JSCEE
Journal for the Study of Cooperative and Experiential Education
JOURNAL FOR THE STUDY OF COOPERATIVE AND EXPERIENTIAL EDUCATION • VOL. 1 • SPRING 2025

Identity Implications for Self-Evaluation of Performance in a Project-Based Leadership Course

MELANIE BUFORD • UNIVERSITY OF MINNESOTA—TWIN CITIES

SHREYA SRIVASTAVA • UNIVERSITY OF MINNESOTA—TWIN CITIES

SHUAI AN • UNIVERSITY OF MINNESOTA—TWIN CITIES

ABSTRACT

Due to the historical and contemporary underrepresentation of women and non-White people in leadership roles and in academic leadership programs in the United States, gender and racial identity carry implications for leadership education and experience-based learning. Gender differences in self-esteem and self-evaluation have been observed in leadership contexts, including in college and university leadership classrooms. Racial differences in self-esteem exist in a number of contexts and non-White leaders continue to experience discrimination in leadership positions. This quantitative study assessed the impact of gender and race on student self-evaluation scores in an undergraduate leadership minor course. Student self-evaluation scores were compared to supervisor scores during a field experience project to determine accuracy of self-evaluation of performance. Quantitative methods were used to assess the significance of gender and race for self-evaluation accuracy. Results indicated that gender had a medium effect on accuracy and racial identity had a small effect (Cohen's D). Implications for inclusive leadership pedagogy and equitable use of self-evaluation assignments are discussed.

Keywords: self-esteem, inclusive pedagogy, leadership, experiential learning, gender, race

INTRODUCTION

In a variety of professional contexts, women and Black, Indigenous, People of Color (BIPOC) are underrepresented among leaders and on leadership teams. Only fifty-three of Global 500 companies were women in 2023. In the United States, organizational leaders overwhelmingly identify as White men. The first woman to join the Fortune 500 list of CEOs did not do so until 1972, and the first BIPOC (Black, Indigenous, People of Color) man did not join until 1981 (Hinchliffe, 2023; Hinchliffe & Abrams, 2023; United Nations Foundation, 2023). Moreover, women and BIPOC leaders consistently identify workplace cultures, bias, and discrimination as factors impacting their presence and success at work and in leadership roles (Lloyd, 2021; McKinsey and Company, 2022; Rosette et al., 2008). There exist significant pay gaps on the basis of gender and race and pay gaps for women and BIPOC women in leadership roles (Bureau of Labor Statistics, 2015; Hoisl & Mariani, 2017; Tsui, 1998). These differences represent an important consideration for academic leadership programs that seek to equitably prepare students to take on leadership roles in the workplace.

Though significant research has been done on gender differences in self-esteem in leadership settings, and studies have been conducted on race and self-esteem in higher education contexts, more work is needed to assess the impact of undergraduate leadership curriculum on students with different identities. This study examines identity differences in self-evaluation of project performance in an undergraduate leadership minor program at a large, public university in

the United States. The researchers aimed to assess the impact of gender and race on how accurately students evaluate their own performance on a leadership field project, in order to explore whether or not students with different identities evaluate their leadership performance differently.

BACKGROUND

Leadership studies is an interdisciplinary and evolving field of study. Overtime, Western academic conceptualizations of leadership have evolved, from centering theories focused on one man with innate leadership qualities to theories that apply a critical lens to the idea of leadership itself and reconceptualize traditional understandings of what effective leadership can look like. Traditional leadership theories have been designed principally by and for leaders who have identified as White men. The overrepresentation of White men in leadership positions underscores the need to explore the impacts of identity in leadership education and self-evaluation of performance (Dugan, 2024).

Gender, Leadership, and Self-Esteem

In a variety of contexts, American men and women tend to demonstrate differences in self-esteem and self-evaluation of performance, especially within fields in which women are underrepresented (Beyer, 1990; Feingold, 1994; Johnson, 1989; Kling et al. 1999; Major et al. 1999; Rentzsch et al. 2016). This has been found to be true in workplace settings, where men tend more often to step into leadership roles (Magee & Upenieks, 2019) and are more commonly found in leadership positions across industries in the United States (McKinsey and Company, 2022).

Eccles et al. (1990) suggested that women and men differ in their gender socialization and are impacted by gender stereotypes. This results in men and boys exhibiting stronger self-esteem than women and girls. In a quantitative study of German adolescents and adults, Rentzsch et al. (2016) found that women exhibited significantly lower self-esteem than men in a few specific dimensions: self-regard, social self-esteem, academic self-esteem, and physical self-esteem. Tsui (1998) suggested that self-confidence positively impacts leadership ability: "Because managerial skills imply the ability to manage, and thus to interact with others, it is reasonable to expect successful managers to be confident in their abilities to socialize with and lead others" (p. 365). Examining 941 individuals in business management careers, Tsui found that individuals who were more confident in their leadership abilities were earning a higher salary, and that men were more likely to have higher self-confidence in the field of management, and to earn more income. Schneer and Reitman (1994) found that pay disparities by gender grow over time as managers move up in their organizations. A Payscale report (2023) found that the gender pay gap still exists following the Covid-19 pandemic, even when we control for job titles.

Some work has been done examining self-evaluation and peer feedback in graduate management (MBA) programs. One study by Mayo et al. (2012) found that students who identified as women were more influenced by critical peer feedback than their peers who identified as men. Women in the study tended to rate themselves lower in self-evaluations of leadership ability after receiving peer feedback than men did. In undergraduate settings, some work exists on self-evaluation of leadership by gender. Sax (2008) suggested that men tend to report more confidence in their leadership abilities than women in college settings. Moreover, Chan and Drasgow (2001) found that self-ratings of leadership ability partially predicted motivation and interest in leadership after college. This mirrors research discussed earlier on the impact of self-confidence on leadership performance in work settings. Blaney (2020) conducted a study of leadership development in undergraduate computing programs and found that men reported more confidence in their leadership abilities than women. Blaney posited that the underrepresentation of women in computing, along with sexism in the wider culture, may have contributed to these findings. Blaney's work in particular has implications for the present study, though it did not take place in a leadership studies department.

Race, Leadership, and Self-Esteem

Research has suggested that Black individuals report higher self-esteem than White individuals, perhaps in part because self-esteem—and the solidarity that comes from membership in a disadvantaged group—may counter negative perceptions or bolster them against discrimination. Hispanic and Asian adults in the United States have been found to have slightly lower self-esteem than White Americans. Among minoritized groups, Black participants had the highest reported self-esteem, in spite of being the most stigmatized and devalued minoritized population. Self-esteem was found to be particularly high among college-aged Black participants, and among Black participants in the Southern United States, suggested that education and proximity to Black cultural centers may support higher self-esteem

(Twenge & Crocker, 2002; Zeigler-Hill et al., 2012).

One 1994 study (Crocker et al.) found that Black students seemed to be able to separate self-esteem in their own racial identity from how the public views their racial group. In contrast, Asian students in the study appeared to have their private feelings of self-esteem in their identity significantly impacted by their perceptions of public sentiment toward their racial group. White students also had a positive correlation between public and private self-esteem in their identity (Crocker et al., 1994). Looking at the intersection of race and gender, de Santibañes et al. (2023) suggested that Indigenous women, and potentially other minoritized leaders, were required to respond to resistance to and denial of their leadership identities and had to do additional identity reflection and work to reconcile these experiences. Greenstein (2000) found that self-efficacy scores may be lowest for Asian American undergraduates, compared with other BIPOC populations, due to the impact of model minority myth expectations.

Due to the deep entanglement of Western conceptualizations of leadership with European colonialism (Liu, 2019), race is a fraught topic in leadership studies. Research has found that the association between Whiteness and leadership has not dissipated over time (Petsko & Rosette, 2023). Harper and Kezar (2021) have critiqued the lack of attention to race, racism, White supremacy and systems of oppression in leadership theory and leadership studies. Wiborg (2020) and Williams et al. (2022) have described the centering and reproduction of Whiteness in leadership programs in higher education, while Ospina and Foldy (2009) have written about the lack of diversity among participants in leadership research. Race-inclusive and anti-racist leadership education requires confronting resistance in the classroom and in administrative spaces, and this resistance can harm BIPOC students (Wiborg, 2020).

PURPOSE OF THE STUDY

Due to the dynamic nature of contemporary careers, there is a growing need for workers to intentionally navigate changing circumstances (Savickas et al., 2009; Savickas, 2013). Self-esteem supports career adaptivity, the capacity to adapt to changing career circumstances (Rudolph et al, 2017; Savickas et al., 2009; Savickas, 2013). Identity differences in self-esteem and self-evaluation of performance warrant examination to better understand how to prepare aspiring leaders for career challenges, especially those who do not identify as White men.

We attempted to fill a gap in the literature by examining the impact of identity on self-evaluation in an undergraduate academic department focused on leadership—specifically, a department within which students are pursuing various majors, but all have declared a minor in leadership studies. Students in this program were diverse in skills, personalities, interests, and identities, and yet had each demonstrated a commitment to leadership through their academic minor. The research team gathered data from an undergraduate leadership minor course and compared students' individual self-evaluation scores to scores provided by students' supervisors. We sought to assess the premise that students who identified as women might rate themselves lower, on average, than students who identified as men. This finding would be consistent with Sax (2008) and Blaney's (2020) findings that women in college settings—including leadership settings—tend to rate themselves lower than men. In terms of racial identity, we sought to explore the relationship between race and self-evaluation accuracy in an undergraduate leadership context, since—to our knowledge—this type of study had not been done at the time of this writing.

This assessment assumed that the difference in self-evaluation accuracy would be reflected in the comparison between individual self-evaluation scores and scores provided by their supervisors. Our research question was: How do students in an undergraduate leadership field course rate their performance on a project, and how might this vary by gender and racial identity?

METHOD

Leadership Field Experience Course

Students in the study were enrolled in an undergraduate leadership minor program at a large, public university in the United States. The leadership minor program has four required core courses, which each involve activity-based learning on topics related to teamwork, emotional awareness, adaptive leadership, social change, and structural leadership. There is a significant focus on self- and other-awareness and learning to lead with ethical integrity and community stewardship. Inclusive leadership is a significant focus of the program.

For this study, the research team examined self-evaluations of leadership on a work-based learning project in the third required course, which includes a field work project. By the time students find themselves in this field experience course, they have completed at least two previous leadership

TABLE 1

Summary Statistics

(a) Continuous Variables

VARIABLE	OBS	MEAN	MEDIAN	STD. DEV.	MIN	MAX
SelfEval	72	92.58	95	9.42	30	100
SuperEval	72	87.42	94.5	16.67	25	100
GPA	72	3.52	3.58	0.35	2.56	4
Accuracy (A)	72	5.17	1	18.11	-60	70

(b) Categorical Variables

GENDER	FREQUENCY	PERCENT
Men	19	26
Women	53	74
Total	72	100
RACE	**FREQUENCY**	**PERCENT**
BIPOC	27	37.5
White	45	62.5
Total	72	100
CLASS YEAR	**FREQUENCY**	**PERCENT**
Sophomore	14	19.5
Junior	34	47
Senior	24	33.5
Total	72	100

courses (6 total credits) in the minor. The field experience course combines a 40-hour, group project conducted in partnership with a local mission-based organization, with coursework on social change leadership. Students engage with their classmates often throughout the 15-week semester, attend classes in person, complete assignments and reading, and practice collaborative communication and decision-making through their field project. This is also the class during which students formally declare the leadership minor, indicating their commitment to the program.

Participants

The research team collected data for 72 students enrolled in four sections of a 3-credit field experience course in the Spring 2024 semester. These 72 students represented nearly all of the population of the 73 students who completed the field experience course that term; one student record was removed from the data set because the student did not identify within the gender binary. Demographic data was gathered through the university's enrollment system. Table 1 provides descriptive information about the 72 students included in the study. Seventy-four percent of the students in the data set identified as women and 26% as men. Sixty-two-point-five percent of the sample identified as White, and 37.5% identified as BIPOC (Black, Indigenous, People of Color). Due to relatively small racial subgroup sizes, we

elected to represent racial identity as a binary: White or BIPOC. The demographic data collected also included class year and cumulative GPA. Students were enrolled in a wide variety of majors, including life sciences, social sciences, health sciences, business, sports management, communication, art and theatre, and others, along with a few students who were undecided. Each student planned to declare a minor in Leadership—and all were in sophomore, junior, or senior class years. Self-evaluation scores and individual supervisor scores for each student were collected and analyzed. The variable "SelfEval" represents students' self-evaluation of performance (as a percentage out of 100%). The variable "SuperEval" represents the percentage score (out of 100%) that students were provided by their project supervisor.

Measurement

This study aimed to assess the impact of self-reported gender identity on the accuracy of leadership studies students' self-evaluation of performance on a leadership project. Of specific interest was whether or not self-evaluation accuracy would differ by gender and racial identity. The study employed a quantitative research design, collecting the scores students assigned to themselves and the scores their supervisors assigned to them at the end of a 15-week leadership field project. Data was collected at the end of the academic term for all four course sections at the field experience level of the program. Participants represent a sample of students, in the sense that data for all students in the program that semester was collected, but this was only one leadership program in a specific context. We hoped to explore implications for the wider population of leadership studies undergraduate students as well and used linear regression in our analysis. Institutional Review Board permission was obtained before data collection.

Individual Self-Evaluation and Supervisor Evaluation

At the end of their time in the field experience course, students completed a self-evaluation of their leadership on the field project. This evaluation was worth approximately 10% of their course grade. Students were asked to reflect on their contributions to the project, and determine a grade for themselves—as a percentage score, with a maximum of 100% and a minimum of 0% (representing no contributions).

Students' group supervisors were asked to provide each individual student with a percentage score—just as students were asked to determine a score for themselves. These scores could range from 0-100%, representing the quality of their leadership contributions throughout the project. Students' individual supervisor-provided scores represented approximately 20% of their final course grades.

Determining Accuracy

Accuracy of student self-evaluation of performance was represented numerically as the difference between a students' self-evaluation score (s) and their supervisor-provided score (p). Beyer (1990) used a similar method to determine gender differences in self-evaluation of performance accuracy. Accuracy (A) = (s) – (p). If a student's accuracy (A) was found to be a negative value: $A < 0$, this indicated an underestimation of their own performance relative to their supervisor's evaluation of their performance. If a student's accuracy (A) was found to be a positive value: $A > 0$, this indicated an overestimation of their own performance relative to their supervisor's evaluation of their performance. If a student's accuracy (A) was 0: $A = 0$, that student assessed their performance consistently with their supervisor, indicating more accurate self-evaluation of performance. The greater the absolute value of students' accuracy (A), the less accurate their self-estimation of performance was relative to their supervisor's evaluation.

This study aimed for content validity (Messick, 1989) in that it was designed to deliver meaningful information about the target population: students who completed the leadership field experience course at a large, public university in Spring of 2024. We did not, therefore, use inferential statistics. Though the findings of this study may not be generalizable to other contexts, they do assess the impact of identity variables among students in this population. It is intended to illuminate the impact of identity on students' self-evaluation of leadership performance in this context in ways that may inform efforts to support students of minoritized identities within other leadership programs.

Data Analysis

Analysis of this data required statistical techniques common in educational research. First, descriptive statistics were used to summarize data and illuminate patterns. Then, simple statistical tests were used to assess the research question. We used Pearson Correlation (Cohen, 1988; Ritchey, 2008) to explore the relationship between cumulative GPA and evaluation accuracy (A). We also used linear regression to evaluate the variation in the response variable, accuracy (A),

to assess how much it might be attributable to gender and racial identity as explanatory variables (Richie, 2008). For Gender we set women as the reference group (women = 1, men = 0), for Racial Group we set White as the reference group (White = 0, BIPOC = 1). Though we were interested in the sample of students, we also wanted to explore identity dynamics in the wider population of leadership students.

As a final measure, we calculated Cohen's D in order to evaluate the magnitude of the mean difference in accuracy (A) among the four groups, women and men. Cohen's D was used as a standardized measure of the difference between the variables (Cohen, 1988; McGrath & Meyer, 2006).

TABLE 2

Gender and racial group differences in evaluation and accuracy

	WOMEN	MEN	BIPOC	WHITE
SelfEval				
Mean	92.08	94	90.85	93.62
Median	95	95	94	95
Std. dev.	10.66	4.34	13.21	6.08
Range	70	15	70	30
SuperEval				
Mean	89.62	81.62	88.52	86.76
Median	95	90	92	95
Std. dev.	13.88	22.03	13.82	18.28
Range	50	75	50	75
Accuracy (A) *				
Mean	2.45	12.74	2.33	6.87
Median	0	5	0	2
Std. dev.	16.08	21.55	17.22	18.6
Range	100	79	97	100

**Accuracy values may be positive, negative, or zero since each students' accuracy (A) is equal to "SelfEval"—"SuperEval"*

RESULTS

Descriptive statistics and effect size for each identity group in accuracy

Table 2 provides descriptive statistics for self-evaluation scores and supervisor evaluation scores for each identity group: women, men, BIPOC students, and White students. It also provides descriptive statistics for the accuracy of students' self-evaluation, by identity group. Both mean and median are provided to support a more holistic picture of the data, and a sense of more common responses.

This table suggests that though students belonging to each of the four identity groups provided similar self-evaluation scores, especially as determined by median values, supervisor evaluation scores varied more widely for men and White students. Women and BIPOC students had mean and median accuracy scores closer to zero, when compared with men and White students. For gender, the difference in mean accuracy was 10.29 (women were more accurate). For racial group, the difference in mean accuracy was 4.54 (BIPOC students were slightly more accurate).

The distributions of accuracy scores by gender and racial group are represented in Figures 1 and 2. Negative values indicate when students underestimated their scores. Positive values indicate when students overestimated their scores. Visually, we did not find differences in accuracy distribution between men and women.

FIGURE 1

Accuracy distribution by gender

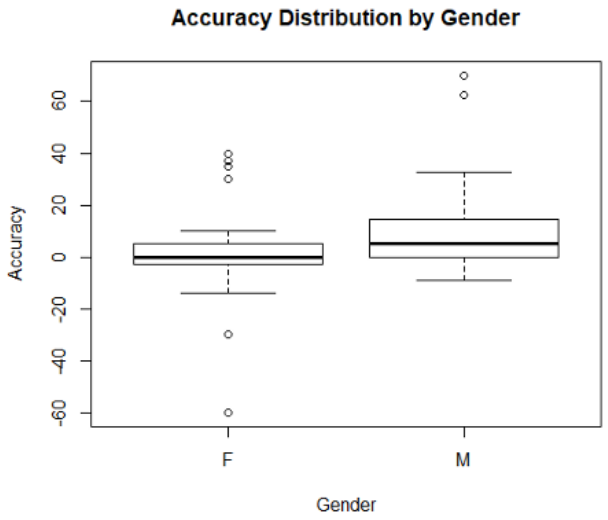

FIGURE 2

Accuracy distribution by racial group

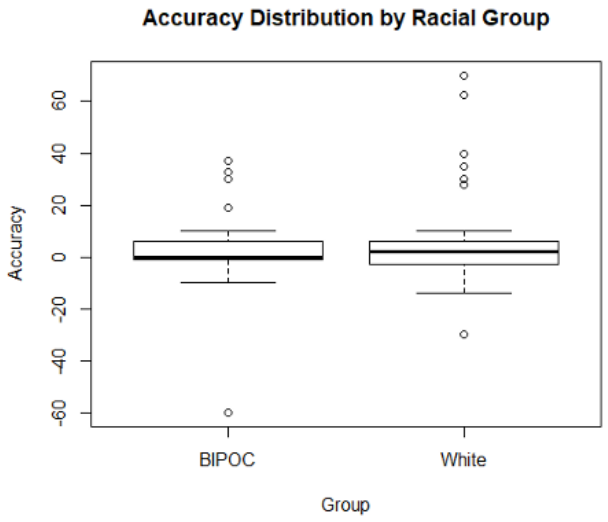

Figure 1 shows the spread of accuracy values by gender. "M" represents the accuracy scores for men in the sample, and "F" represents the accuracy scores for women. Outliers are represented by the dots at the top and bottom of the two plots. It is possible in this plot to see that men generally rated themselves more highly compared with women, and that women's self-evaluation scores more closely aligned with their supervisor scores.

Figure 2 shows the spread of accuracy values by racial group. It is possible in this plot to see that while the mean accuracy values for White and BIPOC students in the sample are similar, there is a small observable difference in outliers. White students had a few outliers with high scores relative to the mean, while one BIPOC outlier score was particularly low.

Cohen's D

Cohen's D was calculated to assess the magnitude of the difference between means for gender and racial group. Table 3 provides Cohen's D values for the difference in means for women and men, and BIPOC and White students. Cohen's D provides an indication of the effect size when comparing the mean accuracy for each group.

TABLE 3

Cohen's D

Cohen's d*	95% CI
Gender	
-0.58**	[-1.11, 0.05]
Racial Group	
-0.25	[-0.73, 0.23]

**Estimated using pooled standard deviation*

***Where d is between 0.2 and 0.5 indicates a small effect size and where d is between 0.5 and 0.8 indicates a medium effect size.*

Confidence intervals are also provided. Effect sizes were drawn from Sullivan and Feinn (2012). The effect size for gender was medium (-0.58) and the effect size for racial group was small (-0.25).

Pearson correlation and linear regression

To complete the picture of how the variables of interest in this study related to each other, Pearson correlation was calculated and simple linear regression analysis was performed. Pearson correlation was determined for the relationship between cumulative GPA and self-evaluation

accuracy. This is reported in order to provide a stronger picture of the data. The Pearson correlation was small: 0.01886. Thus, the linear relationship between accuracy and cumulative GPA is almost non-existent, within the sample.

Simple linear regression was used to explore the relationship between accuracy and identity variables—gender and racial group. When running a simple linear regression of accuracy vs. gender, the regression coefficient of -10.28 (lower mean accuracy percentage for women) had some evidence of statistical significance (standard error = 4.72, t = -2.18, and p-value = 0.31, on 70 degrees of freedom). The corresponding [R-squared] was 0.06.

After running a simple linear regression of accuracy vs. racial group, we found a regression coefficient of -4.53 (lower mean accuracy percentage for BIPOC), which had no evidence of statistical significance beyond the sample (standard error = 4.41, t = -1.03, and p-value = 0.31, on 70 degrees of freedom). The corresponding [R-squared] was 0.01.

The regression coefficients for each variable indicated some association between gender and accuracy and race and accuracy for the sample. The regression coefficient for gender was larger, indicating a stronger relationship. While the p-value for gender was less than 0.05, the p-value for racial group was not. Thus, *gender appeared to have a statistically significant effect on accuracy* in self-evaluation within the sample, and potentially in other similar contexts given the p-value. Racial group did not appear statistically meaningful beyond the sample, though there was a small observable difference in mean accuracy between BIPOC and White students within the sample.

DISCUSSION

After examining the influence of gender and race on self-evaluation accuracy relative to supervisor evaluation, we found that gender appeared to significantly affect accuracy, while racial group did not appear to significantly affect accuracy beyond the sample. The linear relationship between accuracy and cumulative GPA was nearly non-existent. These findings reflect and build upon some of the existing literature on self-evaluation and identity.

The significant relationship between accuracy and gender was consistent with our assumptions based on previous research (Blaney, 2020; Sax, 2008). Research conducted by Berg et al. (2006) suggested that women tend to be more conscious of how others may perceive their attributions of success. In other words, they may seek to demonstrate modesty in explaining their success. The authors observed a self-derogatory attitude among participants. Given that women in this study knew that their self-evaluation would be visible to the teaching team, and would be shared with their teammates and supervisor in final reflection meetings, these participants might have chosen more modest scores. If true, this might imply that women's self-evaluation would be more accurate if they were able to rate themselves privately.

The fact that accuracy and race were not found to have a significant explanatory relationship beyond our sample carries several potential implications. Previous research has found significant differences in self-esteem by racial subgroup (Twenge & Crocker, 2002; Zeigler-Hill et al., 2012). The fact that some minoritized racial groups were found to have higher self-esteem than White groups, and some were found to have lower self-esteem, may have produced a cancelling-out effect in our analysis on self-evaluation of leadership. In fact, there were small observable differences by racial subgroup in our study. Though we chose not to analyze those differences due to small sample sizes, there were small observable differences between the average accuracy of students who identified as Black and students who identified as Asian, consistent with the results of Twenge and Crocker's (2002) study. Of course, Twenge and Crocker looked at self-esteem, not evaluation of performance, a key difference. Self-esteem in general may differ compared with evaluation of performance on a specific project.

Nonetheless, Black students' average accuracy score in our study was higher than White students, while Asian students' average accuracy was the lowest of any demographic in our sample, a negative value. It is also true that Black students' supervisor scores were lower, on average, than any other group. Racial bias was not explored within the scope of this study. Thus, we cannot draw conclusions about self-esteem dynamics by racial subgroup. These subgroup differences do warrant further study in leadership education contexts.

Though women represented a majority in our sample, they nonetheless tended to rate their contributions lower than men. Blaney's (2020) examination of gender and leadership development in undergraduate computing found a significant difference in computing leadership confidence among women in her study, compared to men. Blaney attributed this, in part, to the lack of representation in the field of computing. This was reinforced by her finding that women had more confidence in their general leadership abilities related to their discipline-specific leadership

abilities. The disciplinary context may have impacted women's leadership confidence, according to Blaney. Though women are a growing minority in MBA programs, they still do not represent the majority (Reilly, 2021). Thus, one might expect women in an undergraduate leadership program to evaluate their work more critically than men, due to their underrepresentation. Yet, since women were the majority in the population we assessed, this may suggest that underrepresentation itself is not enough to explain the lower self-evaluations among women in the study.

It is also worth noting that Blaney's (2020) research examined women's ratings of their leadership confidence, not their self-evaluation of actual performance on a domain-specific project. McKee et al. (2018) found that women's self-ratings of leadership were less inflated relative to other ratings than their participants who were men, suggesting that women may have higher self-awareness of their leadership abilities. This is reflected in our study, since women's accuracy was closer to 0 (more accurate) in our study. It may therefore be the case that, rather than women underrating their leadership, our findings suggest that men tended to overrate their leadership performance. It might be that women held a more realistic view of their performance, or a more accurate understanding of how their supervisor might evaluate their work.

Implications for Future Research and Practice

This study contributes to the literature on gender and racial differences in self-evaluation of leadership in undergraduate settings, and carries implications for leadership education research and practice. Blaney's (2020) findings suggested that, for women, feeling a sense of belonging in the computing field was a strong predictor of confidence in leadership ability. Within the leadership studies minor in this study, significant curricular attention was paid to connection and belonging for students. One core principle of the program was "connection before content." In the first and second courses students completed before they reached the field experience course, they spent the beginning of each class session cultivating connection and belonging through structured personal storytelling and other community-building activities. This aspect of the curriculum likely promoted a stronger sense of belonging among students, not just in the course and the minor, but in the field of leadership itself. The program advocated for an "everyone leads" mentality, to counter gender and racial imbalances in the wider field. It may be that the emphasis on belonging in the program supported students' self-esteem and encouraged positive self-evaluation. As we have seen, study findings demonstrate that women's average self-evaluation score was still slightly higher than their supervisor evaluation, despite other research suggesting that women possess self-derogatory beliefs about their success (Herbst, 2020). It is possible that the strong emphasis on belonging supported students' positive self-evaluations, and women's more accurate self-evaluation. This contradicts the assumption that women are less accurate than men in self-perception accuracy, at least in this context.

Regardless, inclusive leadership curricula should represent a goal for leadership programs, at a minimum. White et al. (2021) described best practices for addressing equity in the classroom in the field of chemistry. White et al. provided evidence-based practices for inclusive curricula, and we have described a few below (2021, p. 332):

- Fostering a sense of belonging
- Validating students' scientific identities
- Allowing students to make mistakes
- Cultivating relationships
- Employing active learning and group work

White et al. suggested these practices for reducing equity gaps with respect to identity. The shift to digital learning environments, White et al. suggested, requires even more attention to these practices.

The nature of the field experience curriculum, in addition to the emphasis on relationships and belonging, is focused on project-based learning and group work. Building relationships and embracing uncertainty and failure are written into rubrics in the class, including the self-evaluation. Adaptability and managing uncertainty are also important career capacities (McGowan & Shipley, 2020). The presence of these aspects of inclusive pedagogy in leadership and career education curricula may support less identity difference in self-evaluation of performance. Newer leadership theories suggesting more anti-racist and liberatory practices also hold important value for leadership education (Dugan, 2024; Harper & Kezar, 2021) and may strengthen self-awareness for students with dominant identities.

Another important practice goal in this study was to explore whether or not a self-evaluation assignment in a leadership course, which is incorporated into a students' course grade, may thus affect their final grade in the course. In other words, if women and minoritized students give themselves lower scores, do they therefore have lower course grades?

This is an ethical consideration for educators in these spaces. Though this study did not find significant racial differences in self-evaluation—at least for BIPOC students as a group—it did find significant gender differences, which also include differences for BIPOC women. Educators who ask students to self-evaluate, without awareness of identity differences in self-evaluation, may risk creating identity differences in course grades. Strategies for counteracting identity effects—such as letting students know that men tend to rate their performance more highly—may prompt all students to reflect more deeply on their work. Though self-evaluation of performance provides students with the ability to contribute to their course grades in experiential learning models, and thus can be seen as a form of empowerment, it may also reproduce inequality if not conducted mindfully. Future research might explore these implications.

LIMITATIONS

One limitation of the study is the source of our gender and racial group information, which came from the university's enrollment system, rather than directly from students themselves (with the exception of one student who volunteered their gender identity). Also, in a study on identity differences, it is important to note that the research team fully recognizes and upholds the understanding that gender and race are socially -constructed identities, and that gender is not binary. For the purposes of this study, we do use a binary lens in order to explore the impact of socialized gender in leadership education. Similarly, due to our sample size, we chose not to assess racial difference in accuracy by BIPOC subgroup (Black, Latine, etc.). Instead, we created binary racial groupings, BIPOC and White. This was a significant limitation, especially because studies have indicated differences in self-esteem among BIPOC subpopulations (Twenge & Crocker, 2002; Zeigler-Hill et al., 2012).

Another important limitation lies in the fact that this study did not explore gender or racial bias and discrimination among supervisors. Thus, students' accuracy value represented the alignment between their self-evaluation and the supervisor's evaluation, rather than a truly objective measure of their performance. Further research should explore bias among supervisor evaluations in addition to student perceptions. Another limitation was the focused nature of the inquiry during a specific period of time with a specific course level at a large, public university. We hope that more inquiry will explore the evaluation accuracy of women and minoritized students in other leadership studies contexts, especially as the racial literature in this area is sparse, and add a qualitative lens to a limited empirical picture.

CONCLUSION

This study attempted to fill a gap in the literature on the impact of identity on self-evaluation in undergraduate leadership programs. Findings suggested significant gender differences in self-evaluation relative to supervisor evaluation, where women more accurately evaluated their leadership performance. We did not find significant racial differences for BIPOC students—as a total group—relative to White students. This work contains implications for inclusive pedagogical approaches in gender and racially imbalanced fields, such as leadership education. Curricular focus on relationships, disciplinary belonging, identity representation, active learning, and embracing mistakes (White et al., 2021) may partially address equity gaps in leadership education, and liberatory approaches may further close equity gaps (Dugan, 2024; Harper & Kezar, 2021). At a minimum, these approaches may ameliorate the impact of gender and racial stereotypes in the classroom and in experiential learning settings. We hope that future research will continue to expand upon these themes, addressing identity-based equity issues in higher education.

REFERENCES

Berg, J., Stephan, W, & Dodson, M. (2006). Attributional modesty in women. *Psychology of Women Quarterly*, 5(s5), 711–727. https://www.doi.org/10.1111/j.1471-6402.1981.tb01095.x.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology*, 59(5), 960–970. https://www.doi.org/10.1037/0022-3514.59.5.960

Blaney, J. M. (2020). Gender and leadership development in undergraduate computing: A closer look at women's leadership conceptions. *Computer Science Education, 30(4)*, 469-499. https://doi.org/10.1080/08993408.2020.1816769

Bureau of Labor Statistics, U.S. Department of Labor. (2015). Highlights of women's earnings in 2014. https://www.bls.gov/opub/reports/womens-earnings/archive/highlights-of-womens-earnings-in-2014.pdf

Chan, K. Y., & Drasgow, F. (2001). Toward a theory of individual differences and leadership: Understanding the motivation to lead. *Journal of Applied Psychology, 86(3)*, 481. https://doi.org/ 10.1037/0021-9010.86.3.481

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Lawrence Erlbaum.

Crocker, J., Luhtanen, R., Blaine, B., & Broadnax, S. (1994). Collective self-esteem and psychological well-being among White, Black, and Asian college students. *Personality & Social Psychology Bulletin, 20*(5), 503–513. https://doi.org/10.1177/0146167294205007

de Santibañes, M., Ospina, S. M., Lee, S., Santamaria, A., Evans, M. M., Muelas, D., & Guerrero, N. (2023). The dialectics of leadership identity construction: Case studies from Indigenous women leaders. *Leadership (London, England), 19*(4), 366–390. https://doi.org/10.1177/17427150231169554

Dugan, J. P. (2024). Leadership theory: Cultivating critical perspectives (2nd ed.). Wiley. p. xv.

Eccles, J. S., Jacobs, J. E., & Harold, R. D. (1990). Gender role stereotypes, expectancy effects, and parent's' socialization of gender differences. *Journal of Social Issues, 46*(2), 183-201. https://doi-org.ezp3.lib.umn.edu/10.1111/j.1540-4560.1990.tb01929.x

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin*, *116*(3), 429-456.

Greenstein, B. (2000). *Students perception of the racial climate on campus and in the classroom and the relationship with academic self-efficacy and academic and intellectual development* (Order No. 9967904). ProQuest Dissertations & Theses Global.

Harper, J., & Kezar, A. (2021). *Leadership for liberation: A leadership framework and guide for student affairs professionals.* USC Pullias Center for Higher Education.

Herbst, T. H. H. (2020). Gender differences in self-perception accuracy: The confidence gap and women leaders' underrepresentation in academia. *SA Journal of Industrial Psychology*, 46. https://doi.org/10.4102/sajip.v46i0.1704

Hinchliffe, E. & Abrams, J. (2023, October 3). The Fortune 500 adds a second Black woman CEO in SAIC's Toni Townes Whitley. https://fortune.com/2023/10/03/fortune-500-black-female-ceo-toni-townes-whitley-thasunda-brown-duckett/

Hinchliffe, E. (2023, June 5). Women CEOs run 10.4% of Fortune 500 companies. A quarter of the 52 leaders became CEO in the last year. https://fortune.com/2023/06/05/fortune-500-companies-2023-women-10-percent/

Hoisl, K. & Mariani, M. (2017). It's a man's job: Income and the gender gap in industrial research. *Management Science, 63(*3), 766-790. https://doi.org/10.1287/mnsc.2015.2357

Johnson, J. (1989, February1-5). Effects of successful female role models on young women's attitudes toward traditionally male careers [Conference Report]. Annual Meeting of the Association for Educational Communications and Technology, Dallas, TX, United States.

Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125*(4), 470-500.

Liu, H. (2019). Redoing and abolishing Whiteness in leadership. In B. Carroll, J. Firth, & S. Wilson (Eds.), *After leadership*, 101-114. Routledge.

Lloyd, C. (2021, January 21). One in four Black workers report discrimination at work. Gallup - News. https://news.gallup.com/poll/328394/one-four-Black-workers-report-discrimination-work.aspx

Magee, W., & Upenieks, L. (2019). Gender differences in self-esteem, unvarnished self-evaluation, future orientation, self-enhancement and self-derogation in a U.S. national sample. *Personality and Individual Differences*, 149, 66-77. https://doi.org/10.1016/j.paid.2019.05.016

Major, B., Barr, L., Zubek, J., & Babey, S. H. (1999). Gender and self-esteem: A meta-analysis. In W. B. Swann Jr., J. H. Langlois, & L. A. Gilbert (Eds.). *Sexism and stereotypes in modern society: The gender science of Janet Taylor Spence* (pp. 223–253). American Psychological Association.

Mayo, M., Kakarika, M., Pastor, J. C., Brutus, S. (2012). Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in MBA teams. *Academy of Management Learning & Education, 11*(4), 631-652. https://doi.org/10.5465/amle.2010.0069

McGowan, H. E. & Shipley, C. (2020). *The adaptation advantage: Let go, learn fast, and thrive in the future of work.* Wiley.

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: The case of r and d. *Psychological Methods, 11*(4), 386-401.

McKee, R. A., Lee, Y., Atwater, L., & Antonakis, J. (2018). Effects of personality and gender on self-other agreement in ratings of leadership. *Journal of Occupational and Organizational Psychology, 91*, 285-315. https://www.doi.org/10.1111/joop.12209

McKinsey and Company. (2022, October 18). Women in the workplace 2022 (Report). https://www.mckinsey.com/featured-insights/diversity-and-inclusion/women-in-the-workplace

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). American Council on Education.

Ospina, S., & Foldy, E. (2009). A critical review of race and ethnicity in the leadership literature: Surfacing context, power and the collective dimensions of leadership. T*he Leadership Quarterly, 20*(6), 876–896. https://doi.org/10.1016/j.leaqua.2009.09.005

Payscale. (2023). 2023 gender pay gap report. https://www.payscale.com/research-and-insights/gender-pay-gap/

Petsko, C. D., & Rosette, A. S. (2023). Are leaders still presumed white by default? Racial bias in leader categorization revisited. *Journal of Applied Psychology, 108*(2), 330–340. https://doi.org/10.1037/apl0001020

Reilly, K. (2021, November 12). A record number of women are enrolled in top business schools, but men still dominate MBA programs. *Time.* https://time.com/6116802/women-business-school-record/

Rentzsch, K., Wenzler, M. P., & Schütz, A. (2016). The structure of multidimensional self- esteem across age and gender. *Personality and Individual Differences, 88*, 139-–147. https://doi.org/10.1016/j.paid.2015.09.012

Ritchey, F. (2008). *The statistical imagination: Elementary statistics for the social sciences* (2nd Edition ed.). McGraw-Hill.

Rosette, A. S., Leonardelli, G. J., & Phillips, K. W. (2008). The White standard: Racial bias in leader categorization. *Journal of Applied Psychology, 93*(4), 758-–777. https://doi.org/10.1037/0021-9010.93.4.758

Rudolph, C. W., Lavigne, K. N., & Zacher, H. (2017). Career adaptability: A meta-analysis of relationships with measures of adaptivity, adapting responses, and adaptation results. *Journal of Vocational Behavior*, 98, 17–34. https://doi.org/10.1016/j.jvb.2016.09.002

Savickas, M. L. (2013). Career construction theory and practice. In S. D. Brown & R. W. Lent (Eds.), *Career development and counseling: Putting theory and research to work* (2nd ed., pp. 147-183). Wiley.

Savickas, M. L., Nota, L., Rossier, J., Dauwalder, J.-P., Duarte, M. E., Guichard, J., Soresi, S., Van Esbroeck, R., & van Vianen, A. E. M. (2009). Life designing: A paradigm for career construction in the 21st century. *Journal of Vocational Behavior*, 75(3), 239-250. https://doi.org/10.1016/j.jvb.2009.04.004

Sax, L. (2008). The gender gap in college: Maximizing the developmental potential of women and men. Jossey-Bass.

Schneer, J. A. & Reitman, F. (1994). The importance of gender in mid-career: a longitudinal study of MBAs. *Journal of Organizational Behavior, 15*(3), 199-207. https://www.jstor.org/stable/2488445

Sullivan, G. M., & Feinn, R. (2012). Using effect size-or why the p value is not enough. *Journal of Graduate Medical Education, 4*(3), 279–282. https://doi.org/10.4300/JGME-D-12-00156.1

Tsui, L. (1998). The effects of gender, education, and personal skills self-confidence on income in business management. *Sex Roles, 38*(5-6), 363-373. https://doi.org/10.1023/A:1018701704103

Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: Meta-analyses comparing Whites, Blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000). *Psychological Bulletin, 128*(3), 371-408. https://doi.org/10.1037/0033-2909.128.3.371

United Nations Foundation. (2023). 50 of the world's most sexist laws: A snapshot of gender inequality. https://unfoundation.org/what-we-do/issues/girls-and-women/50-of-the-worlds-most-sexist-laws-a-snapshot-of-gender-inequality/?gad_source=1&gclid=CjwKCAjwvIWzBhAlEiwAHHWgvcRNoTDmSf-Ntc7xOaWdcER59OCkLRFPL_oJoYAkqc6Rnyh-FMIm3rsxoCU2sQAvD_BwE

White, K., Vincent-Layton, K., & Villarreal, B. (2021). Equitable and inclusive practices designed to reduce equity gaps in undergraduate chemistry courses. *Journal of Chemical Education, 98*(2), 330-339. https://doi.org/10.1021/acs.jchemed.0c01094

Wiborg, E. R. (2020). *A critical discourse analysis of leadership learning* [Doctoral dissertation], Florida State University.

Williams, B. M., Hotchkins, B. K., & Evans, M. E. (2022). The permeation of Whiteness in student leadership organizations. In Z. Foste, & T. L. Tevis (Eds.), *Critical Whiteness praxis in higher education* (pp. 188-204). Stylus.

Zeigler-Hill, V., Wallace, M. T., & Myers, E. M. (2012). Racial differences in self-esteem revisited: The role of impression management in the Black self-esteem advantage. *Personality and Individual Differences, 53*(6), 785–789. https://doi.org/10.1016/j.paid.2012.06.007