poorly reflects standards of written English. Numerous errors.	Meaning is lost; excessive errors.	

Percent	**Points**
90% - 100%	18-20 points
80% - 89%	16-17 points
70% - 79%	14-15 points
60% - 69%	12-13 points
0% to 59%	less than 12 points

Grade: ______________ points out of 20

Rubric based on recommendations from Andrade (2005), Hitt & Helms (2009), and AACU (2015)

Appendix D: Grade Sheet (used in the No Rubric condition)

Writing Activity Prompt:
A hobby is any activity that people participate in on a regular basis for the purpose of enjoyment and leisure. There are indoor hobbies, outdoor hobbies, hobbies where things are collected or created, hobbies where things are observed, and hobbies where games are played. Identify a hobby that you enjoy, and explain what it is and why you like it. How could this hobby be turned into an occupation? Would you personally consider making this hobby into a life-long career? Why or why not?

Directions for Research Participant:
After reviewing the writing prompt above, read and grade the writing activity you've been given. Below, fill in the name of the writer whose work you are grading, the total points you are awarding to the assignment (out of 20 total possible points), and use the space below for notes if needed.

Writer's name: ______________________________(do NOT write your name here)

Percent	Points
90% - 100%	18-20 points
80% - 89%	16-17 points
70% - 79%	14-15 points
60% - 69%	12-13 points
0% to 59%	less than 12 points

Grade: ______________ points out of 20

Space for notes (if needed):

Appendix E: Instructions and Stimuli Used for Implicit Association Test

Instructions: In the next task, you will be presented with a set of words to classify into groups. This task requires that you classify items as quickly as you can while making as few mistakes as possible. Going too slow or making too many mistakes will result in an uninterpretable score. This part of the study will take about 5 minutes. The following is a list of category labels and the items that belong to each of those categories.

Category	Items
Male	Man, Boy, Father, Male, Grandpa, Husband, Son, Uncle
Female	Girl, Female, Aunt, Daughter, Wife, Woman, Mother, Grandma
Science	Biology, Physics, Chemistry, Math, Geology, Astronomy, Engineering
Liberal Arts	Philosophy, Humanities, Arts, Literature, English, Music, History

Keep in mind

- Keep your index fingers on the 'e' and 'i' keys to enable rapid response.
- Two labels at the top will tell you which words go with each key.
- Each word has a correct classification. Most of these are easy.
- Sort items by their category membership. Words in green should be categorized with the green labels. Words in white should be categorized with the white labels.
- The test gives no results if you go slow -- Please try to go as fast as possible.
- Expect to make a few mistakes because of going fast. That's OK.

Retrieved from http://www.projectimplicit.net/researchers.html

Appendix F: Modern Sexism Scale (Swim et al., 1995)

Please rate your agreement with the following statements by circling the appropriate number

(1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree).

	SD	**D**	**N**	**A**	**SA**
Discrimination against women is no longer a problem in the United States.	1	2	3	4	5
Women often miss out on good jobs due to sexual discrimination.	1	2	3	4	5
It is rare to see women treated in a sexist manner on television.	1	2	3	4	5
On average, people in our society treat husbands and wives equally.	1	2	3	4	5
Society has reached the point where women and men have equal opportunities for achievement.	1	2	3	4	5
It is easy to understand the anger of women’s groups in America.	1	2	3	4	5
It is easy to understand why women’s groups are still concerned about societal limitations of women’s opportunities.	1	2	3	4	5
Recently, the government and media have shown more concern about the treatment of women than is warranted by women’s	1	2	3	4	5

Appendix G: Women in STEM Stereotype Scale (Jackson et al., 2014)

Please rate your agreement with the following statements by circling the appropriate number (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree).

	SD	D	N	A	SA
Men are more interested in caring for their families than in advancing their careers.	1	2	3	4	5
Men are better at math than women.	1	2	3	4	5
Men are naturally more interested in science than women.	1	2	3	4	5
I prefer male professors more than female professors.	1	2	3	4	5
Men spend more time doing laboratory research than women.	1	2	3	4	5
Only the best professors get promoted, regardless of gender.	1	2	3	4	5
Men are more interested in humanities or liberal arts than women.	1	2	3	4	5
There are fewer women faculty in science because they are less qualified.	1	2	3	4	5
I prefer female professors more than male professors.	1	2	3	4	5
Women are more interested in family than in their careers.	1	2	3	4	5
There are fewer women faculty in science because they are not interested in these fields.	1	2	3	4	5
Women and men are equally good at math.	1	2	3	4	5
Men publish more science research articles than women.	1	2	3	4	5
Compared to men, women are equally qualified to hold positions in science fields.	1	2	3	4	5

Appendix H: Semantic Differential Scale Items

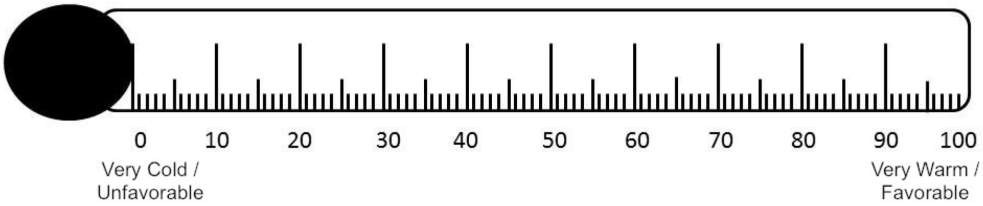
Please rate the following items by circling the number that is closest to your belief.

I think women are:

Analytical	1	2	3	4	5	Emotional
Hostile	1	2	3	4	5	Congenial
Foolish	1	2	3	4	5	Wise
Bad	1	2	3	4	5	Good
Stupid	1	2	3	4	5	Smart
Pleasant	1	2	3	4	5	Unpleasant
Passive	1	2	3	4	5	Assertive
Logical	1	2	3	4	5	Irrational
Favorable	1	2	3	4	5	Unfavorable
Incompetent	1	2	3	4	5	Competent
Committed	1	2	3	4	5	Indifferent
Lazy	1	2	3	4	5	Hard-working

Appendix I: Feeling Thermometer

Please rate each of the following items using the feeling thermometer below. You may use any number from 0 to 100 for a rating. Ratings between 50 and 100 represent a favorable feeling and ratings between 0 and 50 represent an unfavorable feeling.



_______ women

_______ male faculty

_______ female scientists

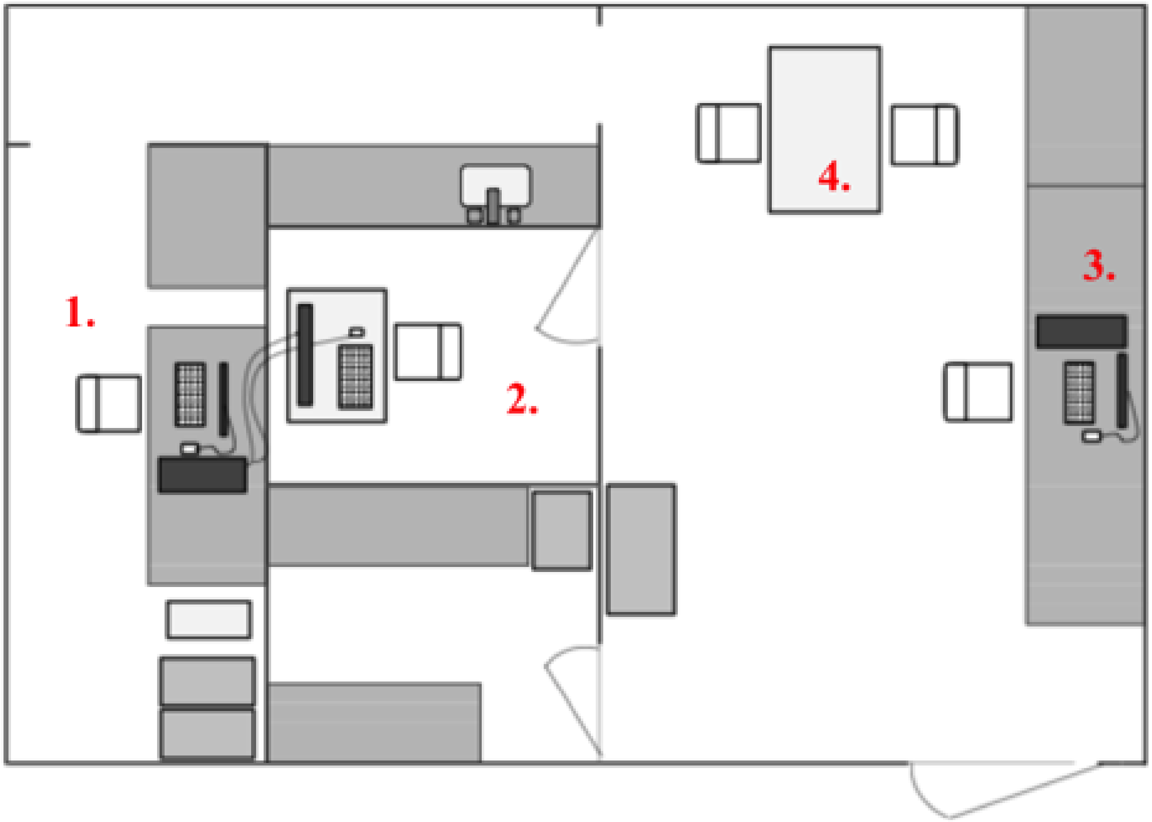
_______ men

_______ male scientists

_______ female faculty

Based on Michigan Feeling Thermometer (Wilcox, Sigelman, & Cook, 1989)

Appendix J: Experimental Laboratory Layout



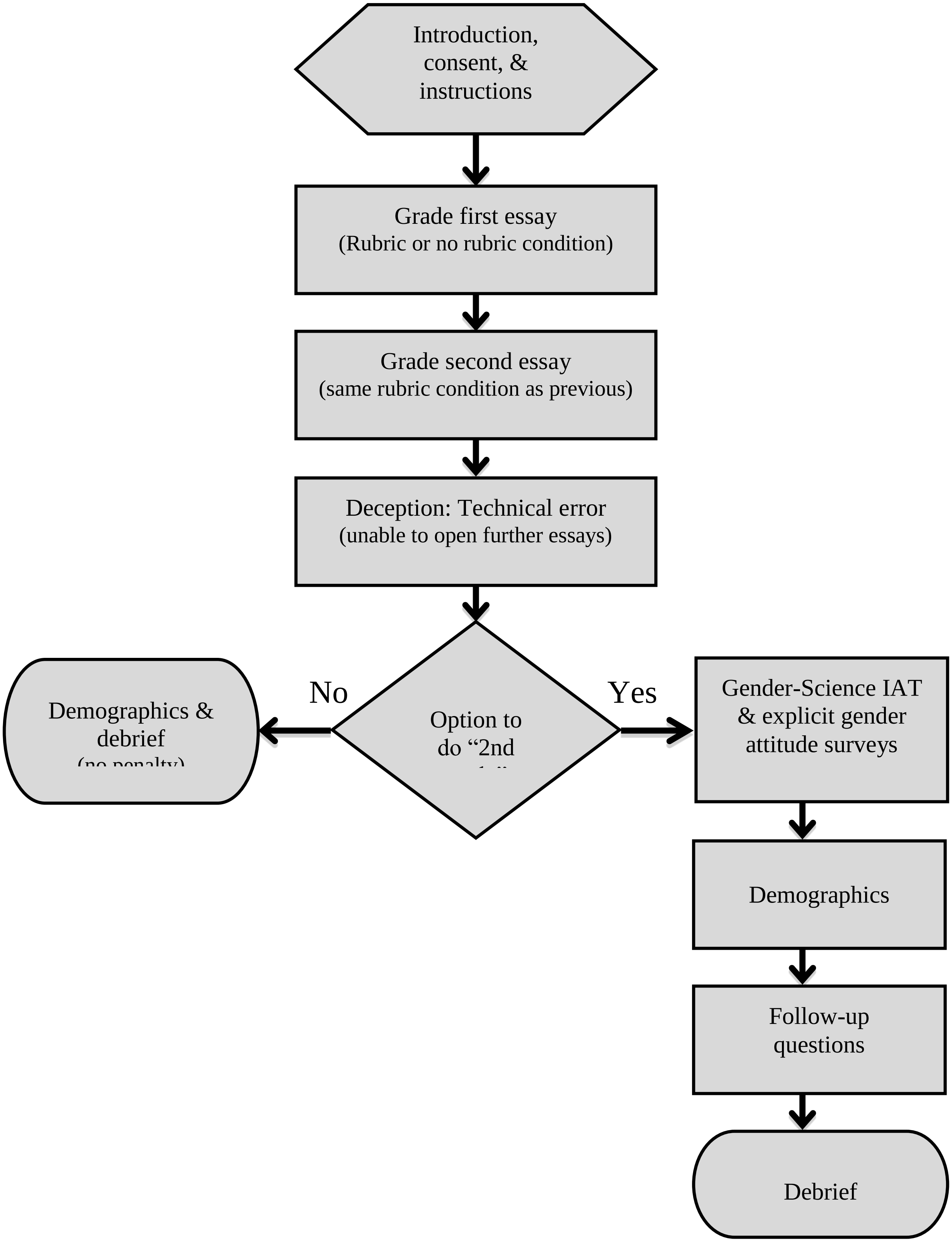
1. Prior to retrieving the participant, the researcher set up the LMS system on the primary computer in the back room and logged in to the online course.

2. Participants entered the lab front door and were escorted to an experiment room where they were seated at a computer.
    a. Here they reviewed the consent information, received instructions, and read and graded the two experimental essays.

3. After completing the essays and encountering the "technical difficulties", participants were given the option of participating in a second, ostensibly unrelated study. If they agreed, the researcher led them to a second computer in a different part of the lab.
    a. Here they completed the Gender-Science IAT, explicit measures, and demographics survey.

4. The researcher debriefed the participant regarding the second study, then asked follow-up questions regarding the first study. Participants were debriefed, thanked for their time, and escorted to the lab door.

Appendix K: Procedural Flowchart


Introduction, consent, & instructions
Grade first essay
(Rubric or no rubric condition)
Grade second essay
(same rubric condition as previous)
Deception: Technical error
(unable to open further essays)
Option to do “2nd
No
Demographics & debrief
(no penalty)
Yes
Gender-Science IAT & explicit gender attitude surveys
Demographics
Follow-up questions
Debrief