Predicted Data Analysis: The Impacts of Implicit Bias on Evaluations of Graduate Student Applications

Barrett Groves Dr. Lisa Willis Ronak Patel

August 13, 2021

ABSTRACT

Despite universities attempting to adopt more EDI (equity, diversity, and inclusion) friendly policies and increasing encouragement for women and marginalized groups to pursue STEM, there continues to be underrepresentation of these individuals in STEM fields and tenure-track faculty positions. This has led to research exploring the impacts of implicit bias on the advancement of marginalized groups in STEM. This research will be focused on how the implicit biases of an evaluator can impact the evaluation of a science graduate student applicant. As the preliminary trial and final experiment has not yet taken place, this paper will be using fake data sets to show how the data from this experiment will be analyzed. The data is partially inspired from similar past research and will display relevant concerns in applicant evaluation. This paper will explore the potential impacts of gender bias, racial bias, and intersectionality on graduate student applicants, while explaining the methods and expectations of an experiment designed to evaluate how sharing personal information on applications can affect the evaluation of the applicant.

INTRODUCTION

When there is a lack of diversity in STEM-related research, mistakes can be made. Studies have shown that a lack of women and marginalized ethnic groups in clinical trials can lead to drugs that are not applicable or potentially harmful to these underrepresented groups (Rouan et al., 2021). A drug designed to treat angina had resulted in lower heart rates in eldery women and women experienced adverse reactions when this drug was taken with a potential 26 other medications, likely due to the lack of female representation in clinical trials (Rouan et al., 2021). The test dummy most commonly used by motor vehicle companies when running safety tests is modeled after a man with an athletic build, based on the individuals who designed it (Ryan et al., 2020). However, a recent study with data collected from 1975 to 2010 has shown that the likelihood for a belt-restrained female driver to sustain severe injuries was 47% higher than for a belt-restrained male driver in the same crash circumstances (Ryan et al., 2020). The anatomical differences between men and women were not taken into consideration when designing the test dummy, and although a female test dummy now exists, it does not accurately represent female anthropometry (Ryan et al., 2020). Therefore, women are more likely to be injured in a car crash as they are excluded from safety tests, and the car is assumed to be safe for women if it is safe for men despite the fact that women are not built the same as men and would be impacted differently in the event of a crash (Ryan et al., 2020). This also remains an issue for individuals who may be male, but do not mimic the body type of the fit dummy, as the safety features are unable to be tested on all body types due to a lack of test dummy variety and a lack of legal reinforcement to test with multiple test dummies (Ryan et al., 2020). These areas where there is accessible and available data on one group but not other ethnic or gender groups, for example, are referred to as a data gap (Duran, 2019). Data gaps can lead to potentially dangerous outcomes as researchers may overlook, or be unable to recognize, that their research is not reflective of the entire population and may be only focusing on one demographic (Duran, 2019). There have been instances where assumptions were made that research would be effective on all gender and ethnic groups after success on one specific ethnic and gender group (Rouan et al., 2021). Diverse research teams are more likely to include multiple demographics in their research and clinical trials, as they may understand the underrepresentation of marginalized groups as well as the consequences for excluding them (Rouan et al., 2021). However, this research that can affect any average citizen who drives a car or takes medication, for example, will not increase in diversity if the individuals reaching these positions do not continue to become more diverse. There are many barriers that members of marginalized groups face throughout their academic and career trajectory, but the most common are the implicit biases of those who decide if an individual is eligible for the next step, such as a lab position, becoming a graduate student, or receiving a tenure-track position. Biases can involve gender or ethnicity, among other things, but the implicit biases that everyone has but may not be aware of, are what creates the most difficult barriers for marginalized individuals pursuing a career in STEM (Eaton et al., 2020).

Gender Bias

The science, technology, engineering, and mathematics field has become one of the most important work sectors; the field provides innovative solutions to the world's problems while the exponential growth provides many job opportunities. However, despite the fact that more women have pursued doctoral degrees in recent years, the percentage of women in tenure-track positions remains low (Eaton et al., 2020). Further, the proportion of science degrees granted to women has increased, but there is still a large disparity between the number of women receiving PhDs and those receiving junior faculty positions (Moss-Racusin et al., 2012). A study done by Moss-Racusin, Dovidio, Brescoll, Graham, and Handelsman found that science faculty evaluating two laboratory manager candidates, one female and one male, with identical applications resulted in the male candidate being evaluated as more competent, hireable, mentorable, and more worthy of a higher starting salary than the female candidate (Moss-Racusin et al., 2012). This was not unique to male faculty, as women faculty were just as likely to give female applicants lower evaluations. These perceptions of females having low STEM competency acts as a barrier for women striving for tenure-track and faculty positions, and even leadership or managerial positions within the STEM workforce.

A large influence on the gender biases present among many science faculty are common stereotypes associated with each gender. For example, a study done in 2016 found that women with a more feminine appearance were judged as less likely to be scientists (Banchefsky et al., 2016). Past research has also shown that both female and male adults rate women as less descriptively similar to successful scientists as opposed to men, and participants identified no overlap between traits associated with women (kind, caring, etc) and traits associated with scientists (competent, assertive, etc) (Carli et al., 2016). Common misconceptions regarding women in science, as addressed by a bias training video series called VIDS (Video Interventions for Diversity in STEM) , includes believing that a pregnant women will not be dedicated to her work during or after her pregnancy, discriminating against men and women who do not follow social stereotypes (men being kind or modest and women being confident), and believing that women are viewed as or assumed to be less competent than men (Moss-Racusin et al., 2018). Many similar experiments have shown that science faculty are more likely to evaluate men as more competent and hireable, while women are evaluated as more likeable (Eaton et al., 2020). This corresponds to the relation that the likeable qualities of women, kindness and warmth, are what restricts them from being seen as an assertive and confident scientist.

Based on the past research that has found that science faculty perceive women to be less competent in STEM and less descriptively similar to scientists, we expect that the results from the experiment will reflect this data. As we predict that men will be given higher competency ratings then women, this result will be mimicked in the fake data set.

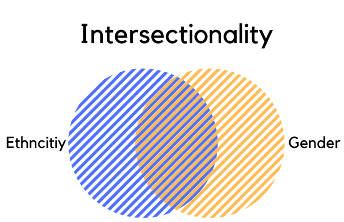
Ethnic Bias

Similar to gender biases, there are also prevalent ethnic stereotypes that can inhibit or enhance an individual's career in STEM. Ethnic bias can result in the discrimmination against individuals based on their ethnicity, which can be dangerous when evaluators hold implicit biases against certain ethnic groups ("APA Dictionary of Psychology", 2021). Individuals of marginalized groups, such as peoples of Black, Asian, or Latinx ethnicities, may face several obstructions or advancements within STEM for a majority of reasons. Black individuals are stereotyped as being less competent than Asian and White individuals, including within the STEM field, which can lead to a loss of feeling as though they belong in STEM and a reduction in intention to major in STEM fields (Blaine, 2013). Latinx individuals are also stereotyped as less competent and lower in STEM ability when compared to their Asian and White counterparts, commonly due to the misconception that Latinxs do not value formal education (Blaine, 2013; Valencia, 2002). The combined beliefs that Latinx individuals are not as competent as White individuals and that they do not value secondary education have led to negative academic consequences. For example, Latinas' concerns over how their professors will stereotype their academic abilities based on ethnicity has led to lower college GPAs and a lower sense of belonging at school (Valencia, 2002). However, Asian men and women are commonly classified as "intelligent," and due to their social ecology are seen as dedicated and hard-working (Eaton et al., 2020). This stereotype may make Asian individuals more appealing graduate and post-graduate applicants. All of these stereotypes, despite how untrue they may be, could severely impact the evaluation of a graduate student if they are a part of an evaluator's implicit, or possibly even explicit, ethnic bias.

Due to these common stereotypes that greatly influence ethnic bias, we expect that the results from the experiment will show that Black and Latinx applicants are seen as less competent compared to White and Asian applicants. These predictions will be displayed in the fake data set.

Intersectionality: The Overlap of Gender and Racial Stereotypes

Intersectionality refers to the complex and cumulative way in which the effects of multiple forms of discrimination, including racism and sexism, intersect primarily in the experiences of marginalized individuals or groups (Merriam-Webster, n.d.). Intersectionality takes into account people's overlapping identities and experiences to truly understand the entirety of the prejudices they face (Merriam-Webster, n.d.). It is essential to study barriers in STEM from an intersectional perspective, as individuals of multiple marginalized groups may encounter various deterrents in STEM as opposed to if they only belonged to one marginalized group (Eaton et al., 2020). For example, a Black woman will likely face more challenges in a STEM career than a Black man, as she would have to face discrimination against her gender and her race. With intersectionality in mind, it is fair to predict that members of marginalized gender and ethnic groups would be the most discriminated against by science faculty.



It is hypothesized that men will be evaluated as more competent than the women in every ethnic group, while Black and Latinx applicants will be evaluated as less competent than the White and Asian applicants. By applying intersectionality and combining these hypotheses, it can be deemed that Latinx women and Black women will have the lowest competency ratings, while White men and Asian men will have the highest competency ratings. Since this has been shown to be true in past research , this result will be shown in the fake data set (Eaton et al., 2020).

METHODS

Since this is predictive data analysis, the experiment has not yet been done. The research will be conducted nationwide, operating completely online. Participants, composed of science faculty from across Canada, will be split into three groups. Although the research question asks the level to which applicants should reveal personal information, this relies heavily on the presence or absence of evaluator bias. The first group will evaluate the reference letter of a graduate student applicant based on their qualifications, competency, and the likelihood of acceptance on a scale of one (not qualified/competent/likely) to ten (very qualified/competent/likely). Although all references will be the same, the names of the applicants will be changed to portray different genders and ethnicities. Thus, we will have a female and male Black applicant, a female and male White applicant, a female and male Latinx applicant, and a female and male Asian applicant. The second group will be asked to read the same reference letter and answer the same questions, but the names from the references will be redacted, making them completely anonymous. The third group will be exposed to an interview video regarding unconscious bias, backed by past research, and then asked to evaluate the named reference letters.

This data analysis will be completed using a fake data set. The data was created by random number input, however, there were guidelines set for each applicant to get results close to what is predicted. The fake data set was made to represent our expectations of group one, as outlined in the introduction. Therefore, the participants would have been given reference letters with names, and would not have seen the bias video intervention, allowing for their implicit biases to be shown. For simplicity, only the competency ratings of applicants will be analyzed. The presumed ethnicity and gender of an applicant have seemed to have the largest effect on the perceived competency of that applicant. This fake data will also be used to determine how large the difference in means must be between the applicants for the differences to be considered significant, as well as provide an ability to determine the process and procedure of the data analysis during the experiment.

RESULTS

There were a total of 800 competency ratings in the data set, 100 for each applicant. The ratings were then analyzed in three different settings; applicant by applicant, combined by ethnic group, and combined by gender. When comparing applicants, our female Black applicant and female Latinx applicant were rated considerably lower than other applicants. The female Black applicant was considered to have a competency rating significantly lower than the male White applicant (MD = 2.70, p = <0.001), the female White applicant (MD = 2.00, p = <0.001), the male Asian applicant (MD = 3.15, p = <0.001), the female Asian applicant (MD = 2.30, p = <0.001), and the Latinx male applicant (MD = 1.80, p = <0.001). The male Black applicant was rated significantly higher than the female Black applicant (MD=1.75, p = <0.001), but significantly less than the male White

Table 1. Average competency ratings (M) and standard deviations (SD) of each applicant.

	Applicant							
	White		Black		Latinx		Asian	
	Male	Female	Male	Female	Male	Female	Male	Female
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M(SD)	M(SD)
Competency Rating	7.30 (1.56)	6.60 (1.29)	6.35 (0.857)	4.60 (0.804)	6.40 (0.0.974)	4.10 (0.0.835)	7.75 (0.947)	6.90 (1.05)

applicant (MD=0.950, *p* = <0.001) and the male Asian applicant (MD = 1.40, *p* = <0.001). The female Latinx applicant was significantly lower than the female Black applicant (MD = 0.500, *p* = <0.001), the male White applicant (MD = 3.20, *p* = <0.001) and the male Asian applicant (MD = 3.65, *p* = <0.001). The male Latinx applicant was rated significantly higher than the female Latinx applicant (MD = 2.30, *p* = <0.001), only slightly lower than the male Black applicant (MD = 0.0500, *p* = 0.70), but considerably lower than the male Asian applicant (MD = 1.35, *p* = <0.001) and the male White applicant (MD = 0.900, *p* = <0.001). Male White applicants were considered more competent than female White applicants (MD = 0.700, *p* = <0.001), and this was also seen between Asian applicants (MD = 0.850, *p* = <0.001).

Table 2. Average competency ratings (M), standard deviations (SD), and comparative P-values (*p*) of each ethnic group and gender.

	Applicant			Applicant				Asian vs.		White vs.	
	Female	Male	Comparison	Asian	White	Black	Latinx	Black	Latinx	Black	Latinx
	M (SD)	M (SD)	*p*	M (SD)	M (SD)	M (SD)	M (SD)	*p*	*p*	*p*	*p*
Competency Rating	5.55 (1.58)	6.95 (1.27)	<0.001	7.33 (1.08)	6.95 (1.47)	5.48 (1.21)	5.25 (1.47)	<0.001	<0.001	<0.001	<0.001

When comparing all female applicants to all male applicants, the male applicants were evaluated as considerably more competent (MD = 1.40, *p* = <0.001). The Asian applicants, both male and female combined, outperformed the other ethnically grouped applicants. The Asian applicants had competency ratings higher than the White applicants (MD = 0.380, *p* = 0.0039) and significantly higher than the Black applicants (MD = 1.85, *p* = <0.001) and the Latinx applicants (MD = 2.08, *p* = <0.001). The White applicants, although considered less competent than the Asian applicants, still scored significantly higher than the Black applicants (MD = 1.47, *p* = <0.001) and the Latinx applicants (MD = 1.70, *p* = <0.001). The Black applicants and Latinx applicants were the closest rated (MD = 0.230, *p* = 0.095), having respective average competency ratings of 5.48 and 5.25.

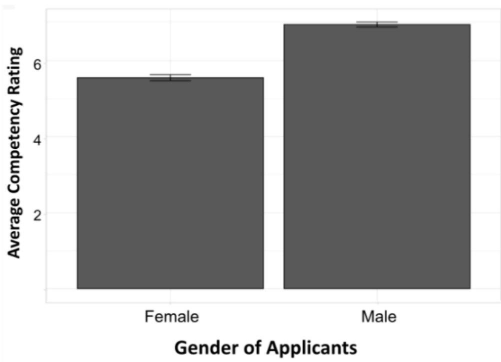


Figure 1. Average competency ratings of applicants grouped by gender as evaluated by science faculty, rated on a scale of 1-10.

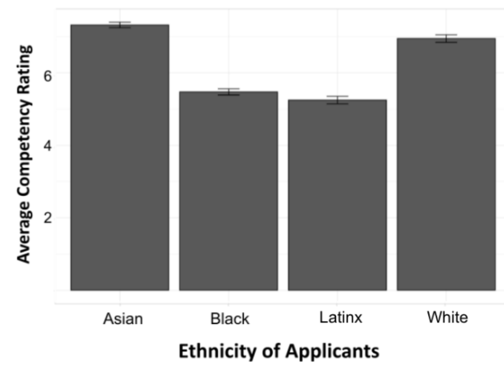


Figure 2. Average competency ratings of applicants grouped by ethnicity as evaluated by science faculty, rated on a scale of 1-10.

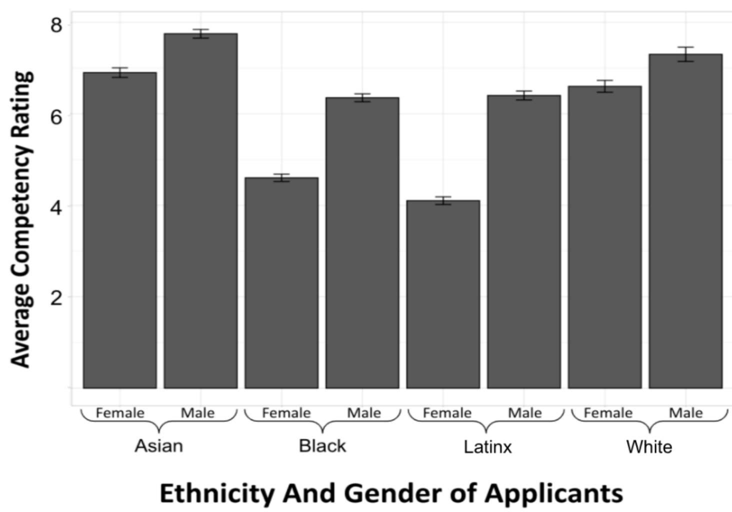


Figure 3. Average competency ratings of of varying ethnicity and gender as evaluated by science faculty, rated on a scale of 1-10

DISCUSSION

As predicted and designed, the data met the expectations outlined during the introduction. The male applicants had an increase in competency rating, as opposed to their identical female applicants, by 25.2%. The Asian applicants and White applicants outperformed the Black applicants and Latinx applicants by significant margins. Although not predicted, the Asian applicants' competency rating slightly exceeded the White applicants. However, the margin of difference was extremely low. If this were to happen within the experiment, it can be attributed to the stereotypes perceiving Asians as dedicated, hard-working, and intelligent. The male White and male Asian applicants were rated as the most competent applicants, while the female Black and female Latinx applicants were rated as the least competent applicants.

Although this data was fake, it provided some expectations for the analysis of the data that will be produced by the experiment. For example, the comparisons between competency ratings set guidelines for the relationship between significance and mean differences. The larger the mean difference between two groups, the smaller the P-value and consequently the more significant the difference, allowing for the objection of the null hypothesis. The lowest mean difference that resulted in a P-value of <0.001 was 0.500. The largest P-value, 0.70, and consequently the least significant difference, was found when comparing two groups with a mean difference of 0.05. However, this is greatly influenced by the size of the sample and variability, which will vary as the data gathered from the experiment will be much more extensive. However, standard error will decrease as the sample size increases and becomes closer to the population, so the more evaluations the experiment receives for each applicant, the more accurate the data will be in representing the faculty population.

Although there were many limitations within the fake data set, such as that all of the numbers were randomly determined within a set range for each applicant and there were only 100 evaluations per applicant, there is still a possibility for limitations within the experiment. In the current social climate, science faculties have been taking steps to reduce biased decisions by their staff. This sudden increase in knowledge regarding EDI and personal bias may result in fair evaluations of our applicants. If our applicants were fairly evaluated in the real experiment, then the differences between competency ratings of applicants would be minimal if present. Nonetheless, it is also relevant which science field the evaluator belongs to. A similar study conducted had separated the biology professors from physics professors and found that biology professors were more likely to give marginalized individuals higher competency ratings compared to the physics professors who favoured the White male applicants (Eaton et al., 2020). This may be related to the fact that biology is already a diverse field, and therefore professors would be more willing to uphold diversity. The opposite applies to physics, as the professors held strong biases favouring male White applicants as the field is dominated by men of the same demographic and is only starting to become more diverse (Eaton et al., 2020).

If the data from the experiment is similar to the data used in the predicted analysis, the presence of implicit bias will be prevalent through the evaluations of competency, as well as qualification and hireability. As both men and women have a tendency to equally discriminate against women and hold implicit biases favouring men, men having a competency rating higher than women is a predicted outcome in the experiment (Moss-Racusin et al., 2012). Therefore, this result was produced in our predicted data analysis, where a mean difference of 1.40 produced a P-value of <0.001, rendering this value significant. A somewhat similar result is expected amongst group one in our experiment, assuming that those individuals have not been exposed to

UNIVERSITY OF ALBERTA

Canada

previous bias training and are not aware of their implicit bias. Due to common misconceptions about ethnic groups, such as that Black individuals are not as intelligent as their White counterparts, Latinx families do not value education, and that Asians are dedicated overachievers, the resulting differences between ethnic groups in our fake data are expected to be repeated in the experiment. Male applicants were evaluated as more competent than their identical female counterparts in every ethnic group, with the mean differences varying from 0.700 between Black applicants to 2.30 between Latinx applicants. When identifying the gender with the lowest competency rating as well as the ethnicities with the lowest competency ratings, we can see intersectionality at play. Theoretically, female Black and female Latinx applicants, both members of the marginalized gender group and a marginalized ethnicity, would be the most discriminated against and therefore would have the lowest competency ratings. Therefore, we can hypothesize that these results will also be present in the experimental results, just as they were shown in the fake data.

The presence and success of marginalized people in STEM does not demonstrate the absence of bias against them, rather these individuals have persisted despite ethnic and gender biases (Moss-Racusin et al., 2018). Females and marginalized groups continue to be outnumbered in graduate programs, tenure-track positions, research labs, and leadership roles in STEM, but this is not due to a lack of interest (Moss-Racusin et al., 2012). Members of marginalized groups are constantly facing unique barriers in STEM that are in large part due to ethnic and gender stereotypes. By conducting this experiment, we hope to find ways that can help promote a level playing field for all applicants, whether that be through anonymizing personal information to avoid any opportunity for bias, or by finding effective video interventions to help faculty identify and find ways to combat their implicit biases. For educators and evaluators who fear their implicit biases have impacted or may impact their teaching or decision making process, it is important to take the time to become educated on implicit bias, identify your own implicit biases, and take actions to counter your biases. Perhaps the best way to do this is by seeking bias training or intervention videos. By conducting this experiment, we hope to help make STEM a field where everyone, from every gender and ethnic group, can feel as though they belong and receive opportunities based on qualification and not the biases of an evaluator. Then perhaps, research will become more diverse and inclusive, further preventing negative impacts on marginalized groups as they no longer will be excluded or overlooked in research, clinical trials, and testing.

ACKNOWLEDGEMENTS

Thank you to my sponsors, Dr. Matina Kalcounis-Rueppell from the University of Alberta's Faculty of Science and Canada Summer Jobs, for their generous support! An important thank you to Dr. Lisa Willis, the Principal Investigator, and Ronak Patel, my supervisor, for having me in their lab this summer and providing guidance and support throughout this project! Thank you to Saige Burnett, for being an incredible lab partner and offering aid

while I worked on this project. A huge thank you to the WISEST team for organizing this fantastic summer program that has made an incredible impact on my life, as well as the lives of my fellow researchers, through experiences in STEM fields that have allowed me to find my true calling!

WORKS CITED

APA Dictionary of Psychology. Dictionary.apa.org. (2021). Retrieved 11 August 2021, from https://dictionary.apa.org/ethnic-bias.

Banchefsky, S., Westfall, J., Park, B., & Judd, C. (2016). But You Don't Look Like A Scientist!: Women Scientists with Feminine Appearance are Deemed Less Likely to be Scientists. *Sex Roles, 75*(3), 95-109. 10.1007/s11199-016-0586-1

Blaine, B.E.(2013). Understanding race, racial stereotypes,and racism. In B.E.Blaine (Ed.), Understanding the psychology of diversity (2nd ed.,pp.87–112). Thousand Oaks, CA:Sage Publications,Inc.

Carli, L. L., Alawa, L., Lee, Y., Zhao, B., & Kim, E. (2016). Stereotypes About Gender and Science. *Psychology of Women Quarterly, 40*(2), 244-260. 10.1177/0361684315622645

Duran, J. G. (2019, February 10). How data gaps Affect Canadians and researchers alike. The Varsity. https://thevarsity.ca/2019/02/10/how-data-gaps-affect-canadians-and-researchers-alike/.

Eaton, A. A., Saunders, J. F., Jacobson, R. K., & West, K. (2020). How Gender and Race Stereotypes Impact the Advancement of Scholars in STEM: Professors' Biased Evaluations of Physics and Biology Post-Doctoral Candidates. *Sex Roles, 82*(3-4), 127-141. 10.1007/s11199-019-01052-w

Merriam-Webster. (n.d.). Intersectionality. In Merriam-Webster.com dictionary. Retrieved July 29, 2021, from https://www.merriam-webster.com/dictionary/intersectionality

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 - PNAS, 109*(41), 16474-16479. 10.1073/pnas.1211286109

Moss-Racusin, C. A., Pietri, E. S., Hennes, E. P., Dovidio, J. F., Brescoll, V. L., Roussos, G., & Handelsman, J. (2018). Reducing STEM Gender Bias With VIDS (Video Interventions for Diversity in STEM). *Journal of Experimental Psychology. Applied, 24*(2), 236-260. 10.1037/xap0000144

Rouan, J., Velazquez, G., Freischlag, J., & Kibbe, M. R. (2021). Publication bias is the consequence of a lack of diversity, equity, and inclusion. *Journal of Vascular Surgery, 74*(2), 111-117S. 10.1016/j.jvs.2021.03.049

Ryan, A., Tainter, F., Fitzpatrick, C., Gazzillo, J., Riessman, R., & Knodler, M. (2020). The impact of sex on motor vehicle crash injury outcomes. *Null,* , 1-25. 10.1080/19439962.2020.1834478

Valencia, R. R. (2002). Mexican Americans Don't Value Education! On the Basis of the Myth, Mythmaking, and Debunking. *Journal of Latinos and Education, 1*(2), 81-103. 10.1207/S1532771XJLE0102_2

References

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 - PNAS, 109*(41), 16474-16479. 10.1073/pnas.1211286109