The usefulness of GRE scores

Many US physics departments are considering dropping the use of Graduate Record Examinations (GREs) in making admissions decisions (see, for example, the commentary by Alexander Rudolph, PHYSICS TODAY, June 2019, page 10). They are concerned that the exams contribute to the profession's nonrepresentative demographics. The American Physical Society (APS) Panel on Public Affairs is looking at adopting a similar position. Those decisions may be influenced by a widely publicized *Science Advances* paper entitled "Typical physics Ph.D. admissions criteria limit access to underrepresented groups but fail to predict doctoral completion," by Casey Miller and coauthors.[1]

Although that paper uses data provided by many physics departments, I found some serious statistical flaws in its analysis. Contrary to its conclusions, proper statistical analysis of even the incomplete published features of the data indicates that an equal-weight sum of the quantitative and physics GREs is somewhat better than undergraduate grade point average at predicting who will graduate.[2]

I believe the key issues raised include the need for more transparency and statistical literacy in handling data, but the effects of graduate admissions policies themselves are also important. Systematic uncertainties in estimating the effects of using GREs in admissions decisions would remain even after a proper analysis of more complete data,[2] as is typical for any attempt to estimate causal parameters from observational data.[3] Therefore, it may be worth trying a more robust way to get information on those effects.

Given the fairly large number of physics departments that are uncertain about what the GRE's role in the admissions process should be, APS could ask for departments to volunteer in a randomized controlled trial. Some departments would be assigned to GRE-aware admissions and others to GRE-blind admissions. Ideally, the assignments would be switched after a year. Beyond graduation rates, various other outcomes of interest could be tracked. Departments could participate in long-term follow-up even if they committed to only two years of randomized admissions policy. Incremental costs above the already labor-intensive selection procedures should be small, perhaps even negative, if one counts the time saved in decision making.

Although the information obtained might be inconclusive, at least the setup could be a model for approaching policy issues scientifically and honestly. That's important when we consider that our credibility on the really big issues—climate, for example—has been challenged by people who wrongly claim we are just pushing political positions disguised as science.

A model of efficient competition in admissions: The more desirable programs attract students who are more likely to succeed.

References

1. C. W. Miller et al., *Sci. Adv.* **5**, eaat7550 (2019); addendum, https://arxiv.org/abs/1906.11618.
2. M. B. Weissman, https://arxiv.org/abs/1902.09442.
3. S. Greenland, *Epidemiology* **14**, 300 (2003).

Michael B. Weissman
(mbw@illinois.edu)
University of Illinois at Urbana-Champaign

Alexander Rudolph's commentary correctly notes that achieving greater diversity in physics requires revamping admissions criteria: You only get what you select for. However, the recommendation against using Graduate Record Examination (GRE) scores draws heavily on a study by Casey Miller and coauthors.[1] That study has three major problems:

- The study measured performance with a binary variable: completion. Eliminating gradations of performance obscures relationships that may be present in more granular data. A large meta-analysis examined student performance with fine-grained measures—for example, research productivity, faculty ratings of student work—and found significant predictive power in GRE scores.[2]
- The work by Miller and coauthors included program rank as an explanatory variable, despite its being strongly correlated with GRE scores. When two or more such variables are strongly correlated, a regression routine cannot easily determine which variable should get the larger coefficient; different coefficient choices could fit the data similarly well. Consequently, coefficient estimates will have large uncertainties.[3] Thus the estimated coefficient of GRE score will almost certainly have a magnitude comparable to the estimate's uncertainty.
- The predictive power of program ranking actually fits a model of efficient competition in admissions: The more desirable programs attract students who are more likely to succeed. Moreover, a student who is weak by one measure can gain admission by demonstrating strength in another measure. Such cases may camouflage correlations between student performance and other explanatory variables.[4] Of course, there are other plausible explanations for the predictive power of program ranking, but nothing in the cited work enables readers to choose among explanations.

Admissions criteria are ultimately about values, and it is wholly appropriate to include diversity of backgrounds among them. However, if performance is also valued, then valid predictors of performance should also be included. The Miller study does not demonstrate that GRE scores lack predictive power, and it should not be cited uncritically.