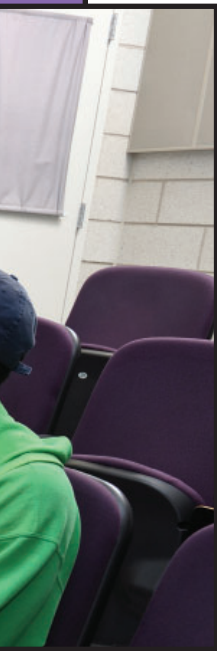


MAKING GRADUATE ADMISSIONS IN PHYSICS MORE EQUITABLE

Physics graduate students at Michigan State University attending class (top), working on a problem set (middle), and solving an in-class problem (bottom). (Courtesy of Harley Seeley.)





Nicholas Young is a postdoc at the Center for Academic Innovation at the University of Michigan in Ann Arbor. **Kirsten Tollefson** is an associate dean in the graduate school and a professor in the department of physics and astronomy, and **Danny Caballero** is an associate professor in the department of physics and astronomy and the department of computational mathematics, science, and engineering, both at Michigan State University in East Lansing.



Nicholas T. Young, Kirsten Tollefson, and Marcos D. Caballero

Preliminary results from the revamping of Michigan State University's physics graduate admissions process suggest that the changes have made the procedure fairer for all.

Over the years, national and international physics organizations, research laboratories, and physics departments have called on leaders in the field to address issues of diversity, equity, and inclusion. In response, the physics community has built programs that increase opportunity and support for a more diverse scientific workforce and aim to address long-standing disparities in who earns degrees in physics; physics educators have developed and implemented pedagogy and curricula that provide more equitable learning opportunities for our students;¹ and physics organizations have developed and implemented codes of conduct across our organizations aimed at making meetings and conferences more inclusive.

But the issues with ensuring a diverse, equitable, and inclusive environment are systemic and pervasive. They are perpetuated often not by actions but by inactions or well-intentioned, yet misplaced, actions.² To make physics more diverse, equitable, and inclusive, we must address systemic issues directly, collaboratively, and reflectively. One area that we have chosen to focus on is graduate admissions. It is especially important to consider diversity there because of its potential ripple effect across science and engineering. Individuals who complete graduate physics degrees are well positioned to become scientific leaders in industry and government, and physicists who pursue academic careers will train the next generation of scientists, engineers, teachers, and even medical doctors.

The admission of graduate students to post-bachelor's physics programs is a complex and challenging system.

Any graduate director, faculty member, or graduate student can recount their own vivid experience with that complicated and, quite often, opaque process.³ At Michigan State University, we set out to understand the admissions process, determine how it was functioning, and make changes to it. We hoped to admit more diverse candidates to our program, evaluate them more equitably and holistically, and, ultimately, create a more inclusive program where each student is valued and supported in their studies. We are far from our ideal collective vision but feel that our progress toward that vision is nonetheless important to share with our colleagues.

The traditional admissions process in physics

Unlike undergraduate admissions, which is typically carried out by a centralized admissions office, individual departments or programs usually handle the review and

GRADUATE ADMISSIONS

evaluation of graduate applications. An applicant to a US physics graduate program will typically submit their CV, undergraduate transcripts, letters of recommendation, and multiple written statements covering such topics as their personal history, research experience, motivations and goals for attending graduate school, and, occasionally, how their experiences and actions contribute to fostering a diverse community.

Depending on the program, an applicant may also be required to submit their scores for the general Graduate Record Examination (GRE) and the GRE subject test in physics (GREP). Currently only around 25% of physics and astronomy graduate programs in the US and Canada require or recommend the submission of GRE scores. Most of the other programs treat score submission as optional.⁴ A group of physics faculty members reviews the applications and extends offers to the selected applicants. Some departments also have postdoctoral researchers and current graduate students assist in reviewing applications, although our department does not.

From various studies conducted on the graduate admissions process in physics, we know that quantitative parts of the application, such as grade point average (GPA) and GREP scores, usually drive the admissions decision. The strength of an applicant's letters of recommendation and the specific physics courses they took as an undergraduate are also important.⁵ Although that approach to admissions is somewhat successful—today the number of physics PhDs awarded annually is near the all-time high—it is inequitable. It favors applicants from groups who are already advantaged in physics and hurts applicants who are underrepresented in the field (see figure 1).

That pattern is most apparent in GREP scores, where Asian and white men tend to score higher than everyone else.⁶ In the past, some physics departments required minimum GREP scores for admission,⁵ which meant that applicants who were not Asian or white men were at a disadvantage.

The GREP also introduces inequities based on an applicant's financial resources. Taking the test and sending scores to each institution one applies to can cost hundreds of dollars. Moreover, applicants at smaller schools might need to travel to a testing location and potentially stay there overnight. For students working jobs at the federal minimum wage of \$7.25 per hour, the cost of taking the exam can easily exceed 40 hours of take-home pay.

Further, some of our previous work found that applicants from larger departments or more selective schools tended to score higher on the GREP than applicants from smaller departments or less selective schools.⁷ Because students at those larger departments or more selective schools tend to be less diverse and more affluent than the college population at large, using the GREP in admissions can further filter out many of the students whom departments are attempting to attract through diversity, equity, and inclusion initiatives.

Other sources of inequity

If the GREP were the only inequitable part of the application process, it would be easy to make admissions more equitable by removing it. Indeed, for all the above reasons, admissions committees had already begun de-emphasizing GREP scores even before pandemic disruptions, such as testing-site closures and the lack of a virtual GREP exam, which meant that an entire cohort of students who never took the test was admitted.

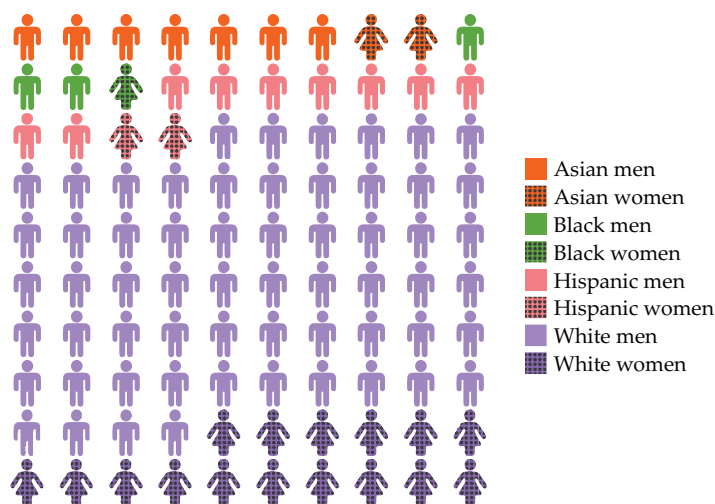


FIGURE 1. A VISUAL REPRESENTATION of the potential applicant pool for physics graduate programs. Each glyph corresponds to 1% of US graduates of various races, genders, and ethnicities who received physics bachelor's degrees from 2016 to 2020. Students who identify as American Indian, Alaska Native, Native Hawaiian, or Other Pacific Islander represent less than 1% of physics graduates and are not shown in the plot. (Data courtesy of the American Physical Society and the Integrated Postsecondary Education Data System.)

Unfortunately, that hasn't changed the admissions pattern. Even when GREP scores are removed, inequity permeates other parts of the application process.

Grades, for example, are a major determining factor in whether an applicant will be admitted, and they, too, have been found to show gender and racial differences. For example, a 2021 study conducted by researchers at the University of Pittsburgh found that students who belong to underrepresented minority groups earned lower grades than even the most disadvantaged students from any other group.⁸

In physics, various investigations have found that women earn lower grades than men. For example, one multi-institutional study found that even after accounting for prior performance, women earned lower grades than men in introductory physics.⁹ That result suggests that differences in grades in those courses are more reflective of grading policies than student ability. It thus follows that selecting applicants based on GPA can hurt the admissions chances of applicants currently underrepresented in physics.

The inequity can also appear implicitly through the bias of those reviewing the applications. For example, in one 2012 study, researchers rated a male applicant for a lab manager position as more competent and hireable than a female applicant with, apart from the name, an identical application. A 2020 follow-up repeated that study with both race and gender. It again found that men were viewed as more competent and hireable than women and also revealed that white and Asian applicants were seen as more competent and hireable than Black and Hispanic candidates.¹⁰ And since applicants are required to submit statements and letters of recommendation from faculty, the nonquantitative portions of the application are also susceptible to contributing to inequitable outcomes.

Because inequities exist throughout the entire admissions

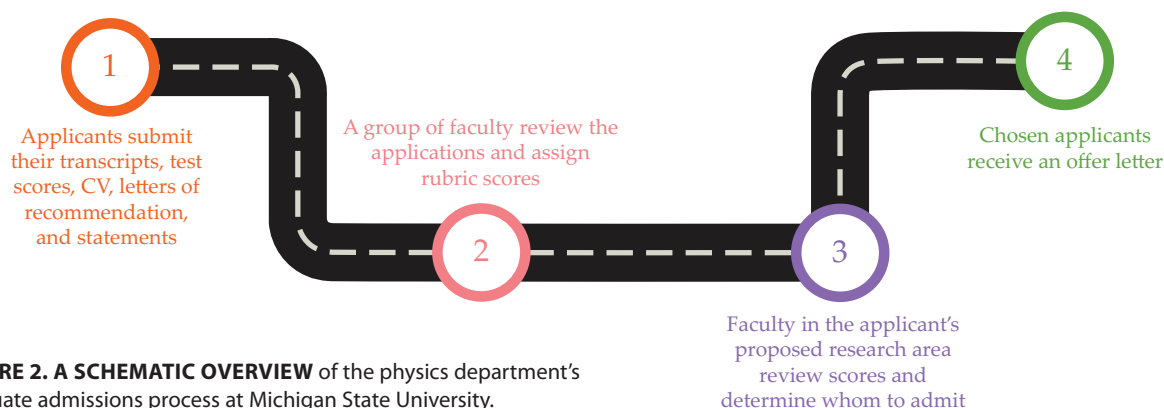


FIGURE 2. A SCHEMATIC OVERVIEW of the physics department's graduate admissions process at Michigan State University.

process, we can't simply change one part of the application process to make it more equitable. Instead, we need to consider a different approach to evaluating applicants and make the process fairer while also acknowledging that students live and learn in inequitable environments. That approach needs to consider the student as a whole and consider broadly what skills and traits an applicant needs to be successful in graduate school.

How we've changed our admissions process

In recognition of those issues, our department began to rethink its graduate admissions process in 2016. Although we expected our students to have strong math and physics skills, we also anticipated that they would be able to learn independently, take initiative, and be resilient in the face of difficulties and unexpected challenges. But our admissions process didn't have a standardized way to assess applicants on the last traits. Normally it didn't even take them into account, and if they were considered at all, they were implicitly determined from the applicant's personal and research statements.

Around the same time, physics and astronomy departments were beginning to think about how to increase diversity in their programs. Many of them started to consider the idea of assessing applicants' noncognitive traits.¹¹ Given the subjective nature of assessing such traits, one recommendation we received about making the process more fair to all applicants was to use a predefined rubric.¹² By using one that defines all the evaluation criteria ahead of time, applicants are compared on the same basis, and evaluators have less to debate about whether an applicant demonstrates the expected trait. Determining the evaluation criteria ahead of time can also reduce subjectivity in evaluations.¹³

The Inclusive Graduate Education Network, an NSF-funded partnership working to increase the participation of racially and ethnically marginalized students in graduate programs in the physical sciences, has conducted work on holistic admissions. After learning about their studies, our department invited two members of its management team, Casey Miller and Julie Posselt, to lead a workshop for faculty serving on our graduate recruiting committee. As a result of that workshop, faculty members decided on five categories for a rubric that aligned with both their previous experience from reviewing applications and the recommendations of the workshop leaders: academic preparation, research experience, noncognitive competencies, fit with program, and GRE scores. (Iterations of

our rubric since the pandemic no longer include GRE scores.)

Each of the categories was then further divided into subcategories that mapped onto specific information about the applicant, such as their technical skills, their GPA in physics courses, and whether their research interests aligned with those of faculty members. Information to assess the subcategories, of which there are 18 in total, comes from the applicant's materials, which include transcripts, a CV, a personal statement, a research statement, and letters of recommendation. (The post-pandemic rubric, which eliminated the use of GRE scores, now contains only 16 subcategories.) To evaluate the nonacademic categories on the rubric, we asked applicants to respond to specific prompts in their personal and research statements. Those prompts broadly map onto at least one subcategory of the rubric.

One of the subcategories rates applicants on their contributions to diversity in physics through their research, teaching, or volunteering efforts (and not simply based on whether they belong to an underrepresented group in physics). Because public universities in Michigan—as in many other US states—are legally prohibited from discriminating against or granting preferential treatment to applicants based on race, sex, or ethnicity, such a scoring system ensures our admissions practices are compliant with state law.

A subset of the admissions committee then rates applicants as low, medium, or high on each subcategory, with clear criteria for what constitutes each level. By using a limited number of ratings on our rubric, we hoped to help admissions committee members avoid getting bogged down debating small differences, such as the distinction between a 3.70 and a 3.75 physics-major GPA. Although it's not included on the rubric, faculty members are also asked to note which subfields in physics and astronomy the applicant expressed interest in.

After each application has been evaluated by members of the admissions committee, a total score is calculated, based on a weighted average of the five categories. The applications, scores, and subscores are then sent to faculty representatives in each of the department's major research areas. They then make a list of applicants in their research area to whom they would like to extend an offer. Because the number of offers depends on funding and research-area needs, we do not make them based on a cutoff rubric score. Instead, we use the total score as a guide for which applicants we might want to admit.

Figure 2 presents a schematic overview of our new admissions process. Initial feedback from faculty who have served

		Most common rubric score		
		Low	Medium	High
ACADEMIC PREPARATION	Physics coursework	✗		✓
	Math coursework		✗	✓
	All other coursework			✓✗
	Academic honors	✗		✓
RESEARCH EXPERIENCE	Variety/duration of research			✓✗
	Quality of work		✗	✓
	Technical skills		✗	✓
	Research dispositions		✗	✓
NONCOGNITIVE COMPETENCIES	Achievement orientation		✗	✓
	Conscientiousness		✗	✓
	Initiative		✗	✓
	Perseverance		✗	✓
FIT WITH PROGRAM	Alignment of research		✗	✓
	Alignment with faculty		✗	✓
	Community contributions		✗	✓
	Diversity contributions	✗		✓
GRE SCORES	General GRE scores			✓✗
	Physics GRE score	✗	✓	

✓ Admitted ✗ Nonadmitted

FIGURE 3. THE MOST COMMON SCORE on each rubric subcategory for admitted and nonadmitted applicants. A rating of high was the most common score for admitted applicants across the subcategories; a rating of medium was the most common score for nonadmitted applicants. Iterations of the rubric since the pandemic no longer include the two GRE-related subcategories. (Adapted from ref. 14; CC BY 4.0.)

on our admissions committee has been positive. They like that the rubric provides guidelines for how to review applications and that it defines the measures of success. They also believe that the rubric has not increased the time it takes for review, which remains between 15 and 30 minutes per application.

Did we succeed?

The goal of rethinking our admissions process was to make it more equitable. To see if that happened, we looked at the initial three years of data. The results are promising.¹⁴ At the time of the study, the university admissions system collected only binary sex data on applicants and no racial or ethnic data. Since then, the system has been updated to allow applicants to disclose their gender, race, and pronouns if they want.

We first looked at how faculty assigned scores to the different applicants. If the rubric was useful for determining whom to admit, we would expect applicants who were admitted to have higher scores than those who were not. That was indeed what we found: Admitted applicants generally had higher scores on rubric subcategories than nonadmitted applicants. The most common rating among admitted applicants was high; among nonadmitted applicants, it was medium (see figure 3).

Next we looked at whether the rubric was equitable with respect to sex. If that were the case, we would expect males and females to have similar scores on average. Aside from a few subcategories, that is what we found, and we believe those exceptions reflect the rubric capturing known systematic issues. For example, males had higher rubric scores than females did on the GREP subcategory, and females had higher rubric scores than males on community and diversity contributions.

But we've long known that males do better on the GREP than females, so it is no surprise that the rubric would measure that. Similarly, females are more likely to serve as volunteers and are often expected to take on more outreach and community-building efforts in academia. So we should not be surprised that females earned higher scores than males on those rubric sections.

We then looked at how applicants from different types of institutions performed on the rubric. Our analysis considered the overall selectivity of the institution and, as a proxy for department size, the typical number of physics degrees it awarded. Based on our experiences, we assumed that applicants from larger departments or more selective institutions had access to more resources and opportunities than applicants from smaller departments or less selective institutions. For example, an applicant from a larger department or more selective institution might have more research opportunities or have access to more advanced physics courses, and those differences might be reflected in scores on the rubric.

But aside from the GREP subcategory, we did not find any consistent differences between applicants from different types of institutions. What was most surprising was that our faculty members did not rate applicants from smaller departments lower than applicants from larger departments on the research subcategories. But prior studies have found that undergraduate students with limited research experience may not apply to graduate school in the first place,¹⁵ which may explain why we did not find differences on those subcategories.

Lessons learned

In addition to thinking about equity in terms of rubric scores, we also considered how the rubric affected the number of female and underrepresented racial minority students who enrolled in our program. Just because applicants receive similar scores doesn't mean that admissions decisions are made along the same lines. For example, if faculty members had to choose between two comparable applicants, they might consider criteria outside the rubric to help make a distinction between the applicants. We did not find that to be the case. Since implementing the rubric, the percentage of admitted applicants who are female has more than doubled, from 13% to 31%, and the percentage of admitted applicants who are of an underrepresented racial minority group has increased from 9% to 12%. But those rates are still far from what we would hope for to achieve parity in representation.

Finally, we examined whether the rubric fundamentally changed our admissions process. We put countless hours into creating the rubric, but we still hadn't determined whether our department was basing its admissions decisions on a broader set of criteria or still relying mainly on grades and test scores. So we used machine learning to create models of our admissions process before and after we started using the rubric. Because we didn't have access to the qualitative parts of the application, such as personal and research statements for both time periods, our models used only quantitative aspects, such as the GRE scores and GPA. It also took into account the applicant's undergraduate institution and their physics subfield of interest.

We found that before we started using the rubric, our model could correctly predict whether three out of every four

See it online

Over the last 20 years, the number of students graduating with bachelor's degrees in physics has nearly doubled, but the number of positions at physics graduate programs has only grown slightly. Visit physicstoday.org/grad-admissions to learn about that trend and its impact on efforts to make admissions more equitable.

applicants would be admitted based on only the applicant's GREP score, GPA, and score on the quantitative portion of the standard GRE. The data from after we started using the rubric are murkier. Those three numbers are no longer determinative of whether an applicant will be admitted, which does make it seem like we are evaluating applicants on a broader set of criteria.

But even when we used the rubric scores to build a model, the resulting simulation was not able to predict whether a given student will be admitted. Perhaps the lack of a few predictive features signifies that our admissions process has become more holistic and that the rubric has created multiple routes to admission. We're currently working on determining what parts of the application are driving admissions decisions so that we can know for sure.

Using the rubric, our department has admitted more applicants from underrepresented groups in physics without increasing the time required for faculty to review applications. Based on that experience, we recommend that other physics departments implement rubrics in their admissions process to help evaluate applicants on a wider range of criteria than simply grades and test scores. But using a rubric will not result in a more equitable admissions process unless it is implemented properly. Departments need to ensure that their process also reflects a commitment to equity.

To do so, admissions committees should ensure that their members do indeed use the rubric to review each application. The committee itself should also be as diverse as possible so that it is reflective of the applicant pool.¹⁶ Finally, we recommend that departments conduct regular reviews of their admissions processes. Just because a department has always done admissions in a certain way does not mean that they need to continue doing so, especially if the data suggest that their process is not aligned with the goals of their program.

What's next

We started using the rubric to evaluate applicants for the class of graduate students who enrolled in fall 2018. Those students are now beginning to complete our program, which means that we are just starting to understand how changes to our admissions process may have affected other areas of our program, such as time to PhD candidacy and time to completion. The initial results suggest that students admitted under the rubric are no more likely to leave the program than students admitted before we began using it.

Some might worry that reducing the role of the GREP and undergraduate GPA will lead to a weaker or less prepared incoming class and, as a result, a longer time to degree, but recent studies in engineering suggest that that hasn't happened in practice.¹⁷ Additional work with rubric-based admissions will confirm whether those concerns are valid in physics graduate programs.

While our work suggests that rubrics with a broader range

of admissions criteria make admissions more equitable, the evaluation of applicants is only one part of the process. Another part of improving equity in physics graduate education is ensuring that all groups have a fair chance in the admissions process, and doing so requires that currently underrepresented groups be included in the applicant pool. That means that future efforts at making graduate admissions more diverse and equitable need to focus on recruitment.

Moreover, once we've admitted diverse students, we need to retain them in our programs. To do so, we need to consider how our courses, qualifying and comprehensive exams, and advising and mentoring structures affect our retention efforts. At Michigan State, based on feedback from students and faculty, we've removed our qualifying exam requirement and changed the timeline for when students take their comprehensive exams. We've also added additional mentoring support for students before they form their thesis committee. We are not alone in making those types of changes: Other physics departments have also changed their exam requirements and added additional mentoring support for graduate students.¹⁸

The future of physics can be diverse, equitable, and inclusive if we work to make it so. Rethinking graduate admissions is one place to start.

REFERENCES

1. C. Mathis, S. Southerland, *Phys. Teach.* **60**, 260 (2022).
2. M. Dancy, A. Hodari, <https://arxiv.org/abs/2210.03522>.
3. J. R. Posselt, *Inside Graduate Admissions: Merit, Diversity, and Faculty Gatekeeping*, Harvard U. Press (2016).
4. J. Guillochon, "GRE requirements and admissions fees for US/Canadian astronomy and physics programs," spreadsheet available online at <https://tinyurl.com/33tdfvjc> (last updated 1 June 2023).
5. G. Potvin, D. Chari, T. Hodapp, *Phys. Rev. Phys. Educ. Res.* **13**, 020142 (2017).
6. C. W. Miller et al., *Sci. Adv.* **5**, eaat7550 (2019).
7. N. J. Mikkelsen, N. T. Young, M. D. Caballero, *Phys. Rev. Phys. Educ. Res.* **17**, 010109 (2021).
8. K. M. Whitcomb, S. Cwik, C. Singh, *AERA Open* (2021), doi:10.1177/23328584211059823.
9. R. L. Matz et al., *AERA Open* (2017), doi:10.1177/2332858417743754.
10. A. A. Eaton et al., *Sex Roles* **82**, 127 (2020).
11. C. W. Miller, *Status: A Report on Women in Astronomy*, January 2015, p. 1.
12. C. Miller, J. Posselt, *Physics* **13**, 199 (2020).
13. I. Bohnet, *What Works: Gender Equality by Design*, Belknap Press (2016).
14. N. T. Young et al., *Phys. Rev. Phys. Educ. Res.* **18**, 020140 (2022).
15. G. L. Cochran, T. Hodapp, E. E. A. Brown, in *2017 Physics Education Research Conference Proceedings*, L. Ding, A. Traxler, Y. Cao, eds., American Association of Physics Teachers (2017), p. 92.
16. S. F. Roberts et al., *Educ. Sci.* **11**, 270 (2021).
17. L. Stiner-Jones, W. Windl, "Work in progress: Aligning what we want with what we seek: Increasing comprehensive review in the graduate admissions process," paper presented at the 2019 ASEE Annual Conference and Exposition, 16–19 June 2019. Available at <https://peer.asee.org/33592>.
18. R. Barthelemy et al., *Phys. Rev. Phys. Educ. Res.* **19**, 010102 (2023). 