COMPARING REU STUDENT VS. ADVISOR RATINGS OF STUDENT PERFORMANCE SUGGESTS LOW STUDENT SELF-EFFICACY

A THESIS SUBMITTED TO
THE GLOBAL ENVIRONMENTAL SCIENCE
UNDERGRADUATE DIVISION IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

BACHELOR OF SCIENCE

IN

GLOBAL ENVIRONMENTAL SCIENCE

MAY 2020

By
CHERRYLE HEU

Thesis Advisor

DR. BARBARA BRUNO

I certify that I have read this thesis and that, in my opinion, it is satisfactory in scope and quality as a thesis for the degree of Bachelor of Science in Global Environmental Science.

THESIS ADVISOR

____

Dr. Barbara Bruno
Hawaii Institute of Geophysics & Planetology

ACKNOWLEDGEMENTS

I would like to thank my thesis advisor, Dr. Barbara Bruno, for introducing me to this project and for being an incredible mentor. I am also grateful to the ‘Ike Wai Education Team, including Dr. Jenny Engels, for their continuous support and for teaching me invaluable lessons. I would also like to thank the entire GES program for being especially supportive and throughout my four years. Finally, I would like to thank my family – Vernon Heu, Atsuko Heu, Jaycob and Kento for their endless love and unconditional support. Without everyone’s support, I would not be where I am today.

This research was supported by the National Science Foundation (NSF/GEO #1565950).

ABSTRACT

This study compares survey responses of 30 undergraduate researchers and their advisors in the School of Ocean and Earth Science and Technology (SOEST). Students and advisors each completed surveys that evaluated the student's research skills and performance at the end of the research program. Student and advisor responses to each question were compared using a paired two-tailed t-test, and then a non-parametric Permutation Test was applied to the entire dataset. I found that on average, students significantly underrated their skills and performance relative to their advisors' rating (p=0.005). I interpret these results in light of self-efficacy, which has been found to be a key factor in undergraduate success. This suggests that SOEST (and perhaps other undergraduate research programs) should focus on not only building research skills and ability, but also building self-efficacy for their students. Further demographic analyses revealed that, of the four subgroups studied, the underrating of female Native Hawaiian/Pacific Islander students was the most significant (p=0.02), suggesting that efforts to build student self-efficacy be especially targeted at students that are of more than one underrepresented group.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 Introduction

1.1 Research Experiences for Undergraduates (REU) program

Research Experiences for Undergraduates (REU) is a program funded by the National Science Foundation (NSF) designed to promote active research participation amongst undergraduate students in various science fields. Most REU programs consist of approximately 10 to 15 undergraduate students working closely with a faculty supervisor on closely mentored research projects. There are over 700 REU sites across the country. NSF invests millions of dollars annually to run this program, with an estimated $76,000,000 of funding for fiscal year 2020 (National Science Foundation, 2019a). The NSF REU program aims to diversify Science, Technology, Engineering, and Mathematics (STEM) fields by increasing participation of those from underrepresented groups, including women, persons with disabilities, and ethnic/racial minorities (African Americans, Hispanics, American Indians, Alaska Natives, Native Hawaiians and Pacific Islanders) (National Science Foundation, 2019b).

Research experiences have been associated with a student's likelihood to persist in science. Compared to their non-REU peers, undergraduate researchers have demonstrated increased understanding, confidence and awareness. In a post-REU survey, most participants indicated their plans to continue to postgraduate education (Russell et al., 2007). Of REU participants, students of underrepresented groups reported higher gains from the program compared to their peers (Lopatto, 2007). This could increase their likelihood to pursue a STEM career, thus eventually diversifying the STEM workforce. However, how much of the difference between REU students versus the non-REU students is causal, as opposed to correlational, is unclear.

1.2 Diversity of the STEM Workforce

While the REU program has continued to train diverse participants since its establishment in 1987, STEM workforce demographics continue to show a lack of diversity. In 2014, women comprised 50% of all STEM undergraduate degree awardees in the U.S. and 47% of the overall U.S. workforce (U.S. Census Bureau 2015 & National Science Foundation 2015). However, women only made up 28% of the STEM workforce (Table 1). Female participation varies greatly by STEM field. Although women make up nearly half of the biological, agricultural and environmental life scientists (48%), they are much less represented in the physical science fields, such as computer and mathematical scientists (26%), engineers (15%), and geologists (21%) (National Science Foundation 2015).

Table 1: Gender and ethnic makeup of U.S. STEM occupations in 2015 (Source: National Science Foundation 2015)

	Female		White	Asian	Hispani	Black	Native Hawaiia Pacific Islander %	Americ Alaska Native %	>1 race
All occupations	47.0	53.0	70.5	12.3	8.0	6.8	0.3	0.3	1.8
All STEM occupations	28.4	71.6	66.6	20.6	6.0	4.8	0.2	0.2	1.6
Computer and mathematical sciences	26.4	73.6	62.3	25.8	5.0	5.1	0.2	0.2	1.6
Engineers	14.5	85.5	70.6	16.2	7.0	4.3	0.2	0.2	1.6
Biological, agricultural, and environmental life scientists	47.9	52.1	71.2	18.1	5.9	2.5	0.2	0.3	1.9
Forestry and conservation scientists	20.3	79.7	93.2	< 0.1	s	< 0.1	s	s	s
Physical and related scientists	27.8	72.2	70.4	19.3	5.1	3.9	s	s	0.9
Atmospheric and space scientists	26.7	66.7	73.3	s	< 0.1	s	s	s	< 0.1
Geologists, including earth scientists	20.8	79.2	89.6	4.2	4.2	2.1	s	s	2.1

Compared to gender, the racial makeup of the STEM workforce is more reflective of the racial makeup of all occupations in the U.S. (Table 1). However, certain fields in STEM are less diverse than others. In 2015, whites accounted for 90% of geologists and 93% of forestry and conservation scientists, while only 62% of computer and mathematical scientists (National Science Foundation 2015).

Meanwhile, the demographics of STEM majors show greater gender and racial diversity than the STEM workforce (Table 2). This indicates that undergraduate STEM majors do not necessarily pursue STEM careers. This suggests that there may be challenges faced by women and minority STEM students and that perhaps increasing research opportunities for these students are not enough to increase their chances of pursuing a STEM career.

Table 2: Gender and ethnic makeup of U.S. STEM Bachelor's degrees awarded in 2014 (Source: National Science Foundation

	Female		White	Asian			Native Hawaiia Islander %	America n Indian Alaska Native %	>1 race
All majors	57	43	61	6	11	10	0.3	0.5	2
STEM majors	50	50	59	9	12	8	0.2	0.5	3
Computer and mathematical sciences	25	75	56	10	9	8	0.2	0.4	2
Engineers	20	80	62	11	10	4	0.2	0.3	2
Biological, agricultural, and environmental life scientists	59	41	58	15	11	7	0.2	0.5	2.8
Physical and related scientists	40	60	62	12	8	6	0.2	0.5	3
Atmospheric and space scientists	34	66	81	2	7	3	0.1	0.4	2
Geologists, including earth scientists	39	61	77	3	7	2	0.1	0.7	2

1.3 Self-Efficacy

The motivation of this study is to see how student self-evaluations of their own skills and performance compare with their advisors' evaluations. This could potentially shed light on student self-efficacy, which is defined as "people's beliefs about their capabilities to produce designated levels of performance that exercise influence over events that affect their lives." (Bandura, 1994).

A person with a strong sense of self-efficacy believes in their capabilities to succeed in a given task and deal with the challenges they encounter. Meanwhile, a person with a weak sense of self-efficacy may underestimate their abilities and feel overwhelmed by challenges that arise. Self-efficacy is more specific than self-confidence. Self-confidence broadly refers to the strength in belief, but not particularly in ability or growth. Self-confidence could, however, largely influence one's self-efficacy (Bandura, 1997).

Self-efficacy is considered to be applicable at all stages in one's life, from a student's belief they will pass a geology exam to an early career scientist's belief that they will have a successful science career. Studies have found that self-efficacy strongly influences various factors such as motivation and self-regulatory processes that can have positive, long-term effects. In an introductory chemistry class, Zusho et al. (2003) found that self-efficacy was the best predictor of final course performance, even when controlling for prior achievement. Similarly, Lent et al. (1986) found a significant correlation between students with high self-efficacy and career persistence in STEM.

Several research studies have shown that men tend to have a stronger sense of self-efficacy than women (Bandura, et al., 2001 & Williams and George-Jackson, 2014). According to the United Nations Educational, Scientific and Cultural Organization (UNESCO, 2015), both boys and girls meet standard proficiency in math and science in primary school. However, gender differences become apparent at the secondary level. The Programme for International Student Assessment (PISA, 2012) revealed that among 15-year-olds in 65 countries, girls in all but three countries were more likely to report feeling "helpless while performing a math problem" compared to their male counterparts, despite performing similarly on a math assessment.

Such gender differences in STEM self-efficacy has been attributed to various factors such as the gendered-constructed nature of STEM fields and the lack of visibility of women in professional STEM careers (Spencer et al., 1999 & Stout et al., 2011). This suggests that building self-efficacy amongst female students is essential for achieving gender parity in STEM fields. Studies have also found that self-efficacy was a strong indicator of academic achievement for indigenous students in higher education (Bryan, 2004 & Frawley, 2017).

Considering that self-efficacy is a strong predictor for student success and career persistence, this study could shed light on whether there is a stronger need for REUs and other STEM undergraduate programs to focus on this important quality. Intentionally working to build self-efficacy among all program participants (and especially women and minorities) may increase their likelihood of success and advancement into a STEM career.

2.0 Data and Methods

2.1 Data

This study surveyed undergraduate students (n=30) participating in the SOEST Scholars program, an REU-type program at the School of Ocean and Earth Science and Technology (SOEST) at the University of Hawaii that runs through the academic year (September to May). The survey was distributed online to two cohorts of participating students and their advisors at the completion of the REU program. A total of 30 student-advisor pairs were surveyed from the 2016-17 and 2017-18 cohorts. Each student-advisor pair was given a unique code to input into the survey to avoid inputting personally-identifying information. This code was used to link the student's survey responses with that of their advisor for analysis.

Of the 30 SOEST Scholars, 16 identified as female and 14 identified as male. No student reported a non-binary gender. Fourteen scholars identified as Native Hawaiian and one as Pacific Islander. Therefore, half of the 30 SOEST scholars reported indigenous Native Hawaiian or Pacific Islander (NHPI) ethnicities. The remaining 15 scholars reported a diverse range of non-indigenous (Non-NHPI) ethnicities, including African-American (2), Asian (5), Caucasian (5), and Filipino (3).

2.2 Survey

Students and advisors each responded to ten survey items designed to evaluate the student's skills and performance during the REU in ten areas (see Table 3). This protocol was approved as exempt by the University of Hawaii Institutional Review Board (#2017-00612).

Table 3: Student Survey Items

1. **Research Productivity:** The amount of work that I accomplish.
2. **Research Quality:** The quality of work that I perform.
3. **Motivation and Initiative:** My self-motivation and willingness to take initiative, as appropriate.
4. **Organizational Skills:** My ability to organize tasks in an efficient manner.
5. **Communication Skills:** My verbal and written communication skills.
6. **Professionalism:** My ability to behave in a professional manner.
7. **Cooperation:** My ability to work as a member of a research group or team.
8. **Independence:** My ability to work independently, as appropriate.
9. **Self-Appraisal:** My ability to analyze my performance and to make constructive efforts to improve.
10. **Time and Attendance:** My ability to maintain agreed-upon research hours and schedule.

(Surveys administered to advisors were worded in the third person, using "the student" instead of "I")

2.3 Data Quantification and Descriptive Statistics

For each survey item, students and advisors were asked to choose a Likert scale response from the following: Unsatisfactory, Fair, Satisfactory, Very Good, and Excellent. To enable quantitative analysis, Likert scale responses were converted to numerical values as follows:

1 – Unsatisfactory; 2 – Fair; 3 – Satisfactory; 4 – Very Good; 5 – Excellent

After converting Likert scale responses, the mean student (S) and advisor (A) response were calculated and their difference was calculated as D = A-S. The corresponding standard error of the Student mean (S SEM) and Advisor mean (A SEM) of each survey item were found.

2.4 Student's Paired T-test

To see if there was a significant difference between the student and advisor mean responses a paired, two-tailed student's t-test was used. Significance was determined using a significance value (α) of 0.05.

Ideally, a paired, two-tailed t-test would similarly be used to determine whether there is a significant difference between the student and advisor mean responses to the dataset as a whole (300 student/advisor pairs). However, doing so would violate the t-test assumption that each pair of data is random and independent of the others. Instead, a non-parametric Permutation Test must be used.

2.5 Non-parametric Permutation Test

To compare the student vs. advisor survey responses to all 10 questions combined (300 pairs), a non-parametric Permutation Test was conducted. While the t-test uses an assumed distribution to calculate a p-value, the Permutation Test generates a p-value by repeatedly sampling the data. By comparing the observed survey responses to a set of randomized data, the significance of the student-advisor differences can be found. (Ross, 2014)

The test was performed using a self-produced code in Matlab (see Appendix 1). In each permutation, a subset of 300 pairs was randomly selected to swap its scholar (s) and advisor (a) rating. For example, if a pair of $s = 3$ and $a = 5$ was randomly selected, the permuted pair's new value becomes $s = 5$ and $a = 3$. If a pair was not selected, its values remain unchanged. The mean difference (D) of the new set of 300 pairs is then computed. This procedure was repeated for

100,000 permutations, which produced 100,000 values of D, and was then plotted on a histogram. Finally, the p-value was computed empirically. For example, if the observed mean difference is 0.12, then the p-value is computed as the sum of the number of D greater than 0.12 and less than -0.12, divided by 100,000 permutations.

2.6 Demographic Analysis

To further investigate any potential patterns in survey results, the responses were separated by student gender and ethnic background. For this demographic analysis, only the Permutation Test was performed on the dataset as a whole (all 10 questions). The t-test on individual questions was not performed due to limited sample size. In the gender analysis, male (n=14) and female (n=16) students were separately analyzed to see whether there were any differences between these groups.

In the ethnicity analysis, the sample size was not sufficient to individually analyze each ethnicity. Instead, the ethnicity analysis was conducted based on whether the scholar was Native Hawaiian and/or Pacific Islander (NHPI, n=15), or non-NHPI (n=15).

To explore the role of the interconnection between gender and ethnicity, an intersectionality analysis was conducted. Intersectionality is “the theory that the overlap of various social identities, as race, gender, sexuality, and class, contributes to the specific type of systemic oppression and discrimination experienced by an individual” (Dictionary.com) People that identify as belonging to more than one underrepresented group often experience “double discrimination" at the intersection of their identities. For example, a Native Hawaiian woman may experience both racist and sexist discrimination.

Scholars were then divided into four intersectional categories based on their gender and ethnicity: Female NHPI (n=8), Male NHPI (n=7), Female Non-NHPI (n=8), and Male Non-NHPI (n=7). Each of these four subgroups were analyzed separately and compared.

3.0 Results

3.1 Overall Analyses

3.1.1 Student's T-test Analysis

Eight out of ten survey items had a positive mean difference (D>0), indicating that advisors on average gave their student a higher rating than the student gave themselves (Figure 1 and Table 3). For one item (#7), the students and advisors gave equal mean ratings (D=0). For the last item (#8), the students gave a slightly higher mean rating than did their advisors (D=-0.03).

Of the eight survey items with a positive mean difference, two were statistically significant: Item #2, Research quality (p=0.04) and Item #4, Organizational skills (p=0.05).

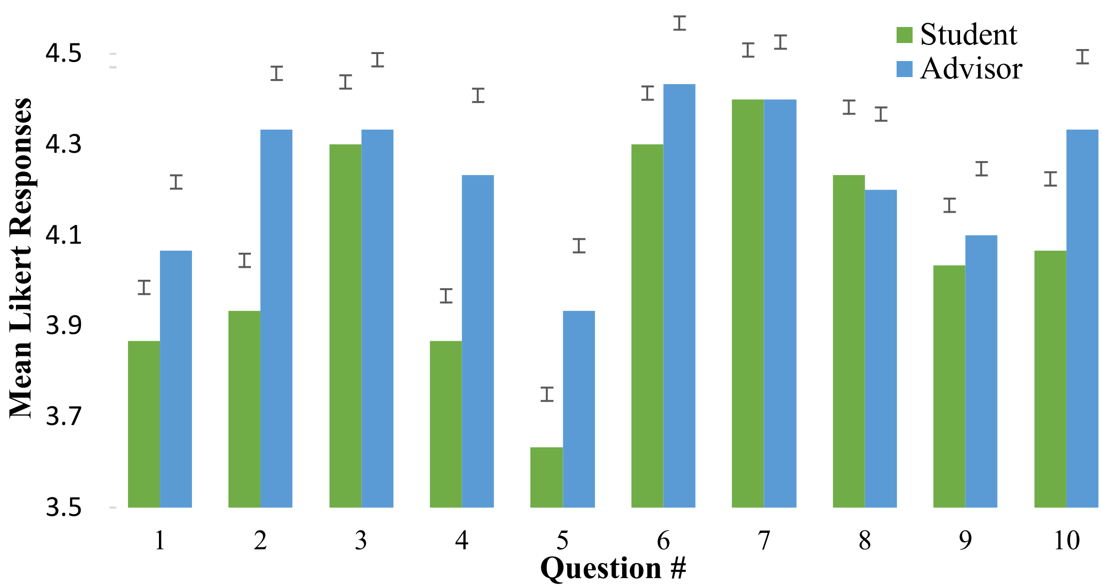

Figure 1: Mean student (green) vs. advisor (blue) ratings for each survey item (see Table 3). Error bars represent +/- standard error of the mean.

Table 4: Comparison of student and advisor ratings for ten survey items.

Survey Item	S[1]	S SEM[2]	A[3]	A SEM[4]	D[5]	p[6]
1. Research Productivity	3.87	0.13	4.07	0.17	0.20	0.28
2. Research Quality	3.93	0.13	4.33	0.14	0.40	0.04*
3. Motivation and Initiative	4.30	0.15	4.33	0.17	0.03	0.87
4. Organizational Skills	3.87	0.11	4.23	0.19	0.37	0.05*
5. Communication Skills	3.63	0.13	3.93	0.16	0.30	0.12
6. Professionalism	4.30	0.13	4.43	0.15	0.13	0.40
7. Cooperation	4.40	0.12	4.40	0.14	0.00	1.00
8. Independence	4.23	0.16	4.20	0.18	-0.03	0.89
9. Self-Appraisal	4.03	0.15	4.10	0.16	0.07	0.75
10. Time and Attendance	4.07	0.17	4.33	0.18	0.27	0.17
All Survey Items	*4.06*	*0.05*	*4.24*	*0.05*	*0.18*	*0.005*

*: $p < 0.05$
[1]S=Mean student rating; [2]Standard Error of S; [3]A=Mean advisor rating; [4]Standard Error of A; [5]D=A-S; [6]p=probability value

3.1.2 Non-Parametric Permutation Test Analysis

Although the mean differences for the individual questions were not generally statistically significant, the trend of eight out of ten items having positive mean differences suggests that students may be underrating their skills and performance as a whole compared to their advisors' assessment. The observed mean difference (D) of student and advisor ratings on the overall dataset was 0.18 (Table 4). To test the statistical significance of the overall dataset, the non-parametric permutation test was performed.

Figure 2 presents the results of the permutation test on the overall dataset (300 student-advisor pairs to 10 questions). Of the 100,000 permutations, ~500 were tailward compared to the observed mean difference (0.18) -- that is, greater than +0.18 or less than -0.18. Thus, I empirically calculate p as 500 / 100,000 = 0.005. This indicates that the positive mean difference was highly significant. Implications of this are discussed below (See Discussion Section).

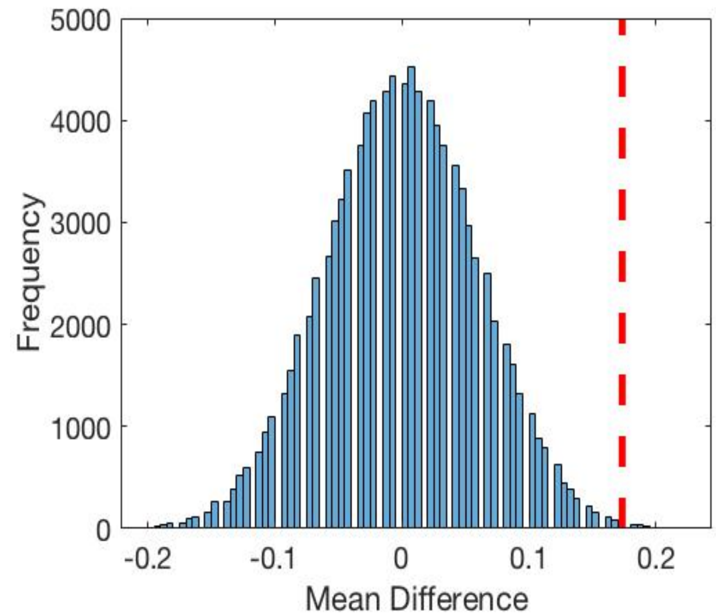

Figure 2: Histogram of distribution of D of permuted data. Red dashed line indicates observed mean difference (D = 0.18)

3.2 Demographic Analyses

To further investigate survey results, the responses were separated by demographic data based on gender and ethnic background. The non-parametric permutation test was conducted for each demographic group to determine whether there were any demographic patterns in survey responses. Results are presented in Table 5.

3.2.1 Gender

First, gender was analyzed. Both male and female students gave themselves a lower mean rating compared to the mean advisor ratings (Male D=0.09 & Female D=0.24). For both male and female students, the mean difference was positive, indicating advisors on average gave higher ratings than students did for both genders. However, the statistical significance greatly differed for the two genders. Female students' underrating of themselves relative to their advisors' rating was highly significant (p=0.005). In sharp contrast, the difference between the ratings of male students and their advisors was not statistically significant (p=0.31).

3.2.2 Ethnicity

The data were then analyzed by ethnicity. I compared the self-assessment of Native Hawaiian and Pacific Islander (NHPI) students vs. all other non-NHPI students. The latter category includes African-American, Asian, Caucasian, and Filipino students. For both groups, advisors' mean rating was higher than that of the students for both groups, but much more so for NHPI (NHPI D=0.24 & non-NHPI D=0.11). Moreover, the NHPI students' underrating of their performance relative to their advisor's ratings was highly significant (p=0.01) while non-NHPI students' underrating was not statistically significant (p=0.18).

Table 5: Comparison of student and advisor ratings by gender (male vs. female) and ethnicity (NHPI vs. Non-NHPI).

		n	S[1]	S SEM[2]	A[3]	A SEM[4]	D[5]	P[6]
Gender	Male	14	3.97	0.06	4.06	0.08	0.09	0.31
	Female	16	4.14	0.07	4.39	0.07	0.24	0.005
Ethnicity	NHPI	15	3.91	0.06	4.15	0.08	0.24	0.01
	Non-NHPI	15	4.21	0.06	4.32	0.06	0.11	0.18
All		30	4.06	0.05	4.24	0.05	0.18	0.005

[1]S=Mean student rating; [2]Standard Error of S; [3]A=Mean advisor rating; [4]Standard Error of A; [5]D=A-S; [6]p=probability value

3.2.3 Intersectionality

The two demographic groups were divided into four subgroups (Female NHPI, Male NHPI, Female Non-NHPI, and Male Non-NHPI) to look into how the two demographic categories interplayed with one another. All four subgroups underrated themselves relative to their advisor's mean rating (Table 6). Female NHPI students had the highest D of 0.36, while the other three subgroups showed a much lower degree of underrating with D ranging from 0.09 to 0.12.

Furthermore, only female NHPI students' underrating of themselves relative to their advisors' rating was highly significant (p=0.02). The student-advisor differences associated with the other groups were not significant (p=0.18 to 0.59).

Table 6: Intersectional comparison of student and advisor ratings

	n	S[1]	S SEM[2]	A[3]	A SEM[4]	D[5]	p[6]
Female NHPI	8	3.80	0.10	4.16	0.11	0.36	0.02
Male NHPI	7	4.04	0.07	4.14	0.12	0.10	0.43
Female Non-NHPI	8	4.49	0.08	4.61	0.07	0.12	0.18
Male Non-NHPI	7	3.90	0.09	3.99	0.10	0.09	0.59

[1]S=Mean student rating; [2]Standard Error of S; [3]A=Mean advisor rating; [4]Standard Error of A; [5]D=A-S; [6]p=probability value

4.0 Discussion

4.1 Student Underrating

Overall, SOEST scholars significantly underrated their skills and performance compared to their advisors' rating (D=0.18; p=0.005). The underrating was largely driven by students more likely to give themselves a "Very Good" rating while advisors were more likely to give their student an "Excellent." (Figure 3)

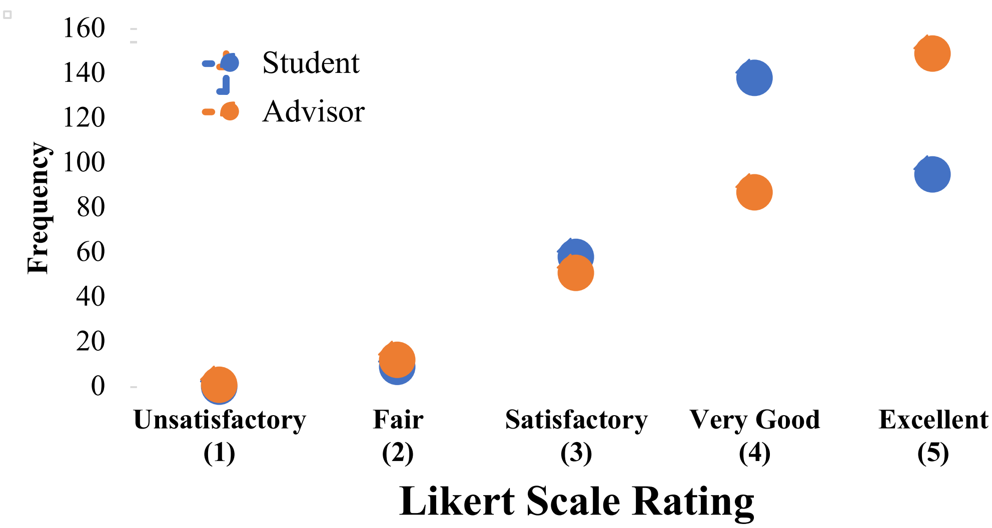

Figure 3: Line graph of the frequency of Likert responses for students (blue) vs. advisors (orange)

Separate demographic analyses of gender and ethnicity indicated that female (D=0.24; p=0.005) and NHPI (D=0.24; p=0.01) each had significantly underrated their skills and performance relative to their advisors, in contrast to the results of male and non-NHPI students. However, the intersectionality analysis revealed that the significant underrating by female and NHPI students was in fact due exclusively to the responses of female NHPI students (D=0.36; p=0.02). That is, female non-NHPI students did not significantly underrate themselves relative to their advisors' ratings (D=0.12; p=0.18); neither did male NHPI students (D=0.10; p=0.43). Male non-NHPI students showed the least magnitude -- and the lowest significance value -- of their underrating (D=0.09; p=0.59).

My finding that female NHPI students were more likely to underrate themselves than their peers highlights the relevance of intersectionality. This may reflect that female NHPI SOEST Scholars, on average, have lower self-efficacy than the other groups studied.

4.2 Suggestions for REU programs

Considering how self-efficacy is a strong predictor for student success in STEM (Zusho et al. 2003), this may imply that the SOEST Scholars program -- and perhaps other undergraduate research programs -- may need to consider intentionally focusing on building self-efficacy of their students, particularly those at the intersection of underrepresented identities. I recommend that REU program managers become familiar with the literature on how to promote self-efficacy.

Bandura (1977, 1994) mentions four main sources of self-efficacy in the following order: Mastery experiences, vicarious experiences, social persuasion, and physiological reactions. Of these, two are the most important: Mastery experiences (which involve success from perseverance and overcoming obstacles) and vicarious experiences (such as seeing social models similar to the student succeed) (Bandura, 1977, 1994).

Kortz (2019) had found that in a geoscience REU, promoting self-efficacy rather than only cognitive factors (e.g. knowing and understanding concepts) made a large difference in student persistence in STEM fields. Kortz found that success in field activities, an example of Bandura's mastery experiences, was the strongest source of self-efficacy for students. Through learning how to use different instruments, collecting samples, running tests to analyze them, etc., students were able to overcome challenges. Self-efficacy also increased in students through vicarious experiences, by observing others' success, particularly in peer leaders and their mentors.

The role of self-efficacy among indigenous undergraduate students has been studied as well. Gloria et al. (2001) found that among 83 American Indian undergraduates, student self-efficacy was strongly influenced by mentoring, positive perception of university environment, and social support interventions. Higher self-esteem and academic self-efficacy were also associated with a higher likelihood of career persistence decisions.

Therefore, REU programs should consider incorporating mastery and vicarious experiences into their programs to promote student self-efficacy and increase the likelihood of student career persistence in STEM.

Limitations

This study has numerous limitations, and three of these are discussed below. First, the survey was not validated to ensure that it measures student self-efficacy. Thus, self-efficacy is only provided as one possible explanation as to why students' self-assessments may be lower than their advisors' assessments.

Second, the sample size was small (30 student-advisor pairs). Ideally, I would have liked to analyze each demographic group separately (e.g., African American, Caucasian, Asian, etc). However, the small sample size prevented me from doing this. Instead, I lumped all non-NHPI ethnicities together into a “non-NHPI category”. In doing so, I combined the results of students from groups that are underrepresented in STEM fields (e.g, African-American) with those of students that are overrepresented in STEM fields (e.g, Caucasian, Asian). Therefore, I recommend caution when interpreting the results of ‘non-NHPI” students. I am more confident in interpreting results from the “NHPI” category, as all of these students share indigenous identities.

Finally, my statistical analyses were dependent on converting responses from a Likert scale to a quantitative number (1-5). This quantification assumes that each Likert item is equally spaced -- i.e., that the difference between unsatisfactory (1) and fair (2) is equivalent to the difference between fair (2) and satisfactory (3); and (b) – which may or may not be true in the minds of students and/or their advisors. There is another type of permutation test (a sign test) that circumvents this assumption by only looking at the sign – and not the magnitude – of the difference between each student-advisor pair rating. However, this test is much less powerful, so I opted to use the magnitude-based permutation test described above.

5.0 Conclusion

Overall, SOEST scholars significantly underrated their skills and performance compared to their advisors' ratings ($p=0.005$). Although the demographic analyses appeared to indicate that female ($D=0.24$; $p=0.005$) and NHPI ($D=0.24$; $p=0.01$) students significantly underrated themselves compared to their advisors, the intersectionality analysis revealed that these results were driven by the responses of female NHPI students ($D=0.36$; $p=0.02$) as opposed to female non-NHPI ($D=0.12$; $p=0.18$) and male NHPI students ($D=0.10$; $p=0.43$). This underrating may reflect low self-efficacy. If this interpretation is correct, then the SOEST Scholars program, and perhaps undergraduate research programs more generally, may need to consider focusing on intentionally building self-efficacy alongside cognitive skills. This appears to be particularly important for female NHPI students, and perhaps others at the intersection of underrepresented identities.

Appendix 1: Self-produced Matlab Code for Non-Parametric Permutation Test

```
function perm_mag(student, advisor, numperms)

observedmean = mean(advisor)-mean(student)

for n = 1:numperms
%Matrix of random zeros or ones
swap = randi([0, 1], [1, length(student)]);

for x = 1:length(swap)
%If swap(1) = 1, sswap(1) will be advisor(1), if swap(1) is 0, sswap(1) =
%student(1)
if swap(x) == 1, sswap(x) = advisor(x);
else sswap(x) = student(x);
end
end
%Converting sswap vector into column
studentswap = sswap';

%Same for advisor
for x = 1:length(swap)
if swap(x) == 1, aswap(x) = student(x);
else aswap(x) = advisor(x);
end
end

advisorswap = aswap';

permmean(n) = mean(advisorswap)-mean(studentswap);

end

%Plot histogram of the means
histogram(permmean)
xlim([-0.15 0.16])
xlabel('Mean Difference')
ylabel('Frequency')
hold
b = plot([observedmean, observedmean], ylim,'--r','LineWidth',5);

%P-value
p = length(find(abs(permmean)>=abs(observedmean)))/(numperms)
```

Literature Cited

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191.

Bandura, A. (1994). Self-efficacy. In V. S. Ramachaudran (Ed.), Encyclopedia of human behavior (Vol. 4, pp. 71-81). New York: Academic Press.

Bandura, A. (1997). Self-efficacy: The exercise of control, Worth Publishers.

Bandura, A., et al. (2001). Self-efficacy beliefs as shapers of children's aspirations and career trajectories. Child development 72(1): 187-206.

Bryan, M. T. (2004). An examination of Navajo cultural identity and its relationship to academic achievement (Doctoral dissertation, Brigham Young University).

Frawley, J., Ober, R., Olcay, M., & Smith, J. A. (2017). Indigenous achievement in higher education and the role of self-efficacy: Rippling stories of success. Perth, South Australia, Australia: National Centre for Student Equity in Higher Education (NCSEHE).

Intersectionality. (2020). Dictionary.com. Retrieved from https://www.dictionary.com/browse/intersectionality

Kortz, K. M., Cardace, D., & Savage, B. (2019). Affective factors during field research that influence intention to persist in the geosciences. Journal of Geoscience Education, 1-19.

Lent, R. w., Brown, S.D., Larkin, K.C. (1986). "Self-efficacy in the prediction of academic performance and perceived career options." Journal of counseling psychology 33(5): 265-269.

Lopatto, D. (2007). Undergraduate research experiences support science career decisions and active learning. CBE—Life Sciences Education 6(4): 297-306.

National Science Foundation (2014). Bachelor's degrees awarded, by citizenship, ethnicity, race, sex, and field. https://www.nsf.gov/statistics/2017/nsf17310/static/data/tab5-7.pdf

National Science Foundation (2015). Women and Minorities in the S&E Workforce. https://nsf.gov/statistics/2018/nsb20181/report/sections/science-and-engineering-labor-force/women-and-minorities-in-the-s-e-workforce

National Science Foundation (2019a). Research Experience for Undergraduates (REU) sites. from https://www.nsf.gov/crssprgm/reu/list_result.jsp.

National Science Foundation (2019b). Research Experiences for Undergraduates (REU) for Students. from https://www.nsf.gov/crssprgm/reu/.

Programme for International Student Assessment. (2012). What Students Know and Can Do. http://www.oecd.org/gender/data/notamathperson.htm

Ross, S. (2004). Introduction to Probability and Statistics for Engineers and Scientists. Academic Press.

Russell, S. H., Hancock, M.P., McCullough, J. (2007). "Benefits of undergraduate research experience." Science 316.

Spencer, S. J., et al. (1999). Stereotype threat and women's math performance. Journal of experimental social psychology 35(1): 4-28.

Stout, J. G., et al. (2011). STEMing the tide: Using ingroup experts to inoculate women's self-concept in science, technology, engineering, and mathematics (STEM). Journal of Personality and Social Psychology 100(2): 255.

U.S. Census Bureau (2015). "American Community Survey, Field of Bachelor's Degree for First Major Table S1502." from https://data.census.gov/cedsci/table?q=education&d=ACS%201-Year%20Estimates%20Subject%20Tables&table=S1502&tid=ACSST1Y2018.S1502&lastDisplayedRow=23.

UNESCO (2015). "World Inequality Database on Education." from https://www.education-inequalities.org.

Williams, M. M. and C. E. George-Jackson (2014). "Using and doing science: Gender, self-efficacy, and science identity of undergraduate students in STEM." Journal of Women and Minorities in Science and Engineering 20(2): 99-126.

Zusho, A., et al. (2003). "Skill and will: The role of motivation and cognition in the learning of college chemistry." International journal of science education 25(9): 1081-1094.