chronicle.com

Congress Should Scrutinize Higher Ed's Use of Predictive Analytics, Watchdog Says
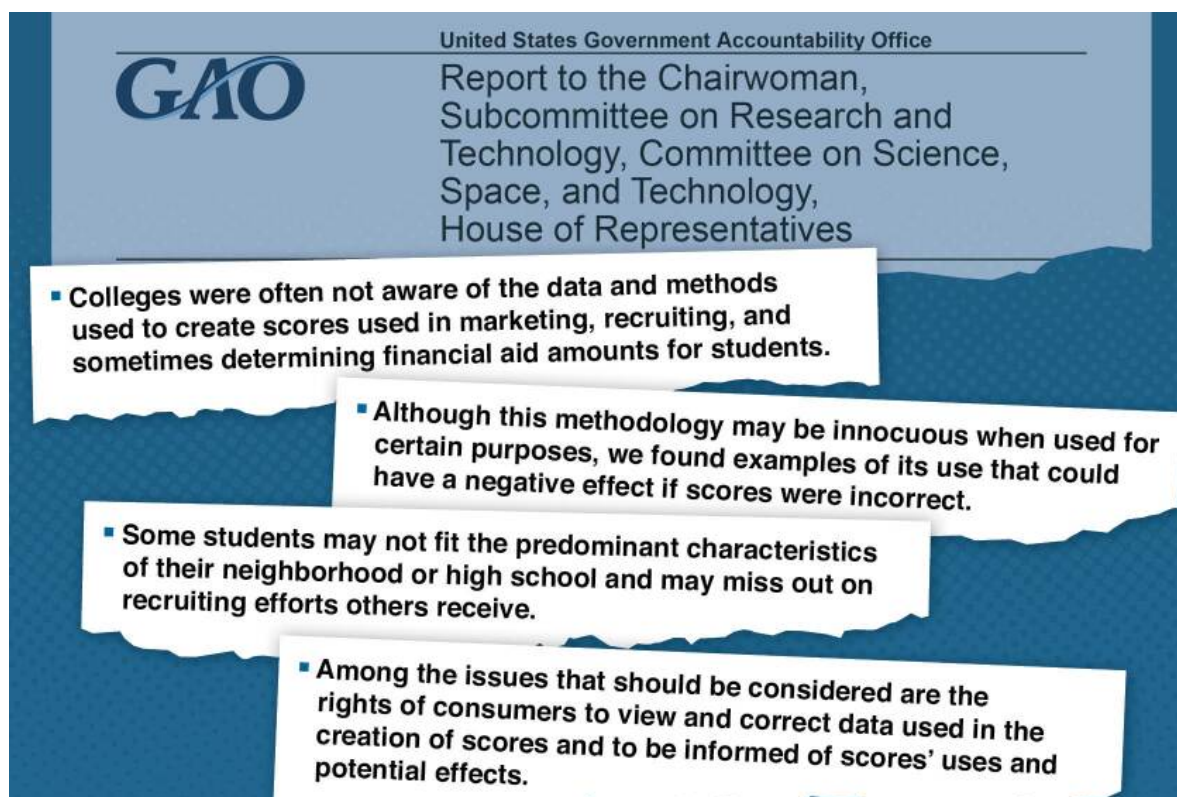
Dan Bauman

6–7 minutes

Illustration by The Chronicle

They're the invisible infrastructure that colleges and universities rely on to target prospective students for recruitment, to build financial-aid offers, and to monitor student behavior. Now, a new report from the Government Accountability Office is urging

Congress to probe how higher education uses these consumer scores, algorithms, and other big-data products, and to consider who stands to benefit most from their use — students or institutions?

The GAO also encouraged Congress to consider bolstering disclosure requirements and other consumer protections relevant to such scores.

"Among the issues that should be considered are the rights of consumers to view and correct data used in the creation of scores and to be informed of scores' uses and potential effects," the office recommended.

Predictive analytics have been heralded as a means to improve many facets of higher education, like bolstering retention and more equitably apportioning institutional aid, but they are not without their detractors. Concerns for student privacy abound. And critics worry poorly designed or understood models can embed and automate discriminatory behavior across an institution's operations.

"Colleges were often not aware of the data and methods used to create scores used in marketing, recruiting, and sometimes determining financial-aid amounts for students," the GAO wrote in its report, summarizing an exchange the agency had with one industry expert, and describing higher ed's uses of predictive analytics that most concerned the office.

The sheer complexity of certain algorithms presented another challenge. After reviewing one scoring product, used to identify and flag students at risk of dropping out or transferring to another college, GAO researchers observed the breadth of variables — "potentially hundreds" — evidently relevant to the underlying model's assessment of risk.

Of most concern to the GAO? The weight assigned by certain models to an individual student's points of origin — including the neighborhood where they live and the high school they attend.

"Although this methodology may be innocuous when used for certain purposes, we found examples of its use that could have a negative effect if scores were incorrect," the agency wrote. Put more bluntly: In a country where race, wealth, and geography are inextricably linked, models and algorithms can rationalize and endorse biases against minority and low-income students, even if such products only factor residency information into their scores and assessments.

As an example, GAO refers to an unnamed scoring product used by admissions offices to identify students who "will be attracted to their college and match their schools' enrollment goals" — in essence, a lead-generation service. A prospective student's neighborhood and high school dictates which lists their contact information will appear on. Each list is in turn assigned its own respective set of scores — measures of each cohort's shared socioeconomic, demographic, and "educationally relevant"

characteristics. Using these scored lists, admissions professionals can deploy recruitment strategies tailored to their respective institution's admissions goals.

But what about high-achieving students enrolled at poor-performing or underfunded high schools? How can a college target those students for recruitment if where they live and learn precludes them from being included in colleges' enrollment efforts? To the Government Accountability Office, it's a recipe for disparate treatment.

"Some students may not fit the predominant characteristics of their neighborhood or high school and may miss out on recruiting efforts others receive," the GAO warns.

To guard against pitfalls like these, colleges and universities should consult with diversity, equity, and inclusion professionals, Jenay Robert, a researcher at Educause, a nonprofit association focused on the intersection of technology and higher education, said in a statement. If analytics staff don't work with diversity experts with their institution's specific needs in mind, "big-data analytics can do more harm than good," she said.

Higher education also lacks broadly accepted policies on this topic, she said: "We've yet to establish a widely used ethical framework that puts forth best practices for engaging in big-data analytics."

In the absence of federal regulations on the use of algorithms, colleges and universities have been left to reconcile how their institutional interests comport with the interests of individual students — and to what extent this usage serves the broader public good. And in theory, there should be no distinction. For

instance, when an institution uses predictive analytics to allocate scholarship aid to those who might have forgone college without it, the public good is served.

But reality is often more complicated. Big-data products and models afford colleges and universities capabilities for fine-grained analysis that may not have been previously available to most admissions offices. In testimony to the GAO, one industry expert posited a scenario in which a college might draw certain conclusions from a prospective student's repeated campus or website visits — conclusions ultimately resulting in less scholarship money awarded to this prospective student relative to similarly situated peers.

For a college, the calculus is simple: Why offer significant scholarship aid to a student who is likely to attend your institution regardless? For the country, though, a different dilemma emerges: Even if there is more scholarship money to go around, is the public good really best served when a student is penalized for making use of campus visits and online research prior to embarking on one of the most significant investments in American life?