FACTORS THAT INFLUENCE PEER GRADING

Mary Moeller

A capstone project submitted in partial fulfillment
of graduating from the Academic Honors Program
at Ashland University

May 2015

Faculty Mentor: Dr. Diane B. V. Bonfiglio, Professor of Psychology
Additional Reader: Dr. Chris Chartier, Assistant Professor of Psychology
Abstract
This study examined whether subject matter, essayist gender, and essayist level of education influenced grading of an essay. In order to examine this, 70 participants were randomly assigned to read an essay with either a stereotypically female topic (about an author’s collected works) or a stereotypically male topic (about politics in sports). Within those conditions, participants were also randomly assigned to be told either that the essay was written by a male essayist or female essayist. Finally, participants were randomly assigned to be told either that the essay was written by a high school student or a graduate student. Participants were then asked to grade the essay they were given according to the SAT grading rubric, which was printed out and provided for them. The hypothesis was that participants would rate the essay most highly when the stereotypical gender topic was the same as the essayist’s gender and they were told the essayist was in graduate school. The results of this study did not support the hypothesis. There appeared to be no interaction effects between gender topic of the essay, gender of the essayist, and education level of the essayist ($F(1, 102) = 0.026; p = 0.871$). However, there was a main effect of education level on the rating of grammar in the essay ($F(1, 102) = 9.099; p = 0.003$), which supports the myriad of research done in this area. This paper explores the implications and possible future directions of this research.
# Table of Contents

- **Introduction** .................................................................................................................. 1
- **Methods** .......................................................................................................................... 11
- **Results** ............................................................................................................................ 13
- **Discussion** ....................................................................................................................... 15
- **Appendix** ........................................................................................................................ 30
  - **Figures** .......................................................................................................................... 30
  - **Materials** ..................................................................................................................... 38
- **Author's Biography** ......................................................................................................... 48
List of Tables and Graphs

Figure 1: Estimated Marginal Means of Rubric Total at Male Essayist…………………….30
Figure 2: Estimated Marginal Means of Rubric Total at Female Essayist ………………….31
Figure 3: Estimated Marginal Means of Rubric Total at Male Essay Topic…………………32
Figure 4: Estimated Marginal Means of Rubric Total at Female Essay Topic ……………..33
Figure 5: Estimated Marginal Means of Rubric Total at High School Education Level…....34
Figure 6: Estimated Marginal Means of Rubric Total at Grad School Education Level……35
Figure 7: Estimated Marginal Means of Grammar Score at Male Essay Topic……………..36
Figure 8: Estimated Marginal Means of Grammar Score at Female Essay Topic ………….37
Table 1: Number of Correct Vs. Incorrect Manipulation Check Education Responses ……38
Factors that Influence Peer Grading

A stereotype is best defined as a fixed, over-generalized belief about a particular group or class of people (Cardwell, 1996). Stereotypes are a pervasive threat in American culture as they can have detrimental effects on interpersonal relationships as well as cause problems internally for individuals (Baumeister & Showers, 1986; Galdi, Cadinu, & Tomasett, 2014; Matheus, 2010; etc.). Americans have been aware of them for decades, and yet, as the research presented in this paper will show, it seems that all of the knowledge gleaned from research has not been fully able to counteract their negative effects.

Stereotypes often come from people noticing traits that seem to be shared among a specific, defined group of people, and there is evidence to suggest that the traits most likely to develop into stereotypes are those that can be used easily to distinguish one defined group from another (Ford & Stangor, 1992). The human mind naturally creates schemas based on previous experience and information in order to process new information more quickly, and these schemas, especially those related to gender, often have stereotypic foundations and easily become assimilated into a culture (Bem, 1981). Once a stereotype is formed, it is incredibly difficult to rid a culture of it, especially if the stereotype suggests a negative trait about a group (Rothbart & Park, 1986). Certain types of people are more likely to agree with stereotypes and expect them to remain true than others; for example, a person who believes that one’s personality is fixed and does not significantly change over the course of his or her life is more likely to agree with the stereotype of a group (Levy, Stroessner, & Dweck, 1998). Research also indicates that if a person believes in a stereotype against one minority group, he or she is also more likely to believe stereotypes presented against other minority groups as well (Crandall, 1994).

Stereotypes can be problematic and potentially dangerous for the stereotyped population
when acted upon explicitly. The Holocaust was sustained, in part, because of the spread of negative stereotypes against Jewish people during that time, and the people that did survive tended to be those who did not match the stereotypes for Jewish physical appearance (Suedfeld, Paterson, Soriano, & Zvic, 2002). However, the dangers of stereotypes can occur in a more individualized context as well. Jurors fall prey to stereotypes and biases when they decide whether a defendant is guilty or innocent. They tend to make a decision of “guilty” with a much higher amount of confidence when the defendant is African American compared to when the defendant is Caucasian (Kleider, Knuycky, & Cavrak, 2012). Additionally, when on trial, defendants that have more stereotypical African American features are sentenced more harshly than African Americans who do not have as many of these features (Blair, Judd, & Chapleau, 2004). As the research shows, the effects of stereotypes can develop both consciously and subconsciously. Not all of the effects, however, come from other people. Stereotypes can cause problems internally as well.

**Stereotypes and Gender**

Stereotypes within the education system especially have had a large impact on American society. For example, females are commonly seen as less competent in math than males, and it has been found that even academically gifted females tend to drop out of science, technology, engineering, and math (STEM) programs more often than males do (Johnsen & Kendrick, 2005). Although some research suggests that the belief in gender differences between subjects in schools are no longer an issue (Spelke, 2005), more recent research suggests that students in many areas throughout the Western world are not only aware of these stereotypes but affected by them every day. For example, a French survey of schoolchildren of different ages found that the stereotype of females, being better at reading than males, was prevalent in all tested age groups
This stereotype is also found in implicit cognitions, which are unconscious factors (e.g. knowledge or memories) that influence a person’s behaviors without him or her being aware of the influence (Greenwald & Banaji, 1995). Women tend to show a stronger negativity toward math, produced by the stereotype that math is an implicitly masculine skill (Nosek & Smith, 2011). These implicit stereotypes correlated with less confidence in math, worse achievement on math tests, and less participation during classes. Another survey done in America found that men and women were not only explicitly aware of these stereotypes, but it also supported that the amount of confidence they felt in different subjects fell directly in line with the stereotypes (Gilbert, O’Brien, Garcia, & Marx, 2014). Women in the study felt a greater amount of confidence in English classes and felt much less confidence in math classes, while men were exactly the opposite. These disparities in confidence occurred even though a meta-analysis of gender research showed that the academic gender stereotypes so ingrained in Western culture both appear to be declining and seem not to have been as definitive as it was once thought (Linn & Hyde, 1989).

Gender stereotypes are experienced cross-culturally (Shafiq, 2001). An analysis of school records in Middle Eastern countries (Azerbaijan, Indonesia, Jordan, the Kyrgyz Republic, Qatar, Tunisia, and Turkey, specifically) found that women in those countries were outperforming men in reading, but that men were outperforming women in mathematics. A similar study performed in South Africa produced similar results: women greatly outperformed men in reading skills, and the authors of the article became interested in the topic due to a noticeable gender-based literacy gap in the country (Zuze & Reddy, 2014). A cross-cultural study of socialization in one hundred ten different cultures worldwide showed that while differentiation of the genders is not significant during infancy, in every culture studied there was cultural pressure for children to fall
into what each culture considered stereotypically male and female for their culture (Barry III, Bacon, & Child, 1957).

The global nature of these studies may suggest that stereotypes might not be the cause of the discrepancies in skills between men and women, that the difference is due perhaps to something biological or implicit. However, research has shown that this is not the case. In their South African literacy study, Zuze and Reddy (2014) also explored some of the factors surrounding reading skill acquirement. They determined that accessibility of reading materials and human resources were both factors that influenced literacy in South African schoolchildren. In addition, they found that human resources appeared either to be made more available to girls or girls were more able to utilize their resources to practice reading. Furthermore, the problem of stereotypes being passed down from teachers to students arises within the school systems. For example, if teachers are implicitly aware of the stereotype that women are better readers than men, they can unintentionally portray that stereotype in their classrooms and inadvertently pass it down to their students (Retelsdorf, Schwartz, & Asbrock, 2014). According to this study, this passing down of stereotypes occurs not only between teachers and students, but also possibly between a child and any adult that child would consider significant in his or her life. It is also important to note that in this study, the positive stereotype towards women and reading did not significantly improve the scores the girls received on tests.

The harm that stereotypes can cause are not always merely academic, and often they are so ingrained within a culture that it is hard to pick them out. For example, Americans tend to find men more trustworthy than women to report on sports, especially if the person rating the trustworthiness of the newscaster is not well-versed in sports him- or herself (Harris, 2013). Additionally, referees, regardless of expertise in their sport, have powerful biases against women
Stereotypes can also affect the occupations chosen by individuals (Matheus, 2010), as well as how well others believe those individuals are at doing their jobs (Bauer & Baltes, 2002). They may also have political ramifications: when running for office, male and female candidates must focus on portraying different traits, those associated with gender stereotypes, in order to earn more votes (Lee, 2014). Schneider and Bos (2014) found that females in politics can sometimes be seen as lacking in the traits stereotypically attributed to women, but that they also are not seen as having any masculine traits either. This is especially detrimental to their political careers, as men and women alike appear to prefer having a man in a leadership position (Rudman & Kiliianski, 2010). This might be because men seem consistently to demonstrate higher levels of knowledge about politics than women do (Gidengil, Giles, & Thomas, 2008; Mondak & Anderson, 2004). Preliminary research suggests that this stereotype may give men more confidence in their political abilities (Pruysers & Blais, 2014), but that women are seen as less competent and more easily deceived than males (Kray, Kennedy, & Van Zant, 2014). Ditonto, Hamilton, and Redlawsk (2013) supported this finding by showing that participants changed what information they looked for based on the gender of a political candidate, searching for more competence-related information about female candidates than they did for male candidates.

If someone is made aware of a negative stereotype about a group to which he or she belongs, all hope is not lost. While there is not yet one definitive, universal way to eliminate stereotype threat, it can be reduced (e.g. Jones, 2011; Weger, Hooper, Meier, & Hopthrow, 2012). As detrimental and powerful as these effects seem, psychologists have found ways to reduce their influence on test scores by either not reminding the participants of the stereotype before they are tested or by exercises in self-affirmation if the stereotype is known (Jones, 2011).
Specifically for education-related gender stereotypes, it can be helpful to portray intelligence as malleable, especially if the person saying it is malleable is someone respected by the student (Good, Aronson, & Inzlicht, 2003). One other way of eliminating the consequences of negative stereotypes is to reframe the threat subtly as a challenge (Alter, Aronson, Darley, Rodriguez, & Ruble, 2010). As the research shows, it is important to target stereotypes and their effects in order to counteract them. It is possible to lessen and, in some cases, even eliminate the detrimental effects of stereotypes, but first there must be research to uncover the stereotype. That is what this present study hopes to accomplish.

**Detrimental Effects of Stereotypes**

It will be helpful for the purposes of this study to talk about the specific detrimental effects of stereotypes, some of which were mentioned briefly earlier in this literature review. One major danger that arises from stereotypes is the stereotype threat. Steele and Aronson (1995) define stereotype threat as “being at risk of confirming, as self-characteristic, a negative stereotype about one’s group” (797). A wide array of researchers have worked to extricate the exact reasons stereotypes are so disparaging, and they have shown that there are many reasons for the threat to be so influential. One reason is anxiety over the threat. If a person is made aware of a negative stereotype about a group he or she belongs to, that person may feel so worried about proving the stereotype correct that it takes away from the ability to process the information he or she is being tested over, which leads to a worse score than if he or she had never been made aware of the stereotype (Baumeister & Showers, 1986). Another reason stereotype threat can be so potent is that it gives the person being tested a reluctance to respond. The person becomes concerned that no matter how he or she answers the test questions, it will never be seen as good enough to defeat the stereotype (Lord & Saenz, 1985). Along these same
lines, there is also a danger of the stereotyped person becoming distracted by the threat or by their own self-consciousness over the threat (Wine, 1971).

This effect is seen in gender research: girls tend to perform worse on math tests after being made aware of the stereotype against them (Galdi, Cadinu, & Tomasetto, 2014; Jones, 2011). The threat’s influence remains even after girls are explicitly warned about the stereotype threat and told that it is this threat, and not physical differences, that contribute to lower test scores (Tomasetto & Appolini, 2012). Not only are test scores affected, but knowing about the stereotype can also have a negative influence on the amount learned in a lesson about the subject one believes he or she is predisposed against (Boucher, Rydell, Van Loo, & Rydell, 2012; Taylor & Walton, 2011).

Additionally, the threat has also been found in race research when African Americans were seen to perform worse on tests when they felt the test would produce results truly representative of their abilities, and therefore put them at risk of confirming the stereotype of African Americans being intellectually inferior to Caucasian Americans (Steele & Aronson, 1995). A disturbing aspect of this threat became known very recently: it is possible for these threats to have an effect even when the person being threatened by the stereotype is not consciously aware that the stereotype exists (Galdi, Cadinu, & Tomasetto, 2014). Fortunately, there has been as much research done about how to prevent or lessen the effects of stereotype threat as there has been done on the stereotypes themselves.

It is important that as much as possible is learned about stereotypes, as many studies have shown that the stereotype threat is both a very real and very detrimental contemporary issue (Galdi, Cadinu, & Tomasetto, 2014). Because these effects are so widespread and cause issues both during tests and during real-world situations, it is imperative that research on stereotypes
continue so that psychologists can find newer and better ways to counteract them. (Temple & Neumann, 2014).

**Priming and Rating Research as it Applies to This Study**

Stereotypes are one specific part of a large group of stimuli that can influence perceptions and actions (Bargh, Chen, & Burrows, 1996). Certain stimuli, when presented, activate knowledge structures that have been experimentally observed and found to have a subconscious effect on a subsequent behavior (Bargh & Pietromonaco, 1982). This phenomenon is known as priming, and can be seen with stereotypes as well as with other stimuli (Bargh, 1982). For example, in one study, participants primed with hostility rated an ambiguously aggressive person in a story as more hostile than those not primed with hostility. For the most part, people do not seem to be aware of how certain phrases or environments may influence their perceptions or opinions (Nisbett & Wilson, 1977). Specifically, according to Nisbett and Wilson’s (1977) study, participants are occasionally not aware that there is a stimulus that greatly influenced their response and, in some cases, not even aware of their response.

In relation to the current study, research has shown that when thinking consciously after being made aware of a stereotype or having a stereotype activated, people tend to rely on the stereotypes when forming an impression of someone, and even suppressed or ignored information that was incongruent with the stereotype they had learned (Bos & Dijksterhuis, 2011). Incidences that confirm the stereotype are more often remembered, meaning that traits a group is stereotyped to have are much more easily confirmed than proven wrong (Biernat & Ma, 2005). The present study takes this idea of unconscious influences and tries to pinpoint the effects they will have on a participant’s rating of an essay.

To speak more specifically about rating, one theory suggests that it is dependent upon
three components: the performance of the person being rated, the observations of the rater, and the ability of the rater to recall those observations (Wherry & Bartlett, 1982). The aspects that this study will manipulate to influence the participants’ ratings are those that the rater will observe about the essayist. There is an important relationship between the perceived source and how good of a rating is given (Bauer & Baltes, 2002; Jones, Paull, & Erskine, 2002; Wimer & Beins, 2008). This is especially clear in humor research. Participants tend to rate jokes as funnier when they think the joke was told by a famous comedian than when they believe the joke came from a non-comedian celebrity (Johnson & Mistry, 2013). These effects are so powerful that they can even manifest physically. When participants tried different foods that all looked and smelled the same and had the same texture, they rated the foods as tasting the same even though there were small variations in the level of saltiness each possessed (Dijksterhuis, Boucon, & Le Berre, 2013). This effect also manifests itself in sports. Soccer referees were more likely to award red and yellow cards to players on a team they had been warned was aggressive than they were when they were told nothing about a team’s level of aggression (Jones, Paull, & Erskine, 2002).

These rating biases are, not surprisingly, also influenced by gender stereotypes. A study of college students found that, when given vignettes about the performances of male and female professors, those with traditional stereotypes evaluated the “female” professors less accurately and more negatively than they evaluated the “male” professors with the same characteristics (Bauer & Baltes, 2002). Additionally, the stereotype that men excel more in math and women in reading and literature affects teachers as well as students, as was shown in a study by Simpson (1974) in which teachers were asked if they thought a man or woman would be better at teaching certain subjects. Female teachers were rated as more competent in teaching language arts and male teachers were rated as more competent in teaching math and science. When a member of a
stereotyped group has traits that do not fit those that match the stereotype, the person’s traits can be reinterpreted to match that stereotype (Hamilton & Rose, 1980). For example, when observing the behaviors of children, adults were more likely to characterize a behavior as aggressive when said behavior was performed by a male versus a female child (Fagot & Hagan, 1985).

This bias is seen outside of an experimental setting in hiring practices. When given the choice between a woman or man to hire to complete a math- or science-related job, hiring managers were twice as likely to hire the man even when no other information besides gender was given (Reuben, Sapienza, & Zingales, 2013). These results held firm whether or not the manager was told how well the job candidate performed on an arithmetic task. Additionally, when recruiting potential employees in a city, managers tend to avoid areas with a high concentration of low-income African Americans (Neckerman & Kirschenman, 1991). When the managers used a skills-based test rather than location and interviews to hire, however, there was a much higher number of African Americans hired. These biases hold true when the stereotype is weight as well (Larkin & Pines, 1979). Overweight people were significantly less likely to be hired than a candidate of average weight with identical qualifications.

Just as being informed about a stereotype does not negate the risk of stereotype threat, being informed of rating biases does not dispel them. In fact, when people are made aware of the risk of rating bias, they tend to rate too far in the other direction (Jorgensen, Backstrom, & Bjorklund, 2013). For example, when exposed to the possibility of being biased, people tended to consider it highly important to avoid overrating white men and underrating people who were not white men, and often ended up overcompensating to avoid doing so. As the research shows, rating biases occur unconsciously in many different aspects of life.

This present study focuses on the gender stereotypes of females having a stronger
intellect for reading and language arts and males having a stronger intellect for sports and politics. The research looks to discover if participants will rate an essay differently if it is about a stereotypically male topic (sports politics) or stereotypically female topic (literature), if they believe the essayist is a man or a woman, and if they believe the essayist is in high school or graduate school. Based on the research, the hypothesis is that participants will rate the essay most highly when the stereotypical gender topic is the same as the essayist’s gender and they are told the essayist is in graduate school.

**Methods**

**Participants**

There were 103 participants recruited from the Ashland University Psychology 101 research pool. The participants signed up using the Ashland University Sona system. All participants were at least 18 years of age, with the mean age being 18.96 with a standard deviation of 1.48. The participants were primarily Caucasian (88.3%) and approximately 66.1% were female. They were compensated with partial course credit for Psychology 101 upon completion of the study. The protocols in this study were approved by the Ashland University Human Subjects Review Board and were determined to be safe and appropriate for use with human subjects.

**Materials**

The materials used in this study were a demographics questionnaire, two essays, one about a stereotypically female topic and one about a stereotypically male topic, short vignettes about the essayists, and a grading rubric for the essays. The demographics questionnaire was a brief survey that gleaned information on race, gender, age, and class rank for possible use in analysis (see Appendix J). The female essay was about the collective works of Thomas E.
Kennedy since language arts is a stereotypically feminine area, as is supported by research completed by Simpson (1974) and Thomas (1994). This essay (Appendix L) was taken from an online undergraduate journal and only a page of it was used so that participants had adequate time to read and grade it (Brenna, 1997). The male essay was about the politics of a professional football team (Okolo, 2014), as both sports and politics are stereotypically masculine topics (Bauer, 2013; Rudman & Killianski, 2000; Bernstein, 2002). Only the first page of this essay was used in order to match it in length with the female essay and to give participants adequate time to read and grade it (see Appendix M). The grading rubric that participants used was taken from the SAT website and is the one administrators use to grade essays for the SAT (SAT Grading Rubric, 2014). It consists of six grading categories and allows participants to circle a number from 1 (the worst) to 6 (the best). The categories have participants focus on the entire essay in general, the development of the essay, its organization, the vocabulary used, the sentence structure, and the grammar. Participants may circle a number from 1 to 6 in each of these categories, for a total possible score of 36. The full rubric and the instructions given to participants can be found in Appendices H and K.

**Procedure**

Participants arrived and completed the study in sets of ten or fewer and were individually assigned to each of the variables by the use of a random number generator. They began by signing a consent form as soon as they entered. Once participants were placed in a condition, they completed the demographics questionnaire. Next, brief essays were given out to the groups. Each participant received either a stereotypically female essay or a stereotypically male essay and a short vignette about the author. The vignettes (Appendix K) informed the participant that the essay was written by Brittany Lewis for the female condition or Matthew Lewis for the male
condition. The vignettes also told participants what level of education Brittany or Matthew had, saying that the essayist was either a senior in high school or a graduate student. Participants were then asked to read the essay they were given and to grade it according to the SAT grading rubric, which was printed out and provided for them. This grading rubric was the same for all conditions. Participants were then given a two-question manipulation check (Appendix O) to be sure that they were aware of the gender and education level of their essayist. They were debriefed, and per section 8.08C of the APA Ethical Principles of Psychologists and Code of Conduct, because this study involved deception, participants were informed at the conclusion of their participation of the nature of the deception and were given the option of withdrawing their data if they so desired. They were then allowed to leave.

In order to determine if the topic of the essay, the gender of the essayist, or the education level of the essayist had any effect on the overall score participants gave the essay, I conducted a 2 (topic) by 2 (essayist gender) by 2 (education level) ANOVA. The dependent variable was the total score participants gave following the rubric. Alpha was set at 0.05 for this analysis.

Results

There appeared to be no interaction effects between gender topic of the essay, gender of the essayist, and education level of the essayist. The overall results of this experiment did not approach significance ($F(1, 102) = 0.781; p = 0.621$). There were also no significant main effects of any of the three independent variables on the rubric total, including Essay Topic ($F(1, 102) = 0.30; p = 0.836$), Essayist Gender $F(1, 102) = 0.678; p = 0.511$, and Education: $F(1, 102) = 2.310; p = 0.133$. The score participants awarded the essay’s grammar was analyzed with the three independent variables, however, and there was a main effect of education level on the rating of grammar in the essay ($F(1, 102) = 9.099; p= 0.003$). The mean rating on grammar (out
of a total possible rating of 6 for that category) given to high school essayists (N=55) was 4.6545 with a standard deviation of 1.02231. The mean rating given to graduate school essayists (N=48) was 5.2083 with a standard deviation of 0.74258. That being said, there were also no significant interaction effects reported from this study, including Education*Essay Topic ($F (1, 102) = 0.87; p = .769$), Essay Topic * Essayist Gender ($F (1, 102) = .052; p = 0.820$), Education * Essayist Gender ($F (1, 102) = 1.048; p = .310$), and Essayist Gender * Essay Topic * Education ($F (1, 102) = 0.026; p = 0.871$). Graphs of these data can be found in Appendices A through H.

During initial analyses, it became clear that more participants than expected were answering incorrectly on the manipulation check. This spurred the creation of the “correctness” variable for analysis in order to help researchers understand possible reasons behind this surprising amount of incorrect answers. When correctness was analyzed as a separate independent variable, however, the overall results were not significant ($F (1,102) =0.601; p=0.857$). Removing participants who did not answer correctly did not change the results, $F (1, 102) = 0.591; p = 0.762$.

In order to see if any of the variables had an effect on whether or not a participant answered the manipulation check correctly, a chi-square analysis was run on the different independent variables. It was found that the only significant effect on whether a participant correctly answered the manipulation check was that of education level of the essayist, $X^2 (1) = 17.199, N = 103; p < 0.001$). Participants were significantly more likely to answer the education manipulation check incorrectly when the essayist was in the graduate school condition compared to the high school condition. A table of these data can be found in Appendix I.

When the incorrect answers were more closely examined, they showed that, out of the 38 participants who answered incorrectly, 25 of them (65.79%) answered in a way that goes along
with the previous research, meaning that they either chose the gender associated with the topic of the essay or they answered the gender of the essayist correctly but chose a lower education level for the essayist if it was the opposite gender topic. When a chi-square analysis of this number versus the expected random chance number of 19 out of the 38 was performed, the results approached but did not reach significance ($X^2(1, N=38) = 3.789; p = 0.052$). While the numbers did not quite reach significance, it seems that descriptively there was a trend in that more participants were choosing either the gender that corresponded with the essay topic or, if they chose the correct gender, they were choosing a lower education level when the gender of the essayist did not match the stereotyped essay topic.

**Discussion**

**General Discussion**

While the expected results of this study were not significant, some interesting data arose during the analysis. First, it was shown that if participants were told that the essayist had a higher level of education (i.e. graduate school), then they would rate the grammar of the essay as better no matter the gender of the essayist or the topic of the essay. This finding falls in line with some other rating bias research, especially that which manipulates information about the source of the item being rated. For example, Johnson and Mistry’s 2013 study showed that participants were more likely to rate a joke as funny if they believed it came from a professional comedian compared to if they believed it came from a non-comedian famous person. There are many other examples of knowledge about the perceived source influencing a rating (Bauer & Baltes, 2002; Wimer and Beins, 2008; etc.). Because grammar is something that must be taught, normally in a classroom setting, it is not surprising that this variable was more effected by education levels than the other, more intuitive parameters on the rubric.
There are some potential reasons why the other expected results did not occur. First, it is possible that the gender of the participants played a role in how susceptible they were to the stereotypes and how they rated the essays. For example, gender played a role in student evaluations of teaching effectiveness (Young, Rush, & Shaw, 2009). Students rated instructors on a variety of interpersonal characteristics, course content characteristics, and pedagogical characteristics. For the latter two, students rated teachers of the same gender as them significantly higher than they rated teachers of the opposite gender. Bachen, McLoughlin, and Garcia (1999) found similar results in that female students in college rated female faculty significantly higher across five categories than they rated male faculty. A further analysis of their results showed that this discrepancy in rating was influenced by the strength of the students’ gender schemas instead of being due solely to a preference for certain teaching styles. Based on this research, the fact that the gender of the essayist was given in this study could have led the female participants to rate the essay as higher when it was written by a woman than they would have had it been written by a man.

Another reason the unexpected results could have arisen is that the task was primarily literature-based. While the topics of the essays were stereotypically catered to different genders, the male participants in the study could have not taken the task as seriously simply because it required reading and grading an essay. Men tend to find reading less stimulating than women do, and are more likely to view the task less positively and find less connection with it (Epting, Rand, & D’Antuono, 2014). This is not to say that all the male participants did not take the reading task seriously; however, if the men in the study did not view reading and grading the essay as an important task, or if they could not connect with it while completing it, this could have impacted the results.
The personality characteristics of the participants also could have influenced the results. Their levels of agreeableness and assertiveness could have influenced their ratings, as those two traits are found to correlate with higher levels of inaccuracy in rating (Bernardin, Cooke, & Villanova, 2000). Specifically, when a rater has high amounts of warm-heartedness and sensitivity, both of which are considered sub-attributes of agreeableness, these traits have been shown to be positively correlated with high rating levels (Bartells & Doverspike, 1997). In regards to peer rating, participants who are very agreeable tend to rate their peers more positively (Wood, Harms, & Vazire, 2010). In addition, high levels of conscientiousness are found to be negatively correlated with ratings, meaning that the more conscientious a participant is, the lower a rating he or she will likely give (Bernardin, Tyler, & Villanova, 2009). Based on this research, the individual personalities of the participants may have influenced the rating they gave the essays, regardless of the perceived gender and education level of the essayist.

One of the most surprising results of this study was the fact that so many participants answered the manipulation check incorrectly. Participants were given no verbal instruction on grading the essays, and therefore they were forced to read the vignettes in order to find out what to do, yet over a third of them did not answer the manipulation check correctly. This could be because participants read one of the essays and fell into the stereotype by assuming that a woman wrote the language arts essay and a man wrote the sports essay. This goes along with the research done by Hamilton and Rose (1980) that showed that if a member of a stereotyped group has traits that do not match the stereotype, people can reinterpret those traits so that they do fit the stereotype. While the research has more support for this trend if the traits are based on gender and not the other way around (Fagot & Hagan, 1985), it is possible that some participants believed the stereotype strongly enough that it changed their perception of the essayist’s gender.
Another surprising result that arose from this study was that participants were significantly more likely to get the education level wrong on the manipulation check when they were told the essayist was in graduate school than when the essayist was in high school. While there is not a lot of previous research that would give insight as to why this may have occurred, one possible explanation is that participants automatically assumed that they would not be asked to grade something written by someone above their level of education, and so they chose the lower level on the manipulation check. Another possible reason for this occurrence is that participants, especially those who were freshmen, might have recognized the SAT Grading Rubric from when they took the SAT and inferred that the essayist was in high school from the fact that only people in high school take the SAT.

Limitations

There are several possible limitations to this study. General limitations focus mainly on the population in which this study occurred. It took place at a small, private, middle-class college in Northern Ohio that has very little diversity. Therefore, the results might not be representative of the general population; however, many psychological studies are completed using similar demographics and have achieved significant results that do apply to more diverse populations (e.g. Gilbert, O’Brien, Garcia, & Marx, 2014; Hively & El-Alayli, 2013; etc.). Additionally, the participants were all students in a General Psychology course and were required to participate in studies for course credit. This could mean that some participants did not take the experiment seriously or were not motivated to try very hard on it. Along this same vein, some participants might find the prospect of reading and editing an essay to be boring or difficult and therefore did not try their hardest on the task.
There are also limitations that are more specific to the experiment performed. For example, seeing as 36.89% of participants answered incorrectly on the manipulation check, it could be that giving vignettes about the essayists was not an explicit enough way to make participants aware of the essayist’s gender and education level. In addition, it is possible that participants, knowing they were in an experiment, were more cautious of deception and therefore did not believe the information given to them about the essayist. There could also have been a reactivity bias present in the study, as Ashland University is a small university and it is possible that previous participants told future participants what to expect before coming to the study.

**Future Directions**

There are a number of future directions this study could take. Because such a large number of participants answered the manipulation check incorrectly, it would be interesting to delve into the reasons that may have occurred. For example, a future study could give participants an essay and have participants then guess the gender of the essayist. Further, a study could also be done that tells participants the gender of the essayist but has them guess the education level. These two manipulations might help uncover the reasons behind the incorrect answers in this original study. Additionally, a study could be done where the gender and education level of the essayist were more explicitly stated instead of being part of vignettes. For example, the researcher could verbally inform the participant that the essayist was a man or a women and was in high school or graduate school and then have the participant repeat the information so he or she is forced to acknowledge it.

Another future direction this study could take would be to test participants for their awareness of the stereotypes being experimented with before performing the study. This would help in seeing how big of a threat stereotypes pose to peer grading. Related to this, participants
could be primed for stereotypes in order to see the effect this would have on how they graded the essays. This would be similar to Graham and Lowery’s (2004) experiment in which they primed police officers for racial stereotypes before having them read vignettes about a hypothetical offender. Those primed for racial stereotypes rated the hypothetical offender significantly more negatively and as more culpable than those who were not primed. It would be interesting to see if these results would hold true in a more academic setting and with peers rather than potential criminals.

Conclusion

While the expected interactions did not occur, several aspects of this study warrant further exploration in future research endeavors. The fact that grammar was significantly affected by education level suggests that perhaps education has more of an influence on certain parameters than it does on others. Additionally, the number of participants who answered incorrectly on the manipulation check suggests that the stereotypical gender topic of the essay might have influenced how they answered the questions. Both areas need to be researched more and in different ways, but the results gleaned from this study lend some support to current research and open doors for future exploration.
References


SAT Grading Rubric. (2014). [Table of SAT writing section scoring guidelines for 2014]. *Data from The College Board SAT website*. Retrieved from [http://www.carrollcc.edu/assets/forms/PTA/SAT%20Writing%20Rubric.pdf](http://www.carrollcc.edu/assets/forms/PTA/SAT%20Writing%20Rubric.pdf)


status gains and losses. *Sociology of Education, 47*(3), 388-398


Appendix
Figure 1: Estimated marginal means of rubric total at essayist gender = male. This graph shows that there were no main or interaction effects on rubric total between the three independent variables.
Figure 2: Estimated marginal means of rubric total at essayist gender= female. This graph shows that there were no main or interaction effects of the three independent variables on the rubric total.
Appendix C

Figure 3: Estimated marginal means of rubric total at male essay topic. This graph shows that there were no main or interaction effects of the three independent variables on the rubric total.
Figure 4: Estimated marginal means of rubric total at female essay topic. This graph shows that there were no main or interaction effects of the three independent variables on the rubric total.
Figure 5: Estimated marginal means of rubric total at essayist education level equals high school.
This graph shows that there was no main effect of the three independent variables on the rubric total. While there appears to have been an interaction effect, it was not statistically significant.
Table 6: Estimated marginal means of rubric total at graduate school level of essayist education. This graph shows that there was no main effect of the three independent variables on the rubric total. While there appears to have been an interaction effect, it was not statistically significant.
Figure 7: Estimated marginal means of Grammar at Essay Gender = Male. This graph shows that there is a main effect of education level on the mean rating given to grammar.
Table 8: Estimated marginal means of Grammar at Essay Topic = Female. This graph shows that there was a significant main effect of education level on grammar scores of the essays.
Appendix I: This table shows the total number of participants who answered correctly and incorrectly for education level on the manipulation check collapsed over the education variable. The chi-square analysis of these data show that there was a significant effect of perceived education level on whether or not a participant answered this question correctly.

<table>
<thead>
<tr>
<th>Education Level of Essayist</th>
<th>Number Correct</th>
<th>Number Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graduate School</td>
<td>23</td>
<td>25</td>
</tr>
<tr>
<td>High School</td>
<td>42</td>
<td>6</td>
</tr>
</tbody>
</table>

Appendix J:

Demographics Questionnaire

1. What is your class rank?
   - Freshman
   - Sophomore
   - Junior
   - Senior

2. What is your gender?
   - Male
   - Female

3. How old are you? __________

4. What is your ethnicity?
   - Caucasian
   - Black
   - Asian
   - Latino
   - Hispanic
   - Multiracial

   Other (please specify) __________
Appendix K:

A.

Essay Grading Instructions

This essay was written by Brittany Lewis as part of her Graduate School Program. Read the essay and grade it using the attached rubric. Circle the box on the rubric that denotes which score you are giving the essay for that particular category. You may take as much time with this as you need. Ask the researcher if you have any questions.

B.

Essay Grading Instructions

This essay was written by Matthew Lewis as part of his Graduate School Program. Read the essay and grade it using the attached rubric. Circle the box on the rubric that denotes which score you are giving the essay for that particular category. You may take as much time with this as you need. Ask the researcher if you have any questions.

C.

Essay Grading Instructions

This essay was written by Brittany Lewis for one of her Senior Year High School classes. Read the essay and grade it using the attached rubric. Circle the box on the rubric that denotes which score you are giving the essay for that particular category. You may take as much time with this as you need. Ask the researcher if you have any questions.
D.

Essay Grading Instructions

This essay was written by Matthew Lewis for one of his Senior Year High School classes. Read the essay and grade it using the attached rubric. Circle the box on the rubric that denotes which score you are giving the essay for that particular category. You may take as much time with this as you need. Ask the researcher if you have any questions.
Appendix L

Stereotyped Essay- Female

Alienation, Imagination, Secrets and Sins in Thomas E. Kennedy

With the publication of a third novel, *The Book of Angels* (Wordcraft of Oregon, 1997), and a second short story collection, *Drive, Dive, Dance & Fight* (BkMk Press, 1997), Thomas E. Kennedy confirms his reputation as a consummate writer and storyteller. Kennedy has been writing from Copenhagen, Denmark for the past 23 years, sending his stories, novels, poems, criticisms, interviews and articles across the Atlantic. His place in American Letters has grown exponentially in the nineties, a decade in which it seems he has come to literary prominence without the help of “Big House” publications, where today’s best-selling novel is often recycled from yesterday’s hack work pumped out for readers who like fiction to be as plot-driven and predictable as television sit-coms.

*Drive, Dive, Dance & Fight’s* eight stories deal with self-tormented, self-exiled men and women. Surrounded by friends or family or co-workers, the message is still the same, though played out in different forms in each story—we are essentially isolated, essentially alone, ever condemned to contest against the loneliness and fear and conformity which society and conscience feed us.

In the first story, “Bonner’s Women,” Bonner sees a former lover in a bar and feels embarrassed by the memory of the intimacy the two once shared. He finds himself ashamed and wishing that the affair had never happened, that he had stayed “home with his wife and children where he should have been.” The former lover has aged and so has Bonner and there is nothing left of the experience except his self-castigating regret that he ever met her in the first place. Seeing her again forces Bonner to remember what the affair had made of him:

*You should hear what they say
About you: cheat cheat . . .*

He recalls the religion he has abandoned and its moral teachings and he thinks of his father kneeling beside a bed at night, saying prayers that Bonner no longer believes in. He remembers how effortlessly his former lover had lied to her husband and how eventually Bonner himself couldn’t tell when she was lying to him, and how one day it hit him that he would never again know what truth was. From that moment on, there was nothing left but a huge dread and the knowledge that the affair would be inside him for the rest of his life, “a smudge” that he could never erase.

“Bonner’s Women” is a moralist’s tale, an exploration of guilt and what guilt does to us and how it controls the perspectives we have and how it seals us in an unforgiving morality. The hidden, secret world inside our minds binds us and makes us condemn ourselves endlessly. Bonner will never get over his failures, his indiscretions, and there will never be any beauty or romance again in his life.

The story closes on his preoccupation with his mother’s and his own inevitable aging and death. He looks out his window on a winter morning at the snow falling, the flakes clicking against the panes and he envisions his children playing in the white “virgin snow.” They will one day follow him to the edge of forever and beyond, just as he in turn will follow his mother. His
children’s lives filled with mistakes and burdens of guilt and spiritual failures will mimic his own. It is the story of every generation condemned to death at the moment of birth and weighed down through life by baggage carried forward from the past.

Appendix M

Stereotyped Essay- Male

Who Owns the Redskins?
In 1932, a money-minded showman named George Preston Marshall bought a team called the Boston Braves in the young, developing National Football League (NFL). His second head coach hired in his second year of ownership—the team began the tradition of new blood under the headset each season much longer ago than many believe—was a man named William “Lone Star” Dietz, reportedly of Sioux descent. Most NFL coaches today are known for their no-nonsense demeanor on game days; Dietz was encouraged by Marshall to adorn himself in a headdress and war paint before his team’s home contests. Marshall was so fascinated by Native American pageantry that in 1933 he decided to honor his coach by renaming his franchise the Boston “Redskins.” He maintained that parallelism was involved as well; the team was playing at Fenway Park at the time, home of the Boston Red Sox.

The less-reported fact, however, is that Marshall was perhaps the most bigoted figure in national sports at the time. By the year 1961, the Washington Redskins were the only NFL team that had yet to sign an African-American player. Marshall didn’t sidestep around the fact: “We’ll start signing Negroes when the Harlem Globetrotters start signing whites.”

Though Marshall is considered the father of “football as entertainment” because of his innovational ideas in the sport—splitting the league into two division whose respective champions would compete for a national championship, introducing the glitzy halftime shows we view as commonplace today—his refusal to integrate cannot be ignored. It was not until the federal government threatened retribution and blacks picketed and boycotted games that Marshall used his number one draft pick to select the first African-American Heisman trophy winner, Ernie Davis.

The current fifteen-year owner of the Washington Redskins, Dan Snyder, is one of the NFL’s billionaires, his fortune primarily built off his lucrative franchise. But facing recent calls to change his team name out of respect for the team’s Native American fans and Native Americans everywhere, Snyder wrote in a October 2013 letter: “On that inaugural Redskins team, four players and our Head Coach were Native Americans. The name was never a label. It was, and continues to be, a badge of honor…We are Redskins Nation and we owe it to our fans and coaches and players, past and present, to preserve that heritage.”

The history of the Washington Redskins coupled with the power of the word “Redskin” has fueled lawsuit and activist campaigns since the name was officially registered in 1967. The calls for change began in earnest over twenty years ago, when famed Native American activist Suzan Shown Harjo filed a lawsuit against the Redskins.

Snyder invokes team history in his defense of the name, but when this history is revealed to be one of racism and stereotype, that claim becomes harder to defend. Still to be explored, however, is the effect a name change would have on team identity or the financial losses that could come from bestowing a new team name. Many franchises have changed their names or
removed offensive iconography with little or no effect on team success—but none have done so as a result of being forced to by protests, the league commissioner, or the Supreme Court; all avenues that have been explored in the effort to retire the word “Redskin.”

For many Native Americans, “Redskin” is a racial slur that won’t leave their mouths. In speaking with the HPR, Joel Barkin, spokesman for the Oneida Nation, actively avoided using the term. The Oneida Nation, a Native American tribe based in New York, has spent the past few years engaged in an educational “Change the Mascot” campaign, urging the NFL to force Snyder’s hand by raising awareness about the power of the team name. In late January, the Onedia Nation took their campaign farther by soliciting the United Nations, believing the situation had entered the realm of human rights. “The fact of the matter is, many people in this country have their first and sometimes only interactions with Native Americans through professional sports,” explained Barkin. “In this case, through a nine-billion-dollar industry that is promoting a racial slur.”
## Essay Grading Rubric

<table>
<thead>
<tr>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>General Category Parameters</strong></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>An essay in this category demonstrates clear and consistent mastery, although it may have a few minor errors. A typical essay:</td>
</tr>
<tr>
<td>5</td>
<td>An essay in this category demonstrates reasonably consistent mastery, although it will have occasional errors or lapses in quality. A typical essay:</td>
</tr>
<tr>
<td>4</td>
<td>An essay in this category demonstrates adequate mastery, and is marked by ONE OR MORE of the following weaknesses:</td>
</tr>
<tr>
<td>3</td>
<td>An essay in this category demonstrates developing mastery, and is marked by ONE OR MORE of the following weaknesses:</td>
</tr>
<tr>
<td>2</td>
<td>An essay in this category demonstrates little mastery, and is flawed by ONE OR MORE of the following weaknesses:</td>
</tr>
<tr>
<td>1</td>
<td>An essay in this category demonstrates very little or no mastery, and is severely flawed by ONE OR MORE of the following weaknesses:</td>
</tr>
</tbody>
</table>

**Development:**
- **Point of view, critical thinking, reasoning, and support**
  - Effectively and insightfully develops a point of view on the issue and demonstrates outstanding critical thinking, using clearly appropriate examples, reasons, and other evidence to support its position.
  - Effectively develops a point of view on the issue and demonstrates strong critical thinking, generally using appropriate examples, reasons, and other evidence to support its position.
  - Develops a point of view on the issue and demonstrates competent critical thinking, using adequate examples, reasons, and other evidence to support its position.
  - Develops a point of view on the issue, demonstrating some critical thinking, but may do so inconsistently or use inadequate examples, reasons, or other evidence to support its position.
  - Develops a point of view on the issue that is vague or seriously limited, and demonstrates weak critical thinking, providing inappropriate or insufficient examples, reasons, or other evidence to support its position.
  - Develops no viable point of view on the issue, or provides little or no evidence to support its position.

**Organization**
- Is well organized and clearly focused, demonstrating clear coherence and smooth progression of ideas.
- Is generally organized and focused, demonstrating some coherence and progression of ideas.
- Is limited in its organization or focus, or may demonstrate some lapses in coherence or progression of ideas.
- Is poorly organized and/or focused, or demonstrates serious problems with coherence or progression of ideas.
- Is disorganized or unfocused, resulting in a disjointed or incoherent essay.

**Vocabulary**
- Exhibits skillful use of language, using a varied, accurate, and apt vocabulary.
- Exhibits facility in the use of language, using appropriate vocabulary.
- Exhibits adequate but inconsistent facility in the use of language, using generally appropriate vocabulary.
- Displays developing facility in the use of language, but sometimes uses weak vocabulary or inappropriate word choice.
- Displays very little facility in the use of language, using very limited vocabulary or incorrect word choice.
- Displays fundamental errors in vocabulary.

**Sentence Structure**
- Demonstrates meaningful variety in sentence structure.
- Demonstrates variety in sentence structure.
- Demonstrates some variety in sentence structure.
- Lacks variety or demonstrates problems in sentence structure.
- Demonstrates frequent problems in sentence structure.
- Demonstrates severe flaws in sentence structure.

**Grammar**
- Is free of most errors in grammar, usage, and mechanics.
- Is generally free of most errors in grammar, usage, and mechanics.
- Has some errors in grammar, usage, and mechanics.
- Contains an accumulation of errors in grammar, usage, and mechanics.
- Contains errors in grammar, usage, and mechanics so serious that meaning is somewhat obscured.
- Contains pervasive errors in grammar, usage, or mechanics that persistently interfere with meaning.
Appendix O: Manipulation Check

Questionnaire 3

1. What level of education did the person who wrote the essay you read have?
   - High School
   - Undergraduate
   - Graduate
   - Unknown

2. What was the gender of the person who wrote your essay?
   - Female
   - Male
Appendix P

CONSENT FORM
Factors Affecting Peer Grading

A. PURPOSE AND BACKGROUND

In this study, we are investigating the relationship between existing attributions and perceived talent in peers. If you choose to participate, you will be asked to read a brief essay, grade the essay, and complete a series of surveys. Participation is expected to last approximately 25 minutes.

B. PROCEDURES

You will be asked to read a brief essay, grade the essay according to a rubric, and complete two short surveys.

C. RISKS/DISCOMFORTS

The risks associated with this study are no greater than those you may encounter in everyday life. All information and data from this study will be kept confidential and will be stored in a locked office and accessible only to study personnel.

D. BENEFITS

There will be no direct benefit to you from participating in this study other than helping to further research in the field of attributions and peer perception. Participation in this study will also earn you one research credit for a Psychology 101 course at Ashland University.

E. COSTS

There will be no costs to you as a result of taking part in this study.

F. QUESTIONS

You may email the researcher, Mary Moeller, at mmoeller@ashland.edu with questions or concerns about this study. If you have any comments or concerns about participation in this study, you should first talk with the researcher. If for some reason you do not wish to do this, you may contact the Human Subjects Review Board, which is concerned with the protection of volunteers in research projects. The HSRB can be reached by email (hsrb-au@ashland.edu) or telephone at (419)-289-5750.

G. CONSENT

If you wish, you may keep a copy of this consent form. Participation in this study is voluntary. You are free to decline to be in this study, or to withdraw from it at any point. Your decision as to whether or not to participate in this study will have no influence on your present or future status as a student or employee at Ashland University. If you agree to participate, you should sign below.

______________________________________________________________
Date                                                                 Signature of Study Participant
TO: Mary Moeller and Dr. Diane Bonfiglio  
FROM: Carol Reece, HSRB Chair  
DATE: September 23, 2014  
SUBJECT: Human Subjects Review Board Approval  
PROJECT TITLE: The Effects of Gender and Experience Level on Peer Grading  
HSRB APPROVAL CODE: 09-08-14-#013

The Human Subjects Review Board has approved your research study. You may proceed with the study as you have outlined in your proposal. The approval is granted for one calendar year. Research participant interaction and/or data collection is to cease at this time, unless application for extension has been submitted and approval for continuance is obtained.

The primary role of the HSRB is to ensure the protection of human research participants. As a result of this mandate, we ask that you adhere to the ethical principles of autonomy, justice, and beneficence. We would also like to remind you of your responsibility to report any violation to participant protections immediately upon discovery. Likewise, we would like to remind you that any alteration to the research proposal as it was approved cannot move forward. Any amendment to the application must be submitted for approval before the project can resume.

We wish you success in your discoveries,

Carol S. Reece DNP, APRN, CPNP  
Ashland University  
Chair Human Subjects Review Board
Author’s Biography

Mary Moeller grew up in Independence, Ohio and graduated from Independence High School in 2011. At Ashland University, she majored in Psychology and minored in Creative Writing and Spanish. She is a member of Alpha Lambda Delta, Psi Chi, and received both the Outstanding Sophomore Award for the Psychology Department and the Senior Academic Honors Award for the Psychology Department. She made the Dean’s List every semester at Ashland, and served as an Executive Chair on AU G.I.V.S. Upon graduation, Mary plans to attend Bowling Green State University’s Clinical Psychology program in pursuit of her Ph. D.