Best vs. All: Equity and Accuracy of Standardized Test Score Reporting

RICE UNIVERSITY
Social Sciences
Department of Economics

Mingzi Niu
(with Sampath Kannan, *University of Pennsylvania*,
Aaron Roth, *University of Pennsylvania*, and,
Rakesh Vohra, *University of Pennsylvania*)

Fairness Concern about Standardized Tests

- Standardized tests are essential to college admission.
- College score-submission policies:
  1. "***Report Max***"(superscoring): applicants can choose which scores to submit.
  2. "***Report All***": applicants must submit all test scores.
- In reality, not all groups retest at the same rates.
- A source of unfairness: only *some* applicants have the resources to take the test multiple times.
  - What happens under different score-submission policies?
  - How do these two policies compare in terms of equity and accuracy?

Model & Notation

- Two types of students, High (H) or Low (L). The prior that a student is of type H is p.
- The test generates a score $s \in \{A, B\}$.
- Test accuracy $Pr(s = A|H) = Pr(s = B|L) = \alpha > 0.5$.
- Two categories of students:
  - Category 1 can only take the test once. The proportion of category 1 students is ϕ.
  - Category 2 can take the test up to k times.
- The College's payoff: 1 for admitting a type H student; -1 for admitting a type L.
- Denote $\hat{p}_k = \frac{\phi(1-\alpha)+(1-\phi)(1-\alpha^k)}{\phi+(1-\phi)[2-\alpha^k-(1-\alpha)^k]}$, $\hat{p}'_k = \frac{\phi\alpha+(1-\phi)\alpha^k}{\phi+(1-\phi)[\alpha^k+(1-\alpha)^k]}$, and $p^*_k = \frac{(1-\alpha)^{k-2}}{\alpha^{k-2}+(1-\alpha)^{k-2}}$.

Assumptions

1. Students know their type, but cannot credibly convey it to the College except through a test.
2. Students differ in their ability to access multiple signals: only Category 2 can make testing decisions *adaptively* to their previous test scores.
3. The type distribution (p) and test accuracy (α) are both category independent.

Main Result

"Report All" is superior to "Report Max" both from the perspective of equity but also from the perspective of the college.

Equity

Compare the false positive (FP) and false negative (FN) rates under "Report Max" with:

- *First-Score Equilibirum* under "Report All":

FN	**(1, *H*)**	**(2, *H*)**
Max	$1-\alpha$	$(1-\alpha)^k$
All	$1-\alpha$	$1-\alpha$

FP	**(1, *L*)**	**(2, *L*)**
Max	$1-\alpha$	$1-\alpha^k$
All	$1-\alpha$	$1-\alpha$

"Report All" achieves parity across categories, whereas "Report Max" always favors the advantaged (Category 2) students.

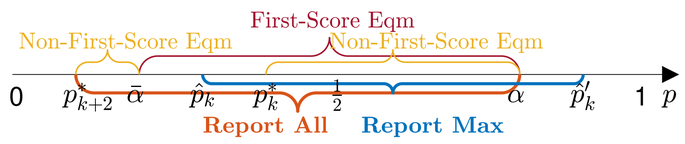
- *Non-First-Score Equilibrium* under "Report All": inequity across categories remains but it is reduced compared to "Report Max".

Accuracy

- The positive predictive value of "Report All" exceeds that of "Report Max": the admitted class has a higher proportion of High types under "Report All".
- The expected payoff to the College is also higher under "Report All" than 'Report Max".

Other Results

Non-Trivial Equilibrium Outcomes



1. Under "Report Max":
   (a) A nontrivial (separating) equilibrium if and only if $p \in [\hat{p}_k, \hat{p}'_k]$.;
   (b) The nontrivial equilibrium is unique: the College accepts a student if the reported score is A and rejects otherwise. Category 2 students take the exam as many times as they need to get an A score (up to k times).
2. Under "Report All":
   (a) A first-score equilibrium in which the admission depends solely on the first score exists if and only if $p \in [1-\alpha, \alpha]$.
   (b) An non-first-score equilibrium exists if and only if $p \in [p^*_{k+2}, 1-\alpha] \cup [p^*_k, \alpha]$.
   (c) A single score of A yields admission and a transcript consisting entirely of B yields rejection for any $p \in (1-\alpha, \alpha)$.

NOTE The equilibrium under "Report All" is unique as the first-score equilibrium if $k = 2$ and $p \in (1-\alpha, \frac{1}{2})$.

Discussion

1. In some cases, "Report All" policy can have the same effect as enforcing that students take the exam only once.
2. "Report All" can also give well-resourced students an advantage, as a population:
   - $(2, H)$: report a more *accurate* signal;
   - $(2, L)$: pool with the lower-resourced students, providing a less accurate signal and an increased chance of admissions.
3. Policy evaluation in equity and accuracy:
   - Same effects of both policies when
     - $\alpha = 1$: the score signal is perfect , or
     - $k = 1$: the access to signals is equal.
   - Disparities between the two policies grow with the test inaccuracy $(1-\alpha)$ and unequal access to the test (k).
4. Tradeoff between equity and accuracy exists among equilibria under "Report All": the first-score equilibrium generates more *equal* admission outcomes across categories yet yields *lower* expected payoff to the College.

References

GRODSKY, E., J. R. WARREN, AND E. FELTS (2008): "Testing and social stratification in American education," in *Annual Review of Sociology*, vol. 34, pp. 385–404.

HUTCHINSON,B., AND M. MITCHELL (2019): "50 years of test (un) fairness: lessons for machine learning," in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 49–58.

IMMORLICA, N., K. LIGETT, AND J. ZIANI: "Access to population-level signaling as a source of inequality.," *Proceedings of the Conference on Fairness, Accountability, and Transparency*. , 2019, January., pp. 249-258.

JUNG, C., S. KANNAN, C. LEE, M. PAI, A. ROTH, AND R. VOHRA (2020): "Fair prediction with endogenous behavior," in *Proceedings of the 21st ACM Conference on Economics and Computation*, pp. 677–678.

MILLI, S., J. MILLER, A. DRAGAN, AND M. HARDT (2019): "The social cost of strategic classification," *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pp. 230–239.

Further Details

Paper available at:
`https://arxiv.org/abs/2102.07809`

Comments, suggestions, contact:

`mingzi.niu@rice.edu`
`kannan@cis.upenn.edu`
`aaroth@cis.upenn.edu`
`rvohra@sas.upenn.edu`