

Reducing Deficit Interpretations of Large-Scale Assessment Results Displays: Exploring Score Reporting Approaches to Support Equity

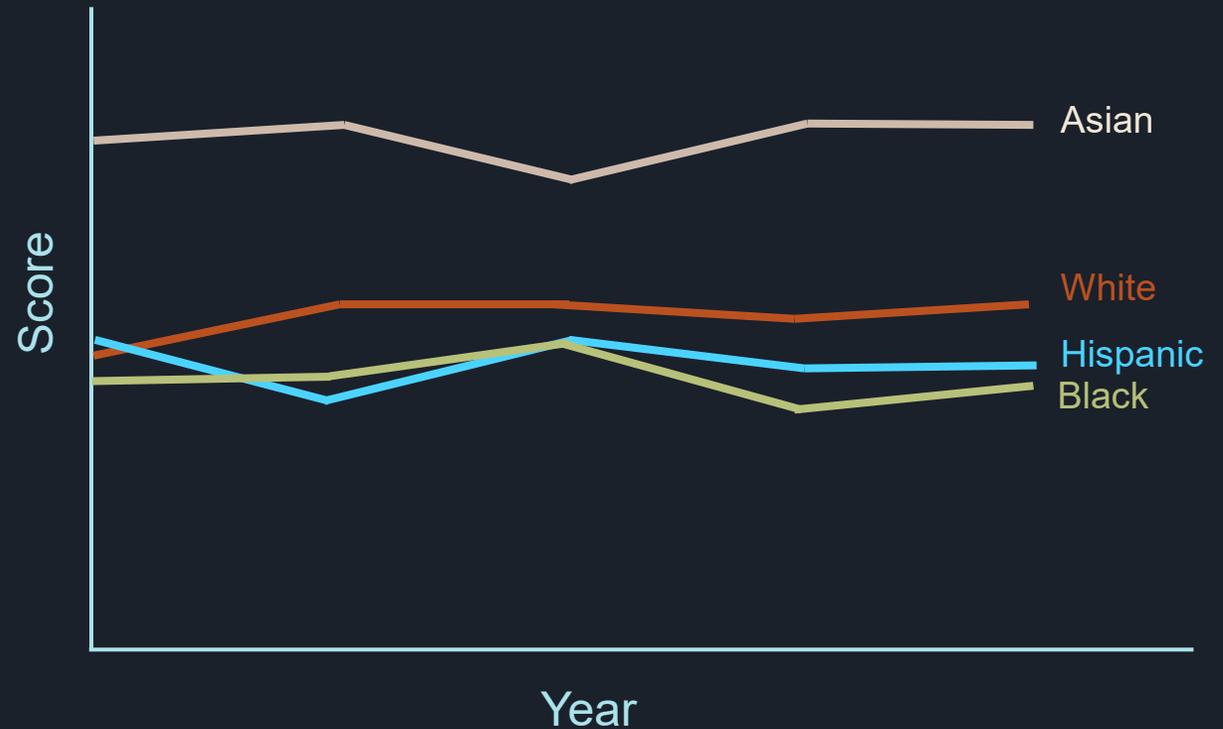
Emily Kerzabi, Katherine Castellano, Mari Kevelson, & Vinetha Belur

Contributors: Eli Holder, Leslie Villegas, Renee Savoie, Jennifer Cain, & Molly Prower

Importance of equity in group-score data displays

- Some of the most frequently used displays for reporting group score data can be the most misleading.

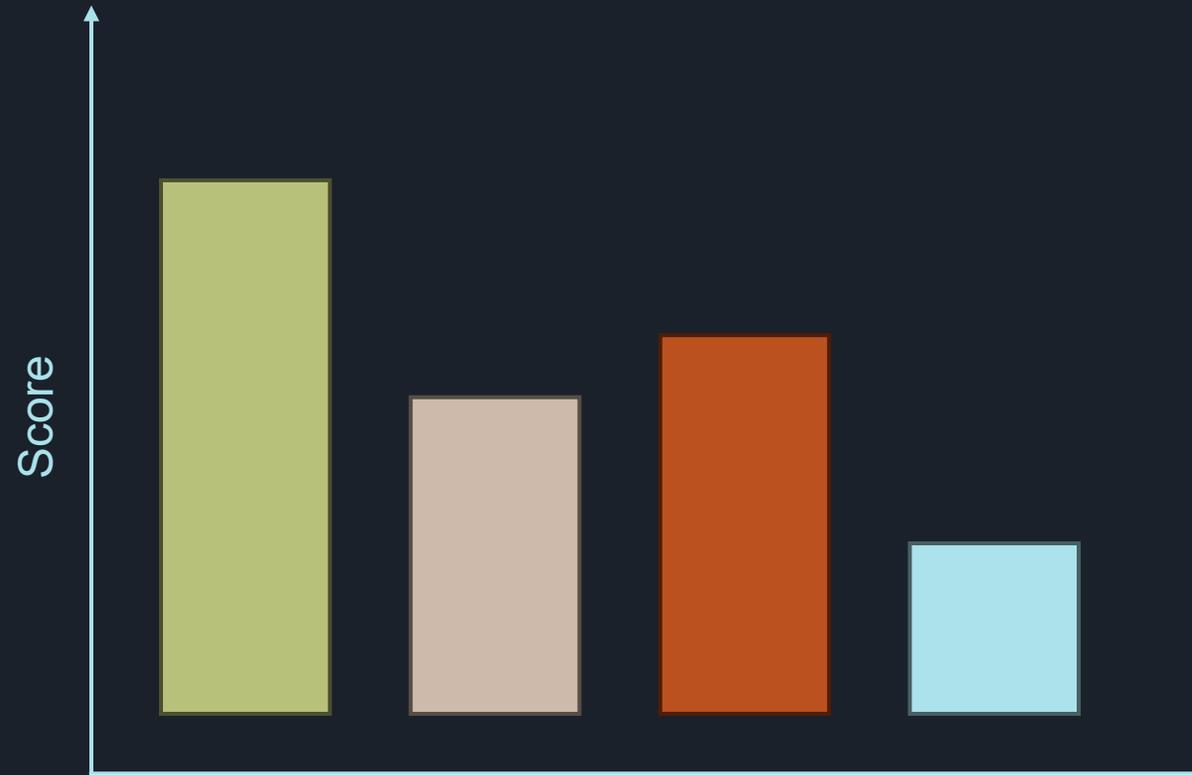
(Looks like all respondents within each group scored at the mean.)



Importance of equity in group-score data displays

- Some of the most frequently used displays for reporting group score data can be the most misleading.

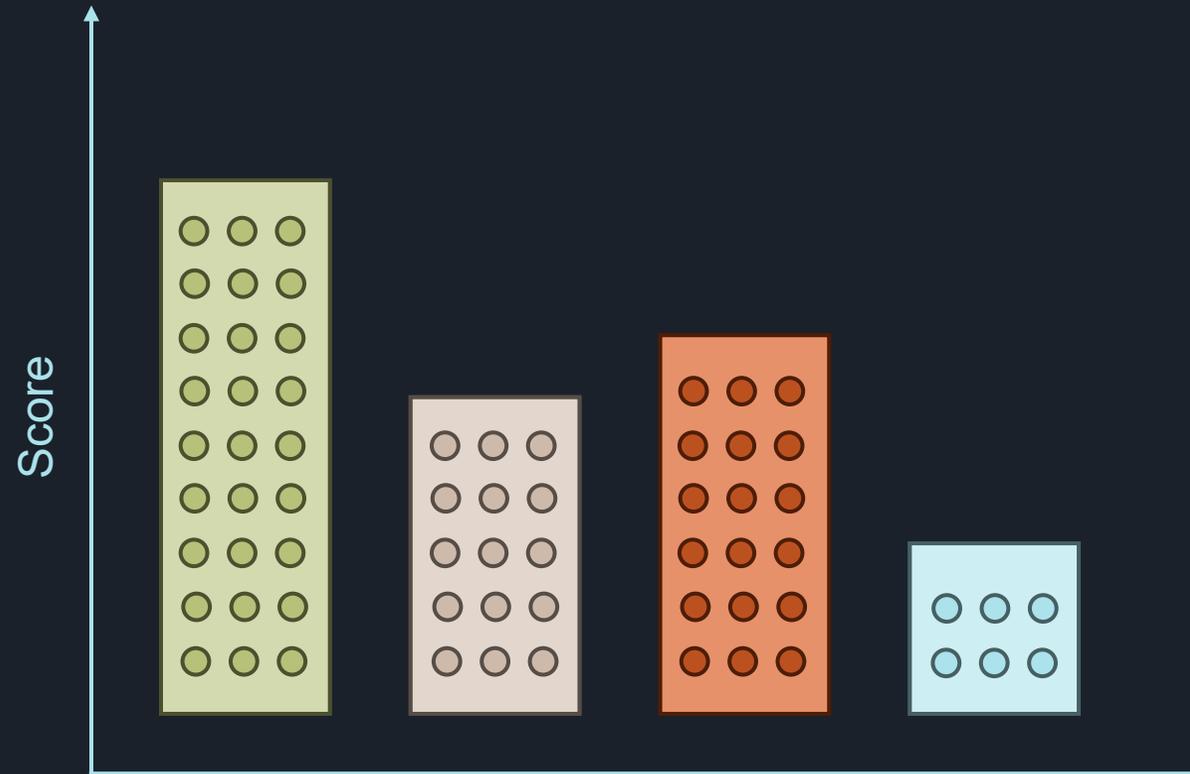
(Looks like all respondents scored within or at the top of each bar.)



Importance of equity in group-score data displays

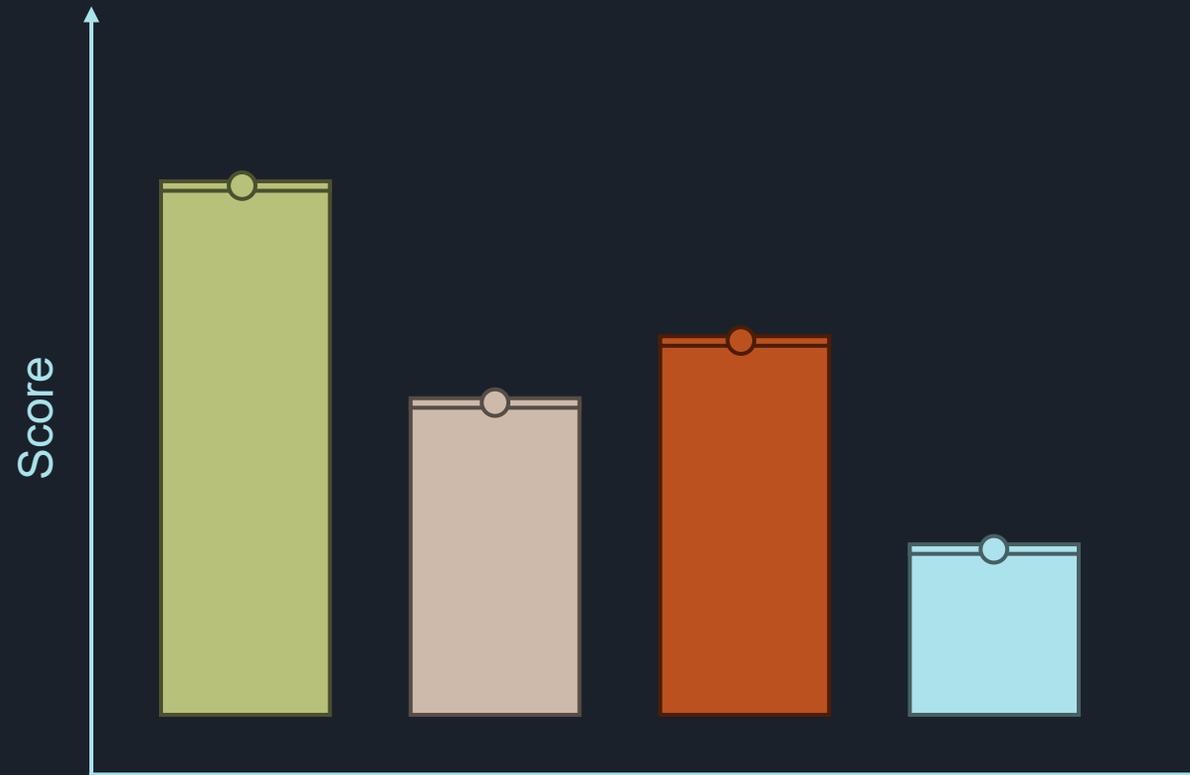
- Some of the most frequently used displays for reporting group score data can be the most misleading.

(Looks like all respondents in each group are uniformly distributed.)



Importance of equity in group-score data displays

- Some of the most frequently used displays for reporting group score data can be the most misleading.
- Reporting mean scores alone hides or distorts variability



Importance of equity in group-score data displays

- Some of the most frequently used displays for reporting group score data can be the most misleading.
- Reporting mean scores alone hides or distorts variability.



Importance of equity in group-score data displays

- Some of the most frequently used displays for reporting group score data can be the most misleading.
- Reporting mean scores alone hides variability.
- Attributing reasons for score differences to racial/ethnic groups can lead to deficit-thinking, attribution bias, and stereotyping.



What is deficit-thinking?

Restaurant Worker Pay



Deficit-thinking

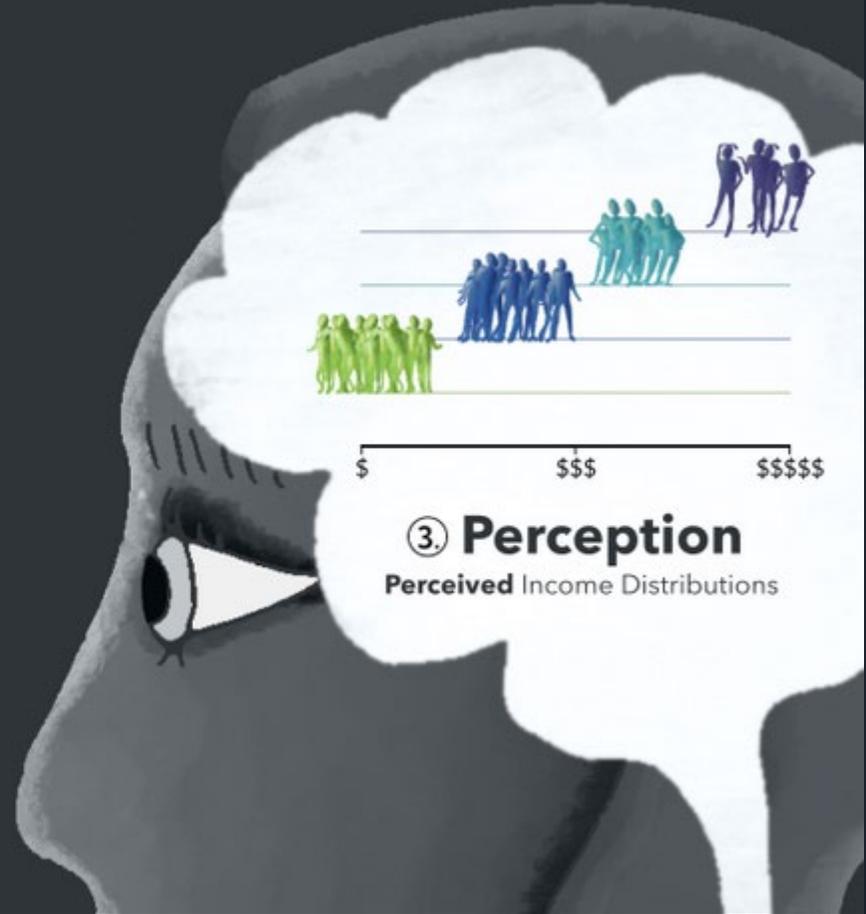
Attributing deficits to personal or internal characteristics

- Group A must work harder than Group B.
- Group A offers the best customer service.
- Group A smiles more.

Recognizing structural disparities

- Group A works in a busier restaurant than Group B

Can charts encourage misperceptions of dispersion?



Reproduced from *Unfair Comparisons* by Holder (2022)

“Deficit framing leads to **blaming**.
Blaming leads to **harmful stereotypes**.”

-- Holder, 2022

Our Investigation

Comprehensive Literature Review

Best practices for:

- Score reporting (44 articles)
- Anti-deficit and achievement gap reporting (22 articles)
- Equity forward data visualization (9 articles)

Stakeholder Interviews

20 interviews with users of large-scale assessment results displays (Dec 2023 - Jan 2024)

- 30% NAEP State Coordinators
- 30% Education Advocacy Groups
- 40% Researchers, Policymakers

Co-design of Data Displays

ETS research team (5)

UI/UX Expert

Equity and data visualization expert

NAEP State Coordinators (2)

Education Senior Policy Analyst

- Met 8 times over 2 months
- Designed equity-forward displays using NAEP math data



Main Takeaways

01

Use “gap” reporting appropriately

02

Display a range of achievement

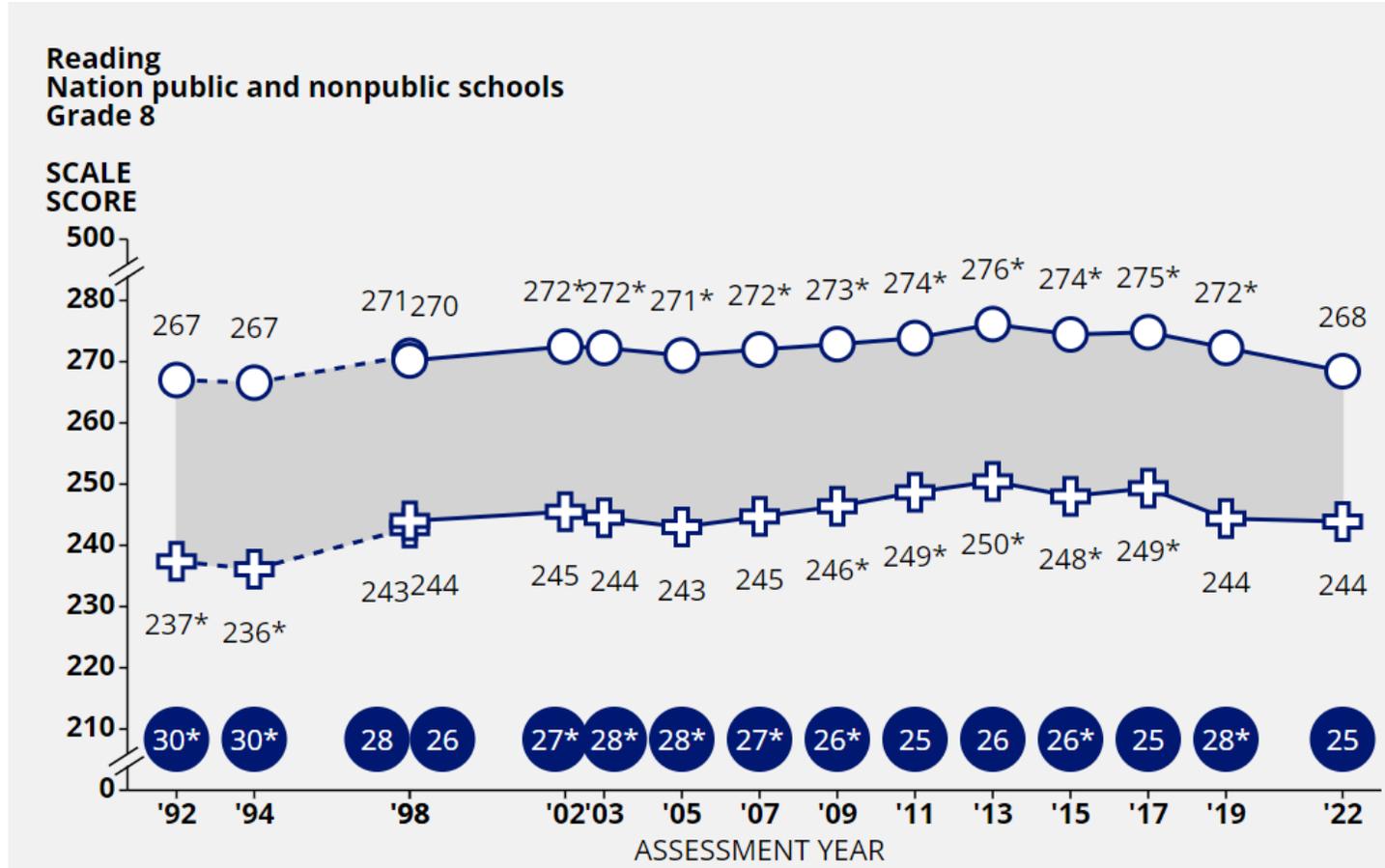
03

Include structural explanations for achievement differences

01

Use “gap” reporting
appropriately

Current Achievement Gap Displays





“[Achievement gap displays] reinforce false stereotypes.”



“Gratuitous reporting on so-called achievement gaps is more likely to do harm than good.”



“People want to know the bad story and try to fix it.”

The gap is really a debt

“When we see the ‘achievement gap’, what we are really seeing is **generations of inequalities that have compounded**. Rather than gap, this is better understood as a debt that should be owed to those whom these inequalities have affected.” — Ladson-Billings, 2006

Negative and racialized

“The term *achievement gap* has a negative and racialized history, and using the term **reinforces a deficit mindset** that is ingrained in U.S. educational systems.” — Shukla, 2022

Less support for equity-focused policies

The term *achievement gap* itself is **detrimental, as teachers and the general public** (Mturk sample) rated it lower as a priority than "ending racial inequality in educational outcomes" — Quinn et al., 2019; Quinn & Desruisseaux, 2022

02

Display a range of
achievement



“[A range display]
humanizes achievement
results.”

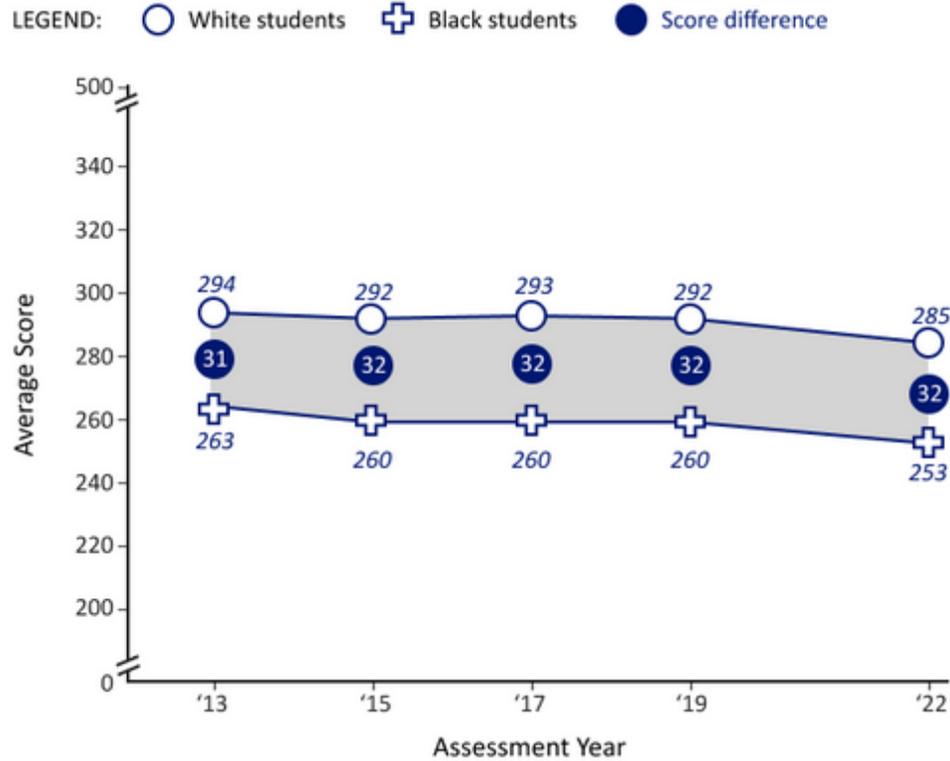


“Using percentiles reveals more than the mean. You can start teasing apart things (rural/urban, by state or location) on the lower end of the distribution.”

Gap vs. Range Reporting

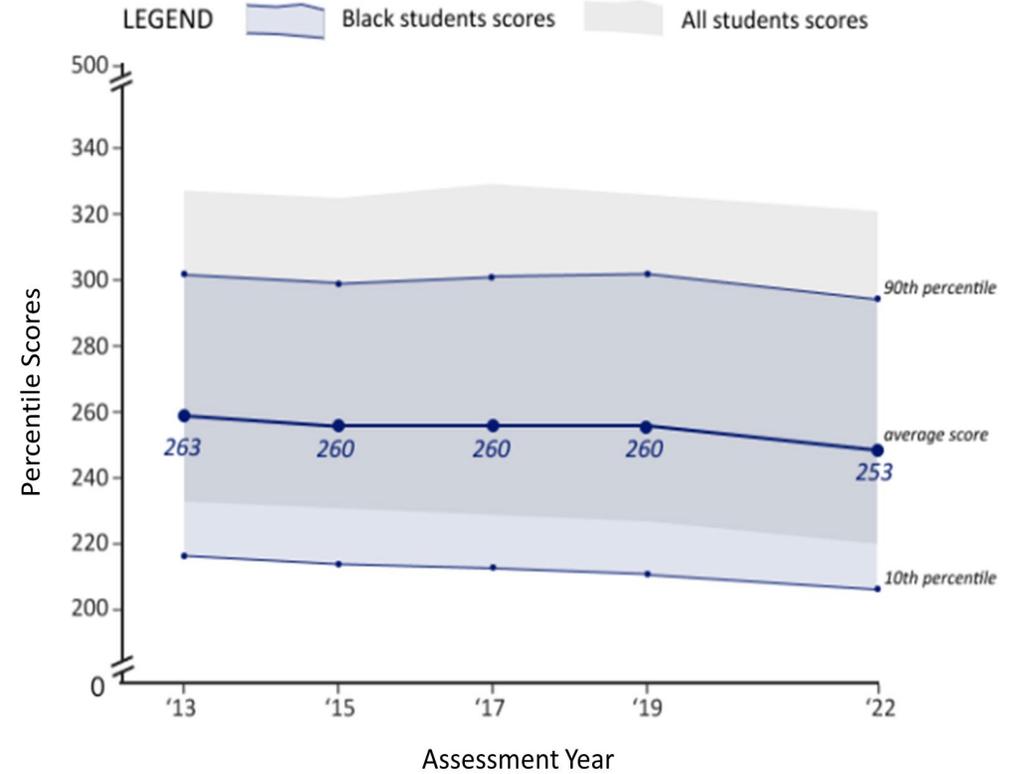
Average Scores

(Black students compared to White students)



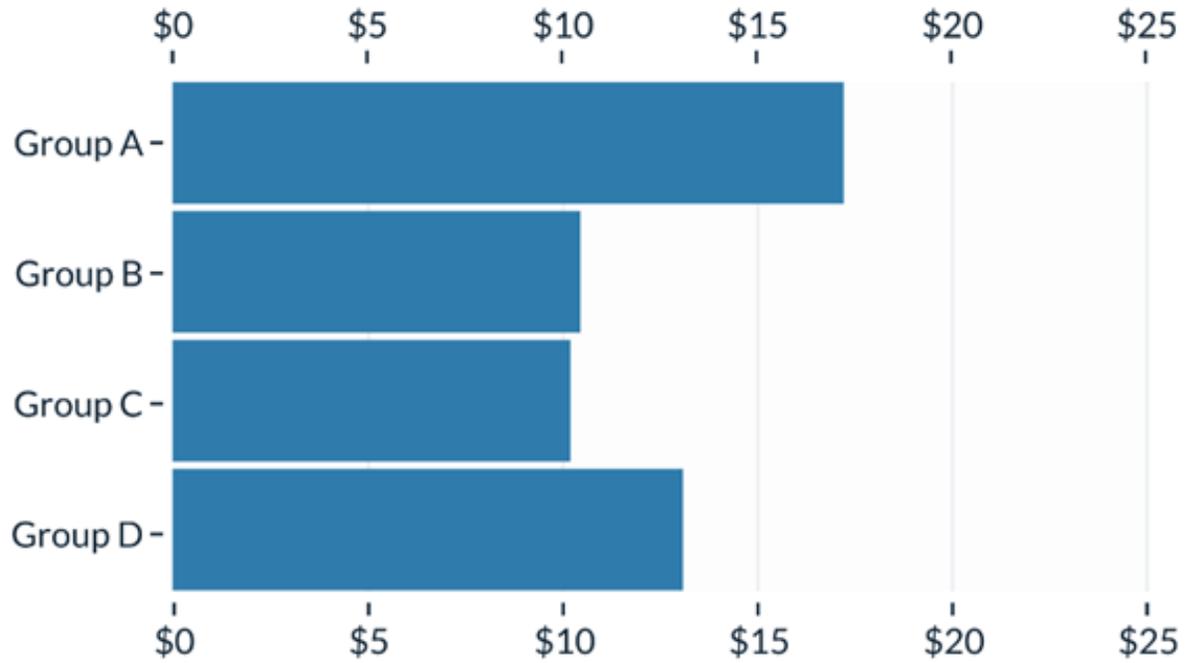
Percentile Scores

(Black students compared to All students)



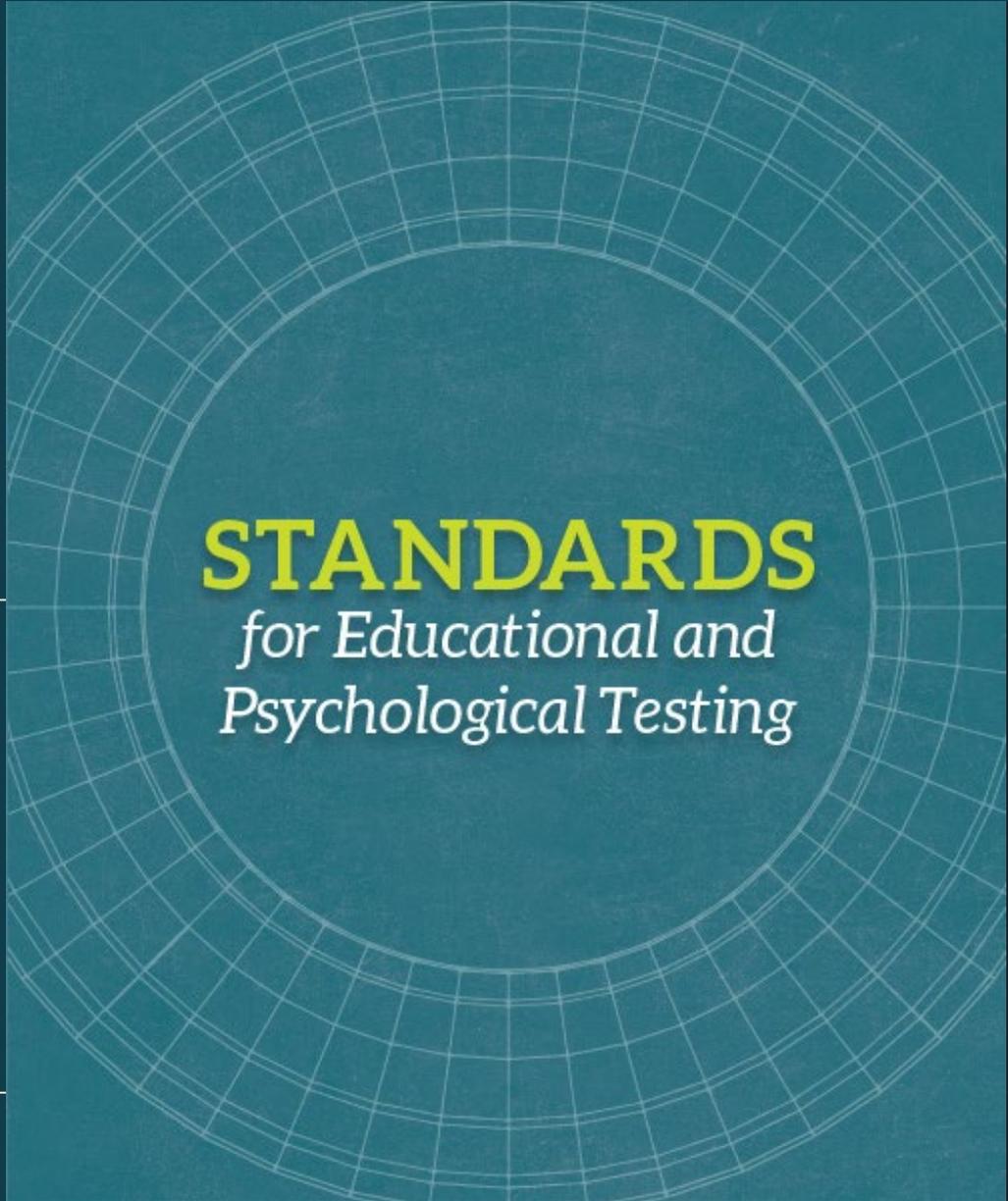
Trend in Grade 8 NAEP Mathematics

Bar vs. Jitter Plot



“When average or summary scores for groups of students are reported, they should be supplemented with additional information about the sample sizes and the shapes or dispersions of score distributions”

-- APA Testing Standard 12.18



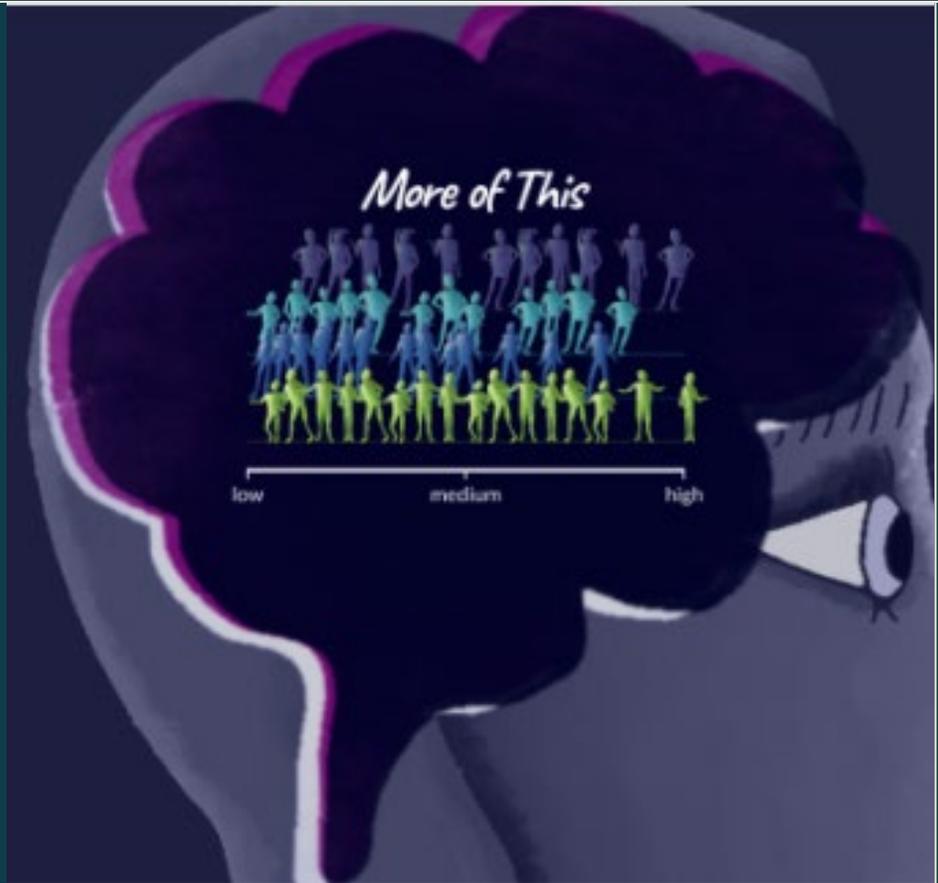
STANDARDS
for Educational and
Psychological Testing

From the Literature: Resist the Trap of False Simplicity



“It’s always better to tell a complex story that’s true than a simple story that’s false.”

-- Holder, 2022



03

Include structural
explanations for
achievement
differences



“Digging into why that context exists is pretty complex and can take time... you may not know exactly what the responses meant or why a particular state/district/region responded that way.”

Context allows for meaningful interpretation

“Reports of group differences in test scores should be accompanied by [relevant contextual information](#), where possible, to enable meaningful interpretation of the differences” – APA Testing Standard 12.17

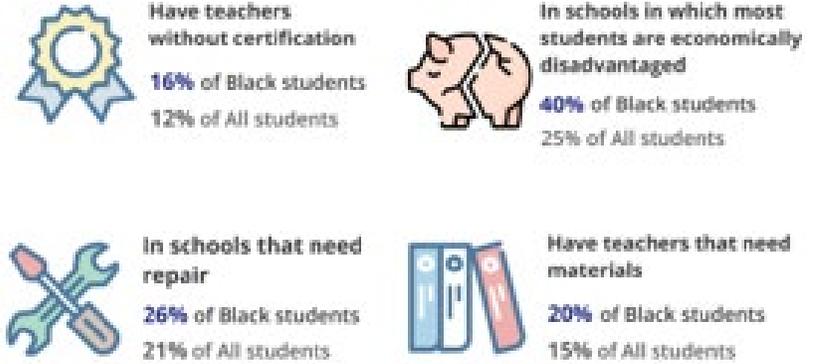
Compare similar contexts

By conducting intersectionality analyses, “using data from a variety of contextual OTL factors as the primary disaggregating variables and using effective visualization to present the within group differences among students from [various demographic subgroups](#) with different opportunities to learn.” – Brown et al., 2023 (p. 2)

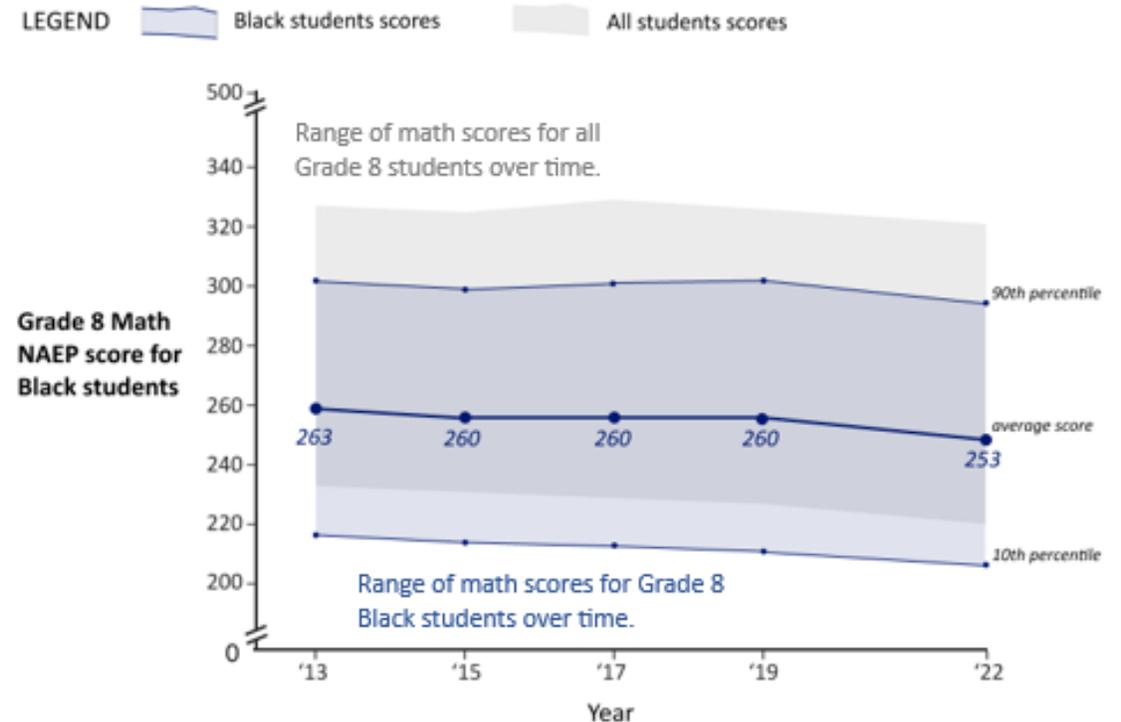
Conduct within-subgroup analyses

Calls for “[more within-subgroup analyses](#) using information from NAEP surveys to determine which factors may explain why some children within the same racial or ethnic subgroup perform better than others” – Hughes, 2023 (p. 26).

Higher percentages of Black students are in schools that lack critical resources, likely contributing to inequitable achievement outcomes

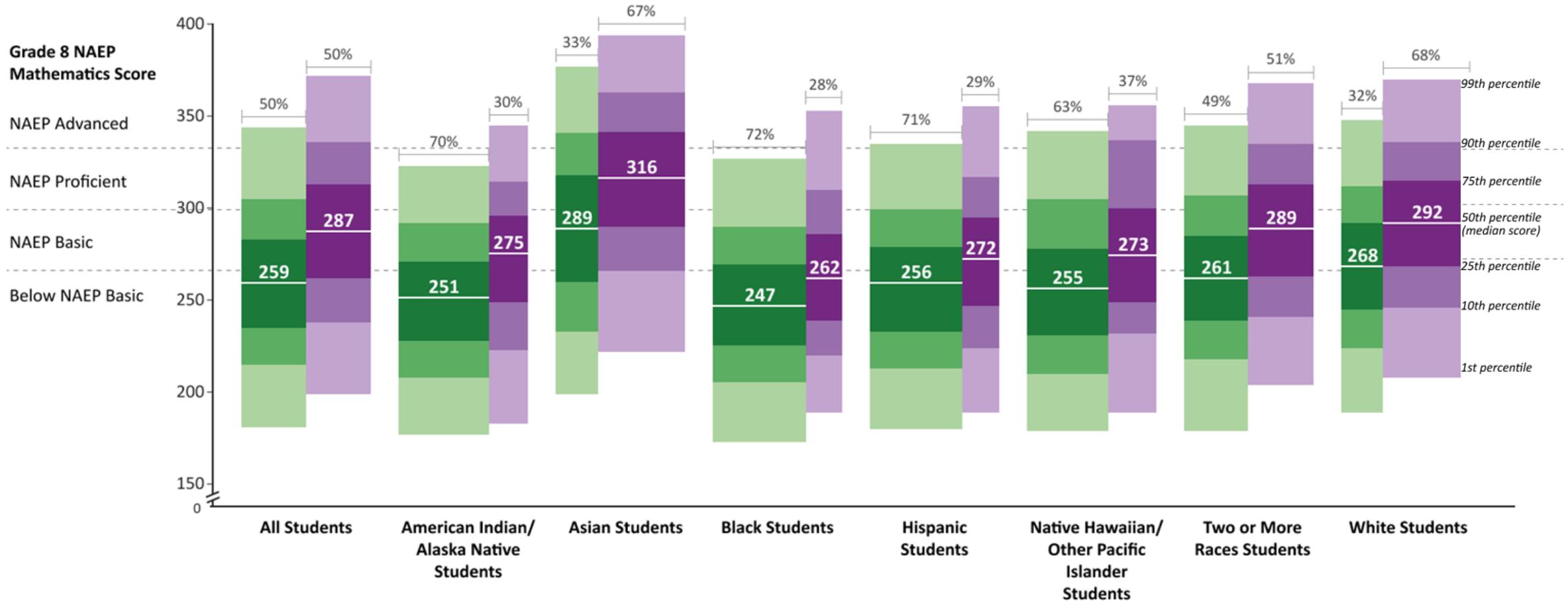


- Proportionally, more Black students are in schools lacking materials and certified teachers and in schools that need repair.
- US schools have not produced equitable outcomes in 8th grade math for Black students, compared to the nation.
- Resource disparities are a contributing factor to these inequitable outcomes.



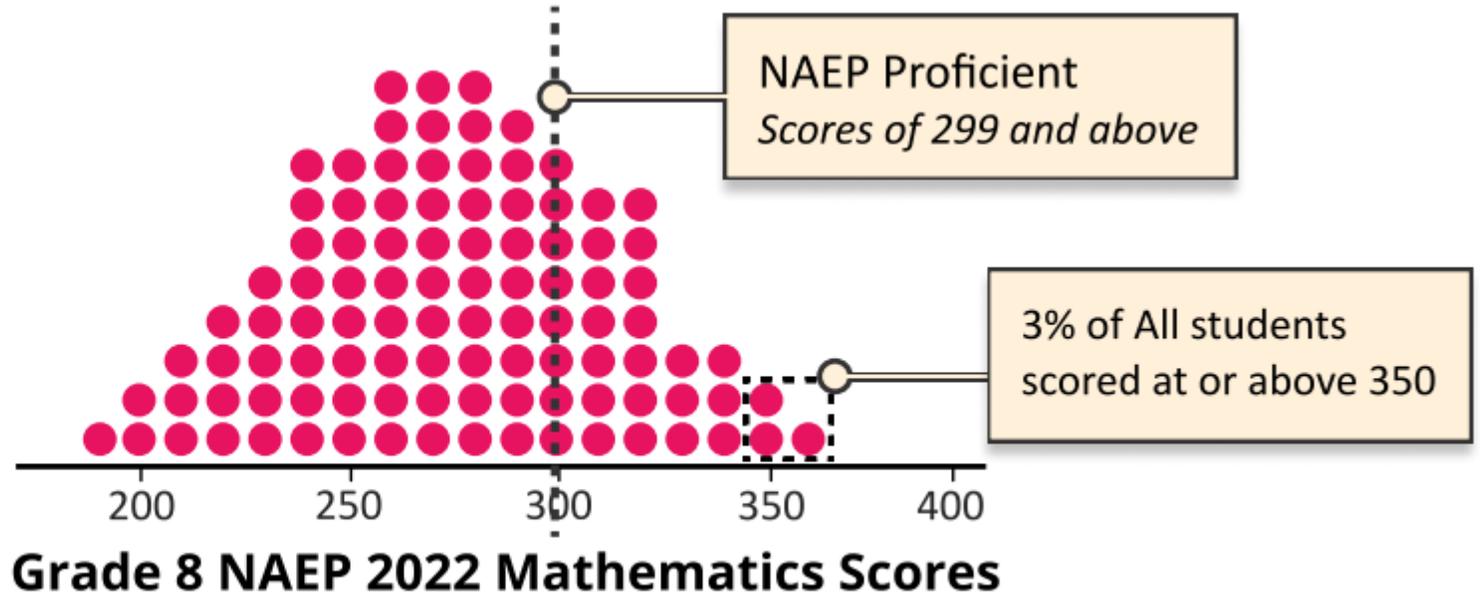
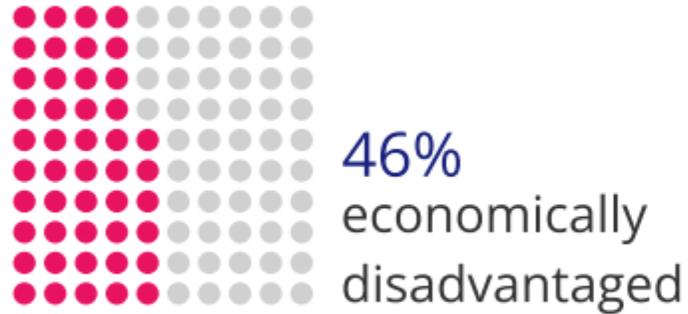
Within-group differences by SES persist across all racial/ethnic groups

FRPL eligible
Not eligible



Relating economic disadvantage to performance distributions

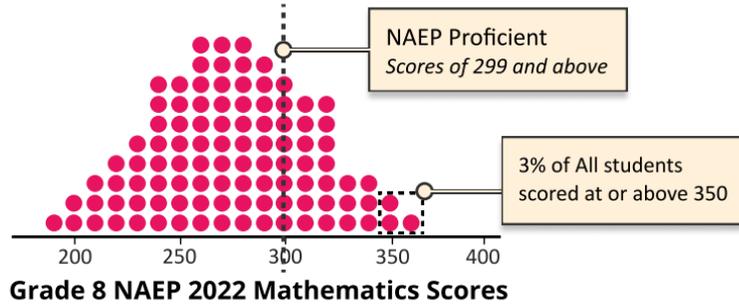
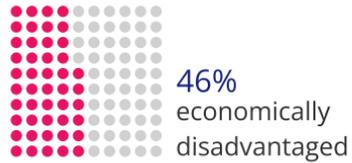
All Students



Key: ● = 1% of students

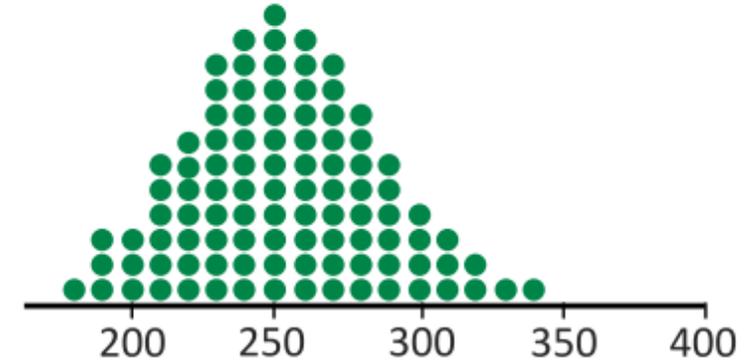
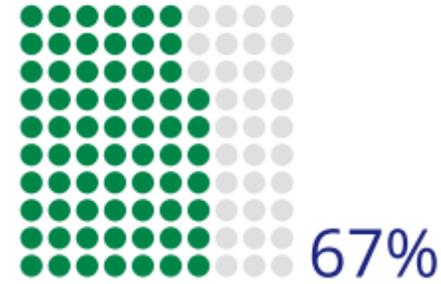
Relating economic disadvantage to performance distributions

All Students

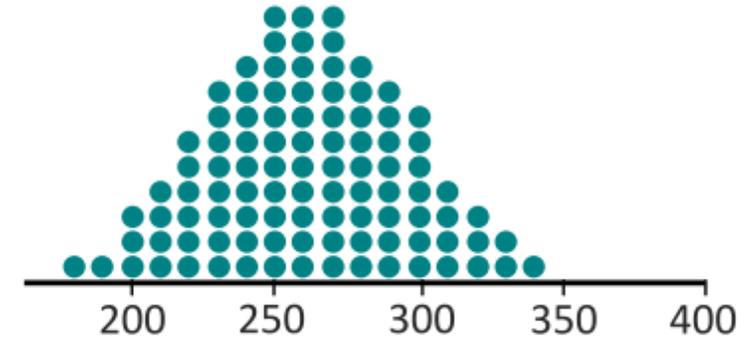
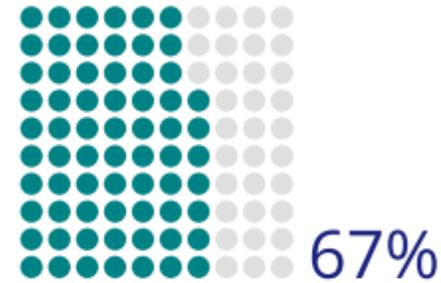


Key: ● = 1% of students

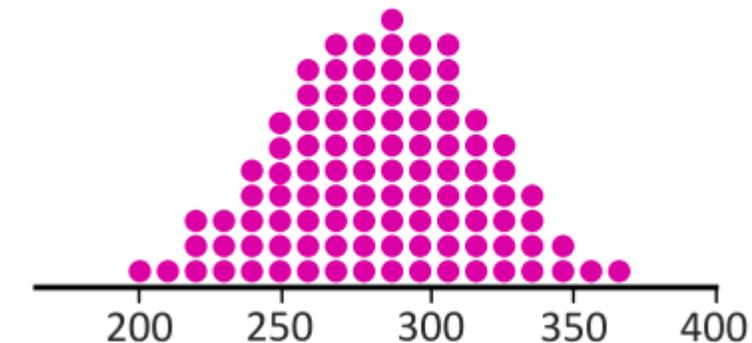
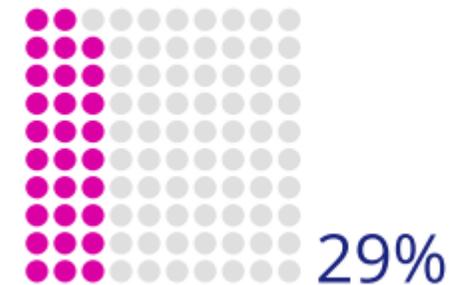
Black Students



Hispanic Students



White Students



Closing remarks



“NAEP is so important because it's the one thing that everyone, all states, take that is the same.”

Recommendations:

1. Improve or remove gap displays



Recommendations:

1. Improve or remove gap displays
2. Add other subgrouping variables





“One of the biggest reasons that we have [NAEP and state] assessments is to be able to understand where there are differences in access to opportunity for high quality learning experiences.”



“There’s a reason why low achievers are hindered. [We] can’t think that doing what we do with high achievers will bring everyone up, and people might find it condescending.”

Recommendations:

1. Improve or remove gap displays
2. Add other subgrouping variables
3. Expand contextual information





“Too often [opportunity to learn] is simplified as poverty, minority, or zip code. Sometimes it... ignores home life when it should, in fact, be much broader than just school opportunities.

Recommendations:

1. Improve or remove gap displays
2. Add other subgrouping variables
3. Expand contextual information
4. Don't oversimplify complex data



Thank you

ekerzabi@ets.org

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). Standards for educational and psychological testing. American Educational Research Association. https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Brown, G. T. L., Kannan, P., Sinharay, S., Zapata-Rivera, D., & Zenisky, A. L. (2023). Challenges and opportunities in score reporting: A panel of personal perspectives. *Frontiers in Education*. <https://doi.org/10.3389/educ.2023.1211580>
- Connecticut State Department of Education (2024, May). Using Accountability Results to Guide Improvement (7th ed.). https://edsight.ct.gov/relatedreports/using_accountability_results_to_guide_improvement.pdf
- Davis, L. P., & Museus, S. D. (2019). What is deficit thinking? An analysis of conceptualizations of deficit thinking and implications for scholarly research. *NCID Currents*, 1(1). <http://dx.doi.org/10.3998/currents.17387731.0001.110>
- Holder, E. (2022). Unfair comparisons: How visualizing social inequality can make it worse. *Nightingale Journal of the Data Visualization Society*. <https://nightingaledvs.com/unfair-comparisons-how-visualizing-social-inequality-can-make-it-worse/>
- Holder, E. & Xiong, C. (2022). Dispersion vs disparity: Hiding variability can encourage stereotyping when visualizing social outcomes. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 624-634. <https://doi.org/10.1109/TVCG.2022.3209377>
- Hughes, G. B. (2023, June). *Improving Equitable Measurement and Reporting in NAEP [Report]*. NAEP Validity Studies (NVS) Panel. American Institutes for Research (AIR). <https://www.air.org/sites/default/files/2023-09/EquitMeasReportNVS-2023-508.pdf>
- Ladson-Billings, G. (2006). From the achievement gap to the education debt: Understanding achievement in US schools. *Educational Researcher*, 35(7), 3-12. <https://www.jstor.org/stable/3876731>
- Merite, G. & Holder (2024, March). Sketching Sketchy Bar Charts. *LinkedIn*. <https://www.linkedin.com/pulse/sketching-sketchy-bar-charts-effect-affect-hxyje>
- Quinn, D. M., Desruisseaux, T., & Nkansah-Armankra, A. (2019). “Achievement Gap” language affects teachers’ issue prioritization. *Educational Researcher*, 48(7), 484-487. <https://doi.org/10.3102/0013189X19863765>
- Quinn, D. M., & Desruisseaux, T.-M. (2022). Replicating and Extending Effects of “Achievement Gap” Discourse. *Educational Researcher*, 51(7), 496-499. <https://doi.org/10.3102/0013189X221118054>
- Shukla, S., Theobald, E. J., Abraham, J. K., & Price, R. M. (2022). Reframing Educational Outcomes: Moving beyond achievement gaps. *CBE – Life Sciences Education*. <https://doi.org/10.1187/cbe.21-05-0130>
- U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) (n.d.). *Achievement Gaps Dashboard*. https://www.nationsreportcard.gov/dashboards/achievement_gaps.aspx



Study Lead
Marisol Kevelson
Mkevelson@ets.org

Opportunity to Improve Educational Assessment Reports

The ETS Innovations Research Programs Policy & Practice group is seeking input from **state education officials, policymakers and their advisors, researchers and others who use national and state assessment reports and data**. Your recommendations will help shape improvements in how educational test data is reported to positively affect student outcomes.

- **Online survey** launches December 2024
- **\$50 incentive payment**
- **Add your name** to the interest list at bit.ly/3VkmsFK

