

The Gender Minority Gaps in Confidence and Self-Evaluations*

Billur Aksoy[†] Christine L. Exley[‡] Judd B. Kessler[§]

October 6, 2022

Abstract

An increasing share of the population identifies as something other than male or female. Yet, we know very little about the economic preferences and beliefs of gender minorities. In this paper, we document a “gender minority gap” in confidence and self-evaluations. Middle and high school students complete a math and science test. Relative to equally performing male and female peers, gender diverse students believe they scored lower and provide more pessimistic subjective self-evaluations of their performance, even after they learn their score. These findings may help to improve our understanding of the socioeconomic disparities that gender minorities experience.

Keywords: gender identity, confidence, self-evaluation

JEL codes: C91, D91, J16

*This paper was supported by Character Lab and facilitated through the Character Lab Research Network, a consortium of schools across the country working collaboratively with scientists to advance scientific insights that help kids thrive. We would like to thank Christopher Carpenter and Dario Sansone for their helpful comments and feedback.

[†]Department of Economics, Rensselaer Polytechnic Institute, Email: aksoyb3@rpi.edu.

[‡]Harvard Business School, Email: clexley@hbs.edu.

[§]The Wharton School, University of Pennsylvania, Email: judd.kessler@wharton.upenn.edu.

1 Introduction

A sizeable share of the population identifies as transgender, nonbinary, or gender diverse (i.e., their sex assigned at birth differs from their current gender identity). At present, about 1–2% of U.S. adults do not identify as cisgender, and there are generational differences: younger adults are more likely to report non-cisgender identity (Jones, 2022; Brown, 2022). For example, while 5.1% of U.S. adults younger than 30 identify as transgender or nonbinary, this rate is 1.6% for 30- to 49-year-olds and 0.3% for those 50 and older (Brown, 2022). This pattern suggests that the share of gender minorities is likely to increase as the population ages.¹ Given these facts, understanding the economics preferences, beliefs, and outcomes of gender minorities is increasingly important.

In terms of outcomes, a small but growing field of LGBTQ+ economics shows that gender minorities have significantly worse economic outcomes compared to the general population (Badgett, Carpenter and Sansone, 2021; Carpenter, Eppink and Gonzales, 2020; Carpenter, Lee and Nettuno, 2022). They are more likely to be unemployed and to be in low-income households and less likely to have health insurance coverage (Badgett, Carpenter and Sansone, 2021). They also have worse educational outcomes: they have lower GPAs, are less likely to graduate high school, and are less likely to have attended or graduated from college (Meyer et al., 2017; Downing and Przedworski, 2018; Sansone, 2019).

In terms of factors that could contribute to these disparities in economic and educational outcomes of gender minorities, recent research documents differences in the treatment of gender minorities (Badgett, Carpenter and Sansone, 2021). Indeed, gender minorities are often described as a “minority within a minority,” and the treatment of—as well as the views toward—gender minorities are found to be worse relative to sexual minorities (Lewis et al., 2017, 2022; Aksoy, Carpenter and Sansone, 2022). In addition, some work points towards the potential role of lower educational expectations, motivation, school engagement, and sense of belonging (Sansone, 2019).

Yet, there is clearly more work to be done to understand what factors may potentially contribute to these disparities and to better understand the economic preferences and beliefs of gender minorities. In this paper, we seek to investigate the beliefs gender minorities hold about their own performance. Specifically, motivated by the large literature on “gender gaps” in confidence and in self-evaluations—a literature that shows women are typically less confident than men in their performance and ability, particularly in stereotypically male-typed domains such as math and science (Lundeberg, Fox and Punčohar, 1994; Niederle

¹Gender minorities include, but are not limited to, those who identify themselves as transgender, nonbinary, genderqueer, or gender diverse. And, sexual minorities include, but are not limited to, those who are gay, lesbian, or bisexual.

and Vesterlund, 2007; Coffman, 2014; Bordalo et al., 2019; Coffman, Collis and Kulkarni, 2019; Exley and Kessler, 2022)—this paper investigates whether we also observe a “gender minority gap” in these beliefs and subjective evaluations.² We investigate whether there are differences in confidence (measured by beliefs about absolute performance) and differences in subjective self-evaluations about performance (both when absolute performance is unknown and when it is known) between those who we classify as gender diverse and those who identify as either male or female.³ To answer these research questions, we collect data from an experiment and supplement it with individual-level administrative data on our study subjects.

One challenge with conducting research on gender diverse individuals relates to data on gender identity either being missing or recorded as binary in administrative records. We overcome this challenge by allowing study subjects to self-identify their gender as something other than male or female in a survey as part of our experiment. Another challenge relates to the difficulty in recruiting a sufficient number of gender diverse individuals to participate in research, given their minority status, particularly among older populations. We overcome this challenge by recruiting a large and young population. Specifically, we partnered with the Character Lab Research Network to recruit 10,637 students who were in grades 6–12 in the United States to complete our experiment; our primary classification identifies 1.6% of this population as gender diverse.

Our experiment proceeds in six stages. First, participants complete a test with 10 math and science questions. Second, we elicit participants’ *beliefs* about their absolute performance by asking them to guess how many questions, out of the 10 questions, they got right on the test, which serves as a measure of *confidence*. Third, we elicit participants’ *uninformed self-evaluations* via four questions that ask them to provide subjective evaluations of their performance on the test, such as by indicating their level of agreement with the statement, “I performed well on the test.” Fourth, we inform participants of the number of questions they answered correctly on the test. Fifth, we elicit participants’ *informed self-evaluations* via the same four questions that ask them to provide subjective evaluations of their performance now that they know their score on the test. Finally, we ask them to complete a short follow-up survey that gathers demographic information, including on gender identity.

The follow-up survey collects data on their gender identity by allowing participants to

²For related reviews, see also Niederle and Vesterlund (2011); Blau and Kahn (2017); Hernandez-Arenaz and Iriberry (2019); Niederle (2016).

³As will be explained in much greater detail, we use different definitions to classify individuals as being gender diverse, but we generally rely on individuals reporting that their gender is “other” when given the option between selecting “male,” “female,” or “other.” Thus, diverse students in our data include those who identify themselves as something other than male or female, such as transgender, non-binary, or gender-fluid.

select “male,” “female,” or “other.” If they select “other,” they have the option to provide free-response text about their gender identity. In our paper, we refer to those who selected “other” as gender diverse and show that our results are robust to various subgroups of gender diverse students, defined according to the free response text that they may have provided on their gender identity. Our results are also robust to the use of gender data obtained in a Character Lab demographics survey that was conducted before our experiment.

We compare gender diverse students to female and male students. In our administrative data on school performance, gender diverse students earn similar GPAs as male students, but female students earn higher GPAs than both gender diverse students and male students. On the math and science test in our experiment, gender diverse students again perform equally to male students, but female students perform worse than both gender diverse students and male students. This pattern of performance might suggest that gender diverse students will look more similar to male students when it comes to performance beliefs and subjective self-evaluations. Our confidence and self-evaluation data paint a very different picture.

When compared to equally-performing male or female counterparts, we observe a gender minority gap: Gender diverse students are less confident in their absolute performance (i.e., they believe they answered fewer questions correctly), and they provide less favorable self-evaluations. They indicate less agreement with the statement that they “performed well” on the test, they report being less inclined to take a class that involves the math and science topics covered on the test, and they believe they would be less likely to succeed in such a class. These differences in self-evaluations persist when students are informed about their test performance, revealing the gender minority gap in self-evaluations is not due to the gap in confidence in absolute performance.

Confidence can shape educational and career choices; individuals are regularly asked to provide self-evaluations of their abilities and performance, which can impact their educational and labor market outcomes. Consequently, the gender minority gaps in confidence and in self-evaluations that we observe are important to document and are worthy of further exploration. For example, future work may investigate why these differences in confidence and self-evaluations exist in the first place. In light of previous research on how social norms impact economic behavior ([Bicchieri, 2005](#); [Field et al., 2021](#); [Pande and Roy, 2021](#)), it is possible that these norms are different for gender minorities, which could impact their beliefs and behavior. We return to avenues for future work in our Conclusion.

2 Design

Our results come from a partnership with the Character Lab Research Network (CLRN).⁴ CLRN helped us recruit 10,637 students in the U.S. who were in grades 6–12 to participate in a short study taken during school. Details about the experiment, including screenshots, can be found in Appendix B. Data from the experiment was also analyzed in Exley and Kessler (2022).⁵

After consenting to participate in the experiment, participants are asked to answer 10 math and science questions from the Armed Services Vocational Aptitude Battery.⁶ Each question appears on a separate page, and students have 30 seconds to answer each question (see Appendix Figure B.2 for an example question). We request that participants try their best when answering these questions, but there are no financial incentives offered as part of this study. After completing the test, participants are asked to answer questions on five additional pages. Responses to these questions are the measures of confidence and self-evaluation that we analyze in this paper.

On the first page, we elicit participants’ beliefs about their absolute performance on the test by asking them how many questions out of 10 they think they got right. This gives us a measure of their confidence in their absolute performance.

On the second page, we elicit participants’ uninformed self-evaluations by requesting that they answer a free response question about their performance and then four self-evaluation questions. Like Exley and Kessler (2022), we focus on the quantitative answers to the self-evaluation questions. In the *performance-bucket* question, participants are asked to indicate how well they think they performed on the test by choosing from the following list of seven

⁴The following has been taken from CLRN website and explains the data collection process in more detail: “This investigation was part of a larger data collection effort that included a variety of studies designed by scientists affiliated with Character Lab Research Network (CLRN)...This study was conducted on school computers during class time in participating schools over the course of a two- to three-week testing window. On a predetermined testing day, a teacher proctor at each school administered the CLRN research activities to students. To introduce the study, teachers read a script that explained to students that all research activities were part of an educational research initiative at their school, that participation was voluntary and they were not being graded, and that teachers would not see their answers. Teachers also instructed students to focus on their own computers and (if relevant) not to look at classmates’ screens. Upon logging into the CLRN platform, all students first viewed an assent screen that reiterated this information and, in addition, explained that parents would not see their responses and that their names and any other unique identifying information would not be shared with researchers. Students who agreed to participate were then directed to the survey.”

⁵See Section V of Exley and Kessler (2022). That paper was focused on gender gaps in self-evaluations and only compared men and women, relying on administrative data that identified each student as one or the other. In this paper, we instead explore students’ self-reported gender elicited during our experiment. We additionally explore how self-reported gender correlates with administrative data on school performance.

⁶Math and science tests are typically considered male-typed tasks. Prior research finds that gender gaps in confidence and self-evaluations are particularly pronounced in male-typed tasks, although less is known about how gender diverse individuals act across different types of environments.

adjectives: terrible, very poor, poor, neutral, good, very good, and exceptional. In the remaining three self-evaluation questions, participants are asked to indicate their agreement—on a scale from 0 (entirely disagree) to 100 (entirely agree)—with various statements. In the *performance* question, participants are asked to indicate their agreement with “I performed well on the test.” In the *class willingness* question, participants are asked to indicate their agreement with “If given an option, I would choose to take a class that involves topics like those covered on the test.” In the *success* question, participants are asked to indicate their agreement with “I would succeed in a class that involves topics like those covered on the test.”

On the third page, we inform participants of how many questions they got right on the test and then require them to correctly report back that number. By informing participants about their absolute performance, we mechanically close any gap in beliefs about absolute performance once we condition on participants having the same score, which we do in our regression analysis in Section 3. Then, on the fourth page, we elicit participants’ informed self-evaluations by re-asking the same set of questions as they are asked on the first page.

Finally, on the fifth page, we ask participants to complete a short follow-up survey to gather demographic information, including a question about their gender where participants select “male,” “female,” or “other.” If they select other, they may choose to provide free response text about their gender identity. As explained below, we use these responses to classify individuals as gender diverse. Figure B.7 shows how we ask participants to self-report their gender.

3 Results

A total of 10,637 students completed our experiment in Fall 2020. Out of these 10,637 students, 48% selected male ($n=5,106$), 50% selected female ($n=5,329$), and 2% selected other ($n=202$) when asked about their gender. Out of the 202 students who selected other, we exclude 28 students who provided offensive responses in the corresponding free response text box.⁷ We classify the remaining 174 students as *gender diverse* (since they selected other as their gender identity and did not provide an offensive free response answer). Specifically, we classify 71 students as *explicitly gender diverse* since—when asked about their gender—they selected other and then explicitly provided further details in the corresponding class willingnessfree response text box that indicated the nature of their gender identity.⁸ We then classify the remaining 103 students as *implicitly gender diverse* since—when asked

⁷These excluded students are dropped from our analysis, but we confirm the robustness of our findings to using all 202 students (see Section 3.4).

⁸Most of these students mentioned being something other than male or female such as non-binary, transgender, agender, demigirl, demiboy, gender fluid, or pangender; while others provided their gender pronouns (such as she/they, he/they, they/them); and a few of them mentioned that they were still questioning.

about their gender—they selected other and then provided either no response or response that was not specific enough for us to classify them as explicitly gender diverse.⁹

To study our research questions, we compare the behavior of our gender diverse students to those who identify as either female or male.¹⁰ When presenting our findings, the omitted category is always the students who identify as male. This allows us to also show coefficients comparing female students to male students to demonstrate that we replicate well-documented gender gaps in our data. It also allows for a direct comparison between gender diverse students and male students. Finally, our analysis always shows tests comparing gender diverse students to female students, so we can make comparisons across all three groups.

To check the robustness of our findings, our main tables also report analyses using only explicitly gender diverse students, since these students provided informative responses about their gender identity in the corresponding free response text box. It is important to note that there is heterogeneity in the gender identities of individuals who we group together as (explicitly) gender diverse. However, because of small sample size issues, we cannot separately study the confidence and self evaluations of more narrowly defined groups (e.g., we cannot study transgender individuals separately from non-binary individuals), but we note that doing so is an important avenue for future work.

3.1 Performance

As performance outcomes, we study school GPAs, school attendance, and performance on the math and science test that was part of our experiment. Table 1 presents OLS regression analysis results where the dependent variables are indicated in the column titles. Columns 1–6 present GPAs during the Fall quarter (when our data was collected) for all classes (in column 1), core courses (column 2), math courses (column 3), science courses (column 4), social studies courses (column 5), and English and language arts courses (ELA, column 6). Column 7 presents the number of absences during the same quarter. Finally, column 8 shows test performance in the experiment.

As noted above, Table 1 presents results using two measures of gender diversity. Panel A includes indicators for whether a student identifies as female and whether a student identifies as (explicitly or implicitly) gender diverse. Panel B drops data from students who are implicitly gender diverse and includes indicators for whether a student identifies as female and

⁹Specifically, 96 of them left the text box empty, 1 wrote "boy", 1 wrote "kid", 1 wrote "uhhhhh", 1 mentioned that they answered this question already, and 3 mentioned that they prefer not to say.

¹⁰We would like to note that some transgender students (especially those who are far along in their transitioning process) might have chosen their current gender identity rather than "other. Since we do not have data on their sex assigned at birth, which would have made it possible to identify these individuals, our gender diverse population might be undercounted.

Table 1: Performance Outcomes

	GPA in the Following Course(s)							
	All	Core	Math	Science	Soc Stud.	ELA	Absence	Test Perf.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A								
Gender Diverse	-0.83 (0.80)	-0.88 (0.81)	-1.76* (0.95)	-1.38 (1.04)	-1.08 (1.04)	0.04 (1.08)	0.26 (0.34)	-0.05 (0.18)
Female	2.59*** (0.20)	2.68*** (0.21)	2.56*** (0.26)	2.23*** (0.25)	2.65*** (0.26)	3.17*** (0.25)	-0.13* (0.08)	-0.46*** (0.04)
Constant								5.90*** (0.03)
N	10590	10566	10348	9924	9882	10367	10512	10609
$\Delta_{GenderDiverse-Female}$	-3.42	-3.55	-4.31	-3.60	-3.73	-3.13	0.39	0.41
$\Delta_{GenderDiverse-Female}$ p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.25	0.02
Panel B								
Explicitly Gender Diverse	0.65 (1.17)	0.69 (1.18)	-0.10 (1.44)	-0.64 (1.60)	0.81 (1.56)	1.83 (1.68)	0.31 (0.61)	0.81*** (0.23)
Female	2.58*** (0.20)	2.67*** (0.21)	2.55*** (0.26)	2.22*** (0.25)	2.64*** (0.26)	3.15*** (0.25)	-0.13* (0.08)	-0.46*** (0.04)
Constant								5.90*** (0.03)
N	10487	10463	10246	9831	9781	10264	10409	10506
$\Delta_{ExpGendDiv-Female}$	-1.93	-1.98	-2.65	-2.86	-1.83	-1.32	0.45	1.26
$\Delta_{ExpGendDiv-Female}$ p-value	0.10	0.09	0.07	0.07	0.24	0.43	0.46	<0.01
School×Grade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the dependent variable noted in the column. In Columns 1–6, the dependent variable is a student’s GPA in the Fall quarter (when our data was collected) for the following set of courses: all courses, core courses, math courses, science courses, social studies courses, and English & language arts courses, respectively. In Column 7, the dependent variable is a student’s number of absences in the Fall quarter. In Column 8, the dependent variable is a student’s number of questions they answered correctly on the math and science test in our experiment. *Female* is an indicator for the student selecting female when asked about their gender. *Gender Diverse* is an indicator for the student implicitly or explicitly identifying as gender diverse. *Explicitly Gender Diverse* is an indicator for the student explicitly identifying as gender diverse. In each panel, Δ is the difference between the first two coefficient estimates and Δ p-value presents the corresponding p-value for a two-sided t-test of these two coefficient estimates being different. School×Grade FEs are dummies for each school interacted by the student’s year in school. These dummies account for school-grade combinations having different performance standards. The dummies are omitted for the analysis of our math and science test because the test is identical across all students in the study. Both panels exclude the 28 students who selected other and provided an offensive response when asked about their gender. Panel A presents the remaining 10,609 students. Panel B also excludes students who are implicitly gender diverse and presents the remaining 10,506 students. Some columns have smaller sample sizes due to missing values in the administrative data.

whether a student identifies as explicitly gender diverse. Thus, the coefficient estimates on *Gender Diverse* compare gender diverse students to male students, the coefficient estimates on *Female* compare female students to male students, and the Δ at the bottom of each panel compares gender diverse students to female students.

Columns 1–6 of Panel A of Table 1 show that, on average, gender diverse students are similar to male students in terms of their school performance, and Columns 1–6 of Panel B further show that these results are robust when we only focus on those who are explicitly gender diverse.¹¹ On the other hand, gender diverse students’ school performances are (significantly in Panel A and directionally in Panel B) worse compared to their female counterparts. Column 7 shows that gender diverse students have a similar average number of absences compared to both male and female students. It is also the case that average female school performance is significantly better than the average male performance and that female students have a lower number of average absences than male students.

Result 1 (Academic Performance) Gender diverse students perform similarly in school compared to their male counterparts in terms of GPA and attendance. We also have weak evidence suggesting that gender diverse students perform worse than their female counterparts.

We next study performance in the math and science test in the experiment. Results are reported in column 8 of Table 1. Gender diverse students perform similarly to male students and better than female students: gender diverse students got an average of 5.85 questions right, which is statistically indistinguishable from the average male performance of 5.90 and significantly better than the average female performance of 5.44. It is also the case that the average male performance is significantly better than the average female performance. When we only focus on those who are explicitly gender diverse in Panel B, we find that explicitly gender diverse students perform significantly better than both male and female students.

Result 2 (Test Performance) Gender diverse students’ average test performance in the experiment is similar or better than their male counterparts and is better than their female counterparts.

3.2 Beliefs about Performance

In the previous section, we show that—on the math and science test—gender diverse students perform similarly to male students and perform better than female students. In this section, we find that gender diverse students nonetheless have the most pessimistic beliefs about their absolute performance. Specifically, gender diverse students believe they

¹¹The only exception to this is Column 3 in Panel A, which shows that gender diverse students have a marginally significantly lower math GPA compared to male students, however, the difference is no longer significant when we only focus on explicitly gender diverse students in Panel B.

Table 2: Performance Beliefs

	Belief (1)	Belief-Performance (2)
Panel A		
Gender Diverse	-1.37*** (0.21)	-1.37*** (0.23)
Female	-1.03*** (0.04)	-0.80*** (0.05)
Constant		0.75*** (0.03)
N	10609	10609
$\Delta_{GenderDiverse-Female}$	-0.34	-0.57
$\Delta_{GenderDiverse-Female}$ p-value	0.10	0.01
Panel B		
Explicitly Gender Diverse	-1.29*** (0.27)	-1.79*** (0.29)
Female	-1.03*** (0.04)	-0.80*** (0.05)
Constant		0.75*** (0.03)
N	10506	10506
$\Delta_{ExpGendDiv-Female}$	-0.26	-0.99
$\Delta_{ExpGendDiv-Female}$ p-value	0.32	<0.01
Performance FEs	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the dependent variable noted in the column. In Column 1, the dependent variable is a student's believed number of questions that they got right out of the 10 questions on the test. In Column 2, the dependent variable is a student's believed number minus the actual number of questions that they got right out of the 10 questions on the test. *Female* is an indicator for the student selecting female when asked about their gender. *Gender Diverse* is an indicator for the student implicitly or explicitly identifying as gender diverse. *Explicitly Gender Diverse* is an indicator for the student explicitly identifying as gender diverse. In each panel, Δ is the difference between the first two coefficient estimates and Δ p-value presents the corresponding p-value for a two-sided t-test of these two coefficient estimates being different. Performance FEs are dummies for each possible number of questions a student got right out of the 10 questions on the test. Both panels exclude the 28 students who selected other and provided an offensive response when asked about their gender. Panel A presents the remaining 10,609 students. Panel B also excludes students who are implicitly gender diverse and presents the remaining 10,506 students.

got an average of 5.23 questions right, which is significantly lower than both the average male belief of 6.65 questions and the average female belief of 5.39 questions. It is also the

case that the average male belief is significantly higher than the average female belief.

Table 2 further examines these differences in beliefs when controlling for actual performance. Column 1 examines beliefs while including performance fixed effects (i.e., comparing equally performing male, female, and gender diverse participants). Column 2 examines an individual-level variable of absolute performance beliefs minus actual performance. Finally, Table 2 follows the same panel structure as Table 1.

Column 1 of Panel A shows that gender diverse students believe they answered 1.37 significantly fewer questions correct than equally-performing male students and 0.34 (marginally) significantly fewer questions correct than equally-performing female students. In addition, female students believe they answered 1.03 significantly fewer questions right than equally-performing male students, replicating gender gaps in confidence that are typical of male-typed tasks. Column 1 in Panel B shows that these results are very similar when we focus on those who are explicitly gender diverse, although the difference between gender diverse and female students is no longer statistically significant.

Column 2 of Panel A shows that male students are overconfident: they significantly overestimate their performance by 0.75 questions (see the coefficient estimate on the constant). Female students are statistically significantly less confident than male students: they underestimate their performance by 0.05 questions on average ($0.75 - 0.80$ questions). Gender diverse students have even more pessimistic beliefs: they underestimate their performance by 0.62 questions on average ($0.75 - 1.37$ questions). Gender diverse students are statistically significantly less confident than both male students and female students on average. Column 2 of Panel B shows that these results are very similar and all differences remain statistically significant when we focus on participants who are explicitly gender diverse.

Result 3 (Beliefs about Absolute Performance) We identify a “gender minority gap” in confidence. Gender diverse students have significantly more pessimistic beliefs about their absolute performance than both male students and female students.

3.3 Self-Evaluations

Next, we turn to self-evaluations that students provided as part of our experiment. Table 3 presents regression results on students’ uninformed self-evaluations, and Table 4 presents students’ informed self-evaluations. Both tables follow the same panel structure as Table 1.

Let us start by considering students’ uninformed self-evaluations in Table 3. Column 1 of Table 3 presents results for the *performance* question that asked participants to indicate their agreement on a scale from 0 (entirely disagree) to 100 (entirely agree) with having “performed well on the test.” Panel A shows that gender diverse students provide a self-evaluation that is 16.75 points lower than those provided by male students and 5.84 points lower than

Table 3: Uninformed Self-Evaluations

	Performance	Performance- Bucket	Class Willingness	Success
	(1)	(2)	(3)	(4)
Panel A				
Gender Diverse	-16.75*** (2.17)	-0.71*** (0.11)	-9.09*** (2.48)	-15.19*** (2.40)
Female	-10.91*** (0.45)	-0.51*** (0.02)	-4.53*** (0.58)	-7.50*** (0.54)
N	10609	10609	10609	10609
$\Delta_{GenderDiverse-Female}$	-5.84	-0.19	-4.56	-7.68
$\Delta_{GenderDiverse-Female}$ p-value	0.01	0.08	0.07	<0.01
Panel B				
Explicitly Gender Diverse	-19.17*** (2.94)	-0.73*** (0.14)	-8.32** (3.39)	-17.56*** (3.35)
Female	-10.91*** (0.45)	-0.51*** (0.02)	-4.54*** (0.58)	-7.51*** (0.54)
N	10506	10506	10506	10506
$\Delta_{ExpGendDiv-Female}$	-8.26	-0.22	-3.78	-10.05
$\Delta_{ExpGendDiv-Female}$ p-value	0.01	0.13	0.27	<0.01
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of a student's response to the uninformed self-evaluation (elicited before the student learns their test performance) noted in the column. See Table 2 for definitions of the independent variables, Δ and Performance FEs. All panels exclude the 28 who selected other and provided an offensive response when asked about their gender. Panel A presents the remaining 10,609 students, while Panel B excludes students who identify as implicitly gender diverse and present the remaining 10,506 students.

those provided by female students. These differences are both statistically significant and substantial, corresponding to self-evaluations that are 25.8% and 7.4% lower than the average male and female self-evaluations, respectively.¹² In addition, female students provide self-evaluations that are 10.91 points (19.9%) significantly lower than those provided by male students. Panel B further shows that these results are very similar when we only focus on those who are explicitly gender diverse.

Column 2 of Table 3 presents results for the *performance-bucket* question that asked participants to indicate how well they think they performed on the test on a seven-point Likert scale. Panel A shows that the average response by gender diverse students is 0.71

¹²Percentage changes reported in this section are calculated using the corresponding groups' average responses.

points (15.5%) significantly lower than the average response of male students and 0.19 points (2.5%) significantly lower than the average response of female students. In addition, female students' average responses are 0.51 points (13.4%) significantly lower than those provided by male students. Panel B further shows that these results are very similar when we focus on those who are explicitly gender diverse, although the difference between gender diverse and female students loses statistical significance.

Column 3 of Table 3 presents results for the *class willingness* question that asked participants to indicate their agreement on a scale from 0 (entirely disagree) to 100 (entirely agree) with "I would choose to take a class that involves topics like those covered on the test." Panel A shows that the average response by gender diverse students is 9.09 points (16.7%) significantly lower than those provided by male students and 4.56 points (7.7%) significantly lower than those provided by female students. In addition, female students' average responses are 4.53 points (9.8%) significantly lower than those provided by male students. Panel B further shows that these results are very similar when we focus on those who are explicitly gender diverse, although the difference between gender diverse and female students loses statistical significance.

Column 4 of Table 3 presents results for the *success* question that asked participants to indicate their agreement on a scale from 0 (entirely disagree) to 100 (entirely agree) with "I would succeed in a class that involves topics like those covered on the test." Panel A shows that the average response by gender diverse students is 15.19 points (22.6%) and 7.68 points (10.1%) significantly lower than those provided by male and female students, respectively. In addition, female students' average responses are 7.50 points (13.9%) significantly lower than those provided by male students. Panel B further shows that these results are very similar when we focus on those who are explicitly gender diverse.

One potential explanation for the gender minority gap in self-evaluations could come from the results in Section 3.2, which show that gender diverse students have significantly more pessimistic beliefs about their performance compared to both male and female students. To test for the role of absolute beliefs about performance, we inform participants about their performance on the test (i.e., we tell them exactly how many questions they answered correctly on the test) and then ask them to report their performance back to us to confirm they saw their score. We then ask them the same four self-evaluation questions to elicit informed self-evaluations.

The results about informed self-evaluations are presented in Table 4. Similar to the uninformed self-evaluations, we again see that gender diverse students' informed self-evaluations are significantly and substantially lower than their male and female counterparts. Looking at Columns 1, 3, and 4 in Panel A of Table 4 where subjects were asked to indicate their degree

Table 4: Informed Self-Evaluations

	Performance	Performance- Bucket	Class Willingness	Success
	(1)	(2)	(3)	(4)
Panel A				
Gender Diverse	-13.40*** (2.31)	-0.53*** (0.13)	-11.30*** (2.54)	-16.61*** (2.48)
Female	-6.35*** (0.53)	-0.25*** (0.03)	-3.18*** (0.60)	-5.40*** (0.59)
N	10609	10609	10609	10609
$\Delta_{GenderDiverse-Female}$	-7.06	-0.28	-8.12	-11.21
$\Delta_{GenderDiverse-Female}$ p-value	<0.01	0.03	<0.01	<0.01
Panel B				
Explicitly Gender Diverse	-17.90*** (3.08)	-0.89*** (0.17)	-10.70*** (3.81)	-18.31*** (3.71)
Female	-6.33*** (0.53)	-0.25*** (0.03)	-3.18*** (0.60)	-5.39*** (0.59)
N	10506	10506	10506	10506
$\Delta_{ExpGendDiv-Female}$	-11.57	-0.64	-7.53	-12.92
$\Delta_{ExpGendDiv-Female}$ p-value	<0.01	<0.01	0.05	<0.01
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of a student's response to the informed self-evaluation (elicited after the student learns their test performance) noted in the column. See Table 2 for definitions of the independent variables, Δ and Performance FEs. All panels exclude the 28 who selected other and provided an offensive response when asked about their gender. Panel A presents the remaining 10,609 students, while Panel B excludes students who identify as implicitly gender diverse and present the remaining 10,506 students.

of agreement with each of the corresponding statements, we see that the average responses submitted by gender diverse students are 13.40 points (30.0%) and 7.06 points (12.6%) significantly lower for the *performance* question, 11.30 points (22.5%) and 8.12 points (15.1%) significantly lower for the *class willingness* question, and 16.61 points (29.5%) and 11.21 points (18.9%) significantly lower for the *success* question relative to male and female students, respectively. Similarly for the seven-point Likert-scale question in Column 2, gender diverse students provide 0.53 points (15.6%) and 0.28 points (5.0%) significantly lower self-evaluations compared to their male and female counterparts, respectively.¹³ Columns 1–4

¹³In addition, female students' average responses are also 6.35 points (19.9%) significantly lower for the *performance* question, 0.25 points (11.1%) significantly lower for the *performance-bucket* question, 3.18 points (8.6%) significantly lower for the *class willingness* question, and 5.40 points (13.1%) significantly lower for the *success* question relative to male students.

of Panel B of Table 4 further show that these results are very similar when we focus on individuals who are explicitly gender diverse.

We summarize these findings below:

Result 4 We identify a “gender minority gap” in self-evaluations. Gender diverse students provide significantly worse self-evaluations relative to their male and female counterparts. These differences in self-evaluations cannot be explained by differences in beliefs about absolute performance.

3.4 Robustness Checks

In Sections 3.1–3.3, we show results for gender diverse students in each Panel A and for explicitly gender diverse students in each Panel B. In this section, we provide further robustness checks and show that the findings presented above are robust to different ways of identifying gender minority status.

As a first set of robustness checks, we report findings including all 202 students who selected “other” as their gender in our survey (i.e., we do not exclude the 28 students who provided an offensive response when asked about their gender, and we instead classify these students as gender diverse). Appendix Table A.1 reproduces Table 1, Appendix Table A.2 reproduces Table 2, Appendix Table A.3 reproduces Table 3, and Appendix Table A.4 reproduces Table 4 when these additional 28 students are classified as gender diverse and included in the analyses. All of our results prove robust.

As a second set of robustness checks, we instead rely on gender data obtained by the Character Lab Research Network’s demographic survey, which was conducted before our experiment.¹⁴ Appendix Table A.5 reproduces Table 1, Appendix Table A.6 reproduces Table 2, Appendix Table A.7 reproduces Table 4, and Appendix Table A.8 reproduces Table 4 with this alternative definition of gender diverse. All of our results prove robust.

4 Conclusion

A robust literature in economics explores how gender correlates with confidence and self-evaluations. However, this literature mainly focuses on binary gender and often conflates gender identity and “sex.” Given that an increasing share of the population identifies as something other than male or female (Brown, 2022), it is important to study the economic behavior and preferences of gender minorities separately from their male and female counterparts. However, it is challenging to identify and recruit gender diverse individuals, mostly due to their minority status. We are able to overcome this challenge through a partnership

¹⁴As part of their demographics survey, Character Lab asked students to indicate their gender identity. They were given the option to choose “Male,” “Female,” “Other,” or “Prefer not to say.” We exclude those who select “Prefer not to say” and classify those who selected “Other” as gender diverse.

with the Character Lab Research Network, which made it possible for us to recruit a large and young student population.

We document a gender minority gap in confidence. We find that gender diverse middle and high school students have lower confidence in their absolute performance on a math and science test. They also provide more pessimistic self-evaluations about their performance compared to their male and female counterparts. This gender minority gap in self-evaluations stays large and significant even after we inform students about their actual performance on the test. These findings are especially interesting given that gender diverse students perform equally to male students and better than female students on the test. Finally, we show that our results are robust to different ways in which we can identify and classify gender diverse students.

We see the persistence of the self-evaluation gap as particularly worrisome. At various points during a student’s academic experience and subsequent labor market career, they are likely to be asked to provide self-evaluations of their abilities and performance. In college applications, in applications for various scholarships or jobs, and in performance reviews, there are many invitations for individuals to discuss their own strengths and weaknesses. That gender minorities provide worse self-evaluations about their performance and ability could put them at a disadvantage in these environments. Thus, the results in our paper open up many important questions for future work.

First, future work may investigate whether the gender minority gap in confidence influences educational and labor market outcomes. Students who have lower self-confidence in their ability in general may be less likely to pursue college education. Moreover, students who have lower self-confidence in their ability to perform well in *math and science* may gravitate away from STEM fields that require strong math and science skills, which could cause occupational segregation and contribute to the socioeconomic disparities that gender minorities experience. Or, even conditional on securing the same jobs, students who have lower self-confidence may be less likely to negotiate for higher salaries.

Second, future work may investigate *why* the gender minority gaps in confidence and self-evaluations exist in the first place. For instance, in light of previous research on how social norms impact economic behavior, one natural question is whether norms are different for gender minorities in a way that influences their beliefs.

Third, future work may strive to better understand the diversity within gender minorities. While our study was not sufficiently powered to investigate more narrowly defined groups, it is of clear importance to separately study groups of gender minorities such as those who identify as transgender men, transgender women, non-binary, genderqueer,^{10,637} students or gender non-conforming.

References

- Aksoy, Billur, Christopher S. Carpenter, and Dario Sansone.** 2022. “Understanding Labor Market Discrimination Against Transgender People: Evidence from a Double List Experiment and a Survey.” *National Bureau of Economic Research Working Paper Series*, , (30483).
- Badgett, MV, Christopher S Carpenter, and Dario Sansone.** 2021. “LGBTQ economics.” *Journal of Economic Perspectives*, 35(2): 141–70.
- Bicchieri, Cristina.** 2005. *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.
- Blau, Francine D., and Lawrence M. Kahn.** 2017. “The Gender Wage Gap: Extent, Trends, and Explanations.” *Journal of Economic Literature*, 55(3).
- Bordalo, Pedro, Katherine Coffman, Nicola Gennaioli, and Andrei Shleifer.** 2019. “Beliefs about Gender.” *American Economic Review*.
- Brown, Anna.** 2022. “About 5% of young adults in the U.S. say their gender is different from their sex assigned at birth.” *Pew Research Center*.
- Carpenter, Christopher S, Maxine J Lee, and Laura Nettuno.** 2022. “Economic Outcomes for Transgender People and Other Gender Minorities in the United States: First Estimates from a Nationally Representative Sample.” *Available at SSRN*.
- Carpenter, Christopher S, Samuel T Eppink, and Gilbert Gonzales.** 2020. “Transgender status, gender identity, and socioeconomic outcomes in the United States.” *ILR Review*, 73(3): 573–599.
- Coffman, Katherine Baldiga.** 2014. “Evidence on Self-Stereotyping and the Contribution of Ideas.” *The Quarterly Journal of Economics*, 129(4): 1625–1660.
- Coffman, Katherine, Manuela Collis, and Leena Kulkarni.** 2019. “Stereotypes and Belief Updating.” *Working Paper*.
- Downing, Janelle M., and Julia M. Przedworski.** 2018. “Health of transgender adults in the US, 2014–2016.” *American Journal of Preventive Medicine*, 55(3): 336–344.
- Exley, Christine L., and Judd B. Kessler.** 2022. “The Gender Gap in Self-Promotion.” *Quarterly Journal of Economics*.
- Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore.** 2021. “On her own account: How strengthening women’s financial control impacts labor supply and gender norms.” *American Economic Review*, 111(7): 2342–75.

- Hernandez-Arenaz, Iñigo, and Nagore Iriberrí.** 2019. “A review of gender differences in negotiation.” *Oxford Research Encyclopedia of Economics and Finance*.
- Jones, Jeffrey M.** 2022. “LGBT Identification in U.S. Ticks Up to 7.1%.”
- Lewis, Daniel C, Andrew R Flores, Donald P Haider-Markel, Patrick R Miller, and Jami K Taylor.** 2022. “Transitioning opinion? assessing the dynamics of public attitudes toward transgender rights.” *Public Opinion Quarterly*, 86(2): 343–368.
- Lewis, Daniel C, Andrew R Flores, Donald P Haider-Markel, Patrick R Miller, Barry L Tadlock, and Jami K Taylor.** 2017. “Degrees of acceptance: Variation in public attitudes toward segments of the LGBT community.” *Political Research Quarterly*, 70(4): 861–875.
- Lundeberg, Mary A, Paul W Fox, and Judith Punčochař.** 1994. “Highly confident but wrong: Gender differences and similarities in confidence judgments.” *Journal of educational psychology*, 86(1).
- Meyer, Ilan H, Taylor NT Brown, Jody L Herman, Sari L Reisner, and Walter O. Bockting.** 2017. “Demographic characteristics and health status of transgender adults in select US regions: Behavioral Risk Factor Surveillance System, 2014.” *American journal of public health*, 107(4): 582–589.
- Niederle, Muriel.** 2016. “Gender.” In *Handbook of Experimental Economics*. Vol. 2, , ed. John Kagel and Alvin E. Roth, 481–553. Princeton University Press.
- Niederle, Muriel, and Lise Vesterlund.** 2007. “Do Women shy away from competition? Do men compete too much?” *Quarterly Journal of Economics*, 122(3): 1067–1101.
- Niederle, Muriel, and Lise Vesterlund.** 2011. “Gender and Competition.” *Annual Review of Economics*, 3: 601–630.
- Pande, Rohini, and Helena Roy.** 2021. ““If you compete with us, we shan’t marry you” The (Mary Paley and) Alfred Marshall Lecture.” *Journal of the European Economic Association*, 19(6): 2992–3024.
- Sansone, Dario.** 2019. “LGBT students: New evidence on demographics and educational outcomes.” *Economics of Education Review*, 73(101933).

Appendices (For Online Publication Only)

A Additional Tables and Results Mentioned in Main Text

Table A.1: Performance Outcomes Including Students with Offensive Gender Responses

	GPA in the Following Course(s)							Test Perf
	All	Core	Math	Science	Soc Stud.	ELA	Absence	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender Diverse	-0.73 (0.76)	-0.73 (0.76)	-1.37 (0.90)	-0.95 (0.97)	-1.19 (1.00)	-0.31 (1.01)	0.15 (0.30)	-0.09 (0.17)
Female	2.59*** (0.20)	2.67*** (0.21)	2.56*** (0.26)	2.23*** (0.25)	2.65*** (0.26)	3.16*** (0.25)	-0.13* (0.08)	-0.46*** (0.04)
N	10618	10594	10374	9950	9905	10394	10540	10637
$\Delta_{Gender\ Diverse-Female}$	-3.32	-3.40	-3.93	-3.18	-3.83	-3.47	0.28	0.37
$\Delta_{Gender\ Diverse-Female}$ p-value	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.34	0.03
School*Grade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the dependent variable noted in the column. See Table 1 for definitions of the dependent variable, the independent variables, Δ and School*Grade FEs. The data are from all 10,637 students (including 28 participants who—when asked about their gender—selected other and provided an offensive response).

Table A.2: Performance Beliefs Including Students with Offensive Gender Responses

	Belief (1)	Belief-Performance (2)
Gender Diverse	-1.31*** (0.20)	-1.28*** (0.23)
Female	-1.03*** (0.04)	-0.80*** (0.05)
Constant		0.75*** (0.03)
N	10637	10637
$\Delta_{GenderDiverse-Female}$	-0.28	-0.48
$\Delta_{GenderDiverse-Female}$ p-value	0.17	0.04
Performance FEs	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the dependent variable noted in the column. See Table 2 for definitions of the dependent variables, the independent variables, Δ and Performance FEs. The data are from all 10,637 students (including 28 participants who—when asked about their gender—selected other and provided an offensive response).

Table A.3: Uninformed Self-Evaluations Including Students with Offensive Gender Responses

	Performance (1)	Performance-Bucket (2)	Class Willingness (3)	Success (4)
Gender Diverse	-15.78*** (2.14)	-0.71*** (0.11)	-10.18*** (2.40)	-15.04*** (2.33)
Female	-10.91*** (0.45)	-0.51*** (0.02)	-4.53*** (0.58)	-7.50*** (0.54)
N	10637	10637	10637	10637
$\Delta_{GenderDiverse-Female}$	-4.87	-0.20	-5.65	-7.54
$\Delta_{GenderDiverse-Female}$ p-value	0.02	0.08	0.02	<0.01
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of a student's response to the uninformed self-evaluation (elicited before the student learns their test performance) noted in the column. See Table 3 for definitions of the independent variables, Δ and Performance FEs. Data are from all 10,637 students (including 28 participants who—when asked about their gender—selected other and provided an offensive response).

Table A.4: Informed Self-Evaluations Including Students with Offensive Gender Responses

	Performance	Performance-Bucket	Class Willingness	Success
	(1)	(2)	(3)	(4)
Gender Diverse	-11.10*** (2.24)	-0.42*** (0.12)	-11.55*** (2.43)	-15.03*** (2.39)
Female	-6.35*** (0.53)	-0.25*** (0.03)	-3.18*** (0.60)	-5.39*** (0.59)
N	10637	10637	10637	10637
$\Delta_{GenderDiverse-Female}$	-4.76	-0.17	-8.37	-9.63
$\Delta_{GenderDiverse-Female}$ p-value	0.03	0.17	<0.01	<0.01
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of a student's response to the informed self-evaluation (elicited after the student learns their test performance) noted in the column. See Table 4 for definitions of the independent variables, Δ and Performance FEs. Data are from all 10,637 students (including 28 participants who—when asked about their gender—selected other and provided an offensive response).

Table A.5: Performance Outcomes Using CLRN Gender Data

	GPA in the Following Course(s)							
	All	Core	Math	Science	Soc Stud.	ELA	Absence	Test Perf
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Gender diverse	-1.07 (0.88)	-0.95 (0.88)	-1.45 (1.04)	-0.87 (1.15)	-1.82 (1.14)	-0.35 (1.21)	0.36 (0.37)	0.24 (0.18)
Female	2.64*** (0.20)	2.70*** (0.21)	2.56*** (0.26)	2.27*** (0.26)	2.69*** (0.27)	3.19*** (0.25)	-0.20** (0.08)	-0.47*** (0.05)
N	10087	10063	9857	9447	9400	9872	10017	10106
$\Delta_{GenderDiverse-Female}$	-3.71	-3.65	-4.01	-3.14	-4.51	-3.53	0.56	0.71
$\Delta_{GenderDiverse-Female}$ p-value	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	0.12	<0.01
School*Grade FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the dependent variable noted in the column. See Table 1 for definitions of the dependent variable, the independent variables, Δ and School*Grade FEs. These results rely on the gender data from the Character Lab Research Network survey. *Female* is an indicator for student selecting Female, Femenino, or Mujer when asked about their gender in this survey. *Gender Diverse* is an indicator for the students selecting Other when asked about their gender in this survey. We exclude the 531 students who did not answer the gender question in this survey. The data are from the remaining 10,106 students.

Table A.6: Performance Beliefs Using CLRN Gender Data

	Belief (1)	Belief-Performance (2)
Gender diverse (CL data)	-1.46*** (0.22)	-1.63*** (0.24)
Female (CL data)	-1.01*** (0.04)	-0.78*** (0.05)
Constant		0.72*** (0.03)
N	10106	10106
$\Delta_{GenderDiverse-Female}$	-0.45	-0.85
$\Delta_{GenderDiverse-Female}$ p-value	0.04	<0.01
Performance FEs	Yes	No

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of the dependent variable noted in the column. See Table 2 for definitions of the dependent variable, Δ and Performance FEs. These results rely on the gender data from the Character Lab Research Network survey. *Female* is an indicator for student selecting Female, Femenino, or Mujer when asked about their gender in this survey. *Gender Diverse* is an indicator for the students selecting Other when asked about their gender in this survey. We exclude the 531 students who did not answer the gender question in this survey. The data are from the remaining 10,106 students.

Table A.7: Uninformed Self-Evaluations Using CLRN Gender Data

	Performance (1)	Performance-Bucket (2)	Class Willingness (3)	Success (4)
Gender diverse	-18.11*** (2.37)	-0.82*** (0.12)	-9.42*** (2.61)	-15.78*** (2.52)
Female	-10.76*** (0.47)	-0.50*** (0.02)	-4.22*** (0.60)	-7.36*** (0.55)
N	10106	10106	10106	10106
$\Delta_{GenderDiverse-Female}$	-7.35	-0.32	-5.20	-8.43
$\Delta_{GenderDiverse-Female}$ p-value	<0.01	0.01	0.05	<0.01
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of a student's response to the uninformed self-evaluation (elicited before the student learns their test performance) noted in the column. See Table 3 for definitions of the independent variables, Δ and Performance FEs. These results rely on the gender data from the Character Lab Research Network survey. *Female* is an indicator for student selecting Female, Femenino, or Mujer when asked about their gender in this survey. *Gender Diverse* is an indicator for the students selecting Other when asked about their gender in this survey. We exclude the 531 students who did not answer the gender question in this survey. The data are from the remaining 10,106 students.

Table A.8: Informed Self-Evaluations Using CLRN Gender Data

	Performance	Performance- Bucket	Class Willingness	Success
	(1)	(2)	(3)	(4)
Gender diverse	-14.92*** (2.53)	-0.65*** (0.13)	-10.83*** (2.71)	-16.01*** (2.77)
Female	-6.33*** (0.54)	-0.25*** (0.03)	-2.94*** (0.62)	-5.22*** (0.60)
N	10106	10106	10106	10106
$\Delta_{GenderDiverse-Female}$	-8.59	-0.40	-7.89	-10.79
$\Delta_{GenderDiverse-Female}$ p-value	<0.01	<0.01	<0.01	<0.01
Performance FEs	Yes	Yes	Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. SEs are robust. Results are from OLS regressions of a student's response to the informed self-evaluation (elicited after the student learns their test performance) noted in the column. See Table 4 for definitions of the independent variables, Δ and Performance FEs. These results rely on the gender data from the Character Lab Research Network survey. *Female* is an indicator for student selecting Female, Femenino, or Mujer when asked about their gender in this survey. *Gender Diverse* is an indicator for the students selecting Other when asked about their gender in this survey. We exclude the 531 students who did not answer the gender question in this survey. The data are from the remaining 10,106 students.

B Experimental Instructions

Prior to participating in the study, participants must correctly answer a captcha and consent to participate. At the end of the study, participants must complete a short follow-up survey to gather demographic information. Participants are recruited via the Character Lab Research Network and complete this study as part of the curriculum at school. There are no payments associated with this study.

The study begins by informing each participant about the test that they will take. The instructions for the test are displayed in Figure B.1 and an example of a question on the test is displayed in Figure B.2 (note that the timer in that screenshot indicates the participant has 24 seconds left to answer the question although the timer starts at 30 seconds). After completing the test, participants are asked to complete five additional pages of the study.

On the first page, they are asked about their absolute performance belief (see Figure B.3). On the second page, they are asked the self-evaluation questions (see Figure B.4). On the third page, participants are provided with perfect information on their absolute performance and are required to correctly report back their absolute performance (see Figure B.5). On the fourth page, they are asked the self-evaluation questions again (see Figure B.6). On the fifth page, they are asked for demographic information including their gender identity (see Figure B.7).

Figure B.1: Instructions for the test

Information about the Test:

On the test, you will be asked to answer up to 10 questions from the Armed Services Vocational Aptitude Battery (ASVAB). Each question will test your aptitude in one of the following five categories: General Science, Arithmetic Reasoning, Math Knowledge, Mechanical Comprehension, and Assembling Objects. In addition to being used by the military to determine which jobs armed service members are qualified for, performance on the ASVAB is often used as a measure of cognitive ability by academic researchers.

You will be presented with each of the 10 questions on separate pages. You will be given up to 30 seconds to answer each question, although you may push the arrow at the bottom of the page to answer a question before the 30 seconds are up.

Please try to answer each question as best as you can.

Figure B.2: Example question on the test

English ▼

24

Question 2 out of 10:

MATH KNOWLEDGE: Which number has the greatest value?

9,299

903 tens

93 hundreds

9 thousands

Figure B.3: Absolute Performance Belief Question

Page 1 out of 5

Please answer the following question.

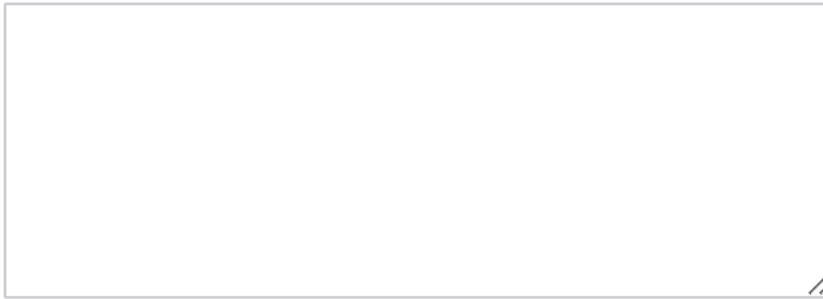
Out of the 10 questions on the test, how many questions do you think you answered correctly?

Figure B.4: Self-Evaluation Questions

Page 2 out of 5

Please answer the following questions.

Please describe how well you think you performed on the test and why.



Please indicate how well you think you performed on the test.

Terrible	Very Poor	Poor	Neutral	Good	Very Good	Exceptional
----------	-----------	------	---------	------	-----------	-------------

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement:

Entirely	Strongly		Somewhat	Neither	Somewhat		Strongly	Entirely		
Disagree	Disagree	Disagree	Disagree	Disagree	Agree	Agree	Agree	Agree		
0	10	20	30	40	50	60	70	80	90	100

I performed well on the test.



If given an option, I would choose to take a class that involves topics like those covered on the test.



I would succeed in a class that involves topics like those covered on the test.



Figure B.5: Absolute Performance Information

Page 3 out of 5

On the test, you answered **0 questions correctly out of the 20 questions.** To confirm that you read the prior sentence, please answer the following question.

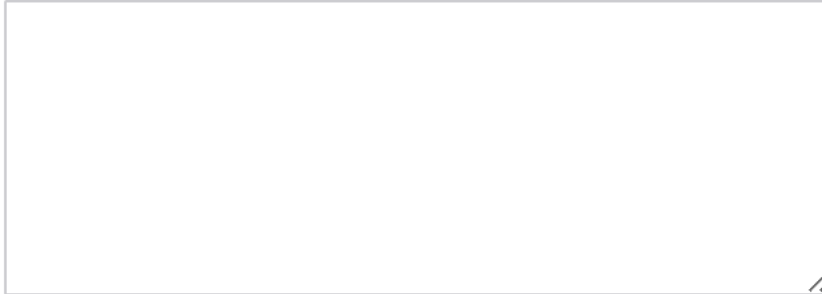
Of the 10 questions on the test you took in part 1, how many questions did you answer correctly?

Figure B.6: Informed Self-Evaluation Questions

Page 4 out of 5

Now that you have information on your test performance, please answer the following questions again. Your answers may be the same or different than your previous answers.

Please describe how well you think you performed on the test and why.



Please indicate how well you think you performed on the test.

Terrible	Very Poor	Poor	Neutral	Good	Very Good	Exceptional
----------	-----------	------	---------	------	-----------	-------------

On a scale from 0 (entirely disagree) to 100 (entirely agree), please indicate the extent to which you agree with the following statement:

Entirely Disagree	Strongly Disagree	Disagree	Somewhat Disagree	Neither Disagree Nor Agree	Somewhat Agree	Agree	Strongly Agree	Entirely Agree		
0	10	20	30	40	50	60	70	80	90	100

I performed well on the test.



If given an option, I would choose to take a class that involves topics like those covered on the test.



I would succeed in a class that involves topics like those covered on the test.



Figure B.7: Screenshot of gender question

Please select your gender.

Male

Female

Other