IES Institute of Education Sciences

STUDY SNAPSHOT
February 2023

Regional Educational Laboratory Mid-Atlantic
At Mathematica

Stabilizing Subgroup Proficiency Results to Improve the Identification of Low-Performing Schools

The Every Student Succeeds Act requires states to designate schools with low-performing student subgroups for Targeted Support and Improvement (TSI) or Additional Targeted Support and Improvement (ATSI). Identifying the schools that most need support hinges on accountability data that reliably measure school performance. However, random differences between students' true abilities and their test scores—called measurement error—can obscure a school's true performance. This is especially likely in small schools or student subgroups, where random factors that affect a small number of students can have an outsized impact on the school's or subgroup's average score. Measurement error reduces the statistical reliability of the performance measures used to identify schools for these designations, introducing a risk that the identified schools are unlucky rather than truly low performing. Enhancing the reliability of school performance measures will advance state and local education agencies toward their goal of providing support to the schools and students that need it most.

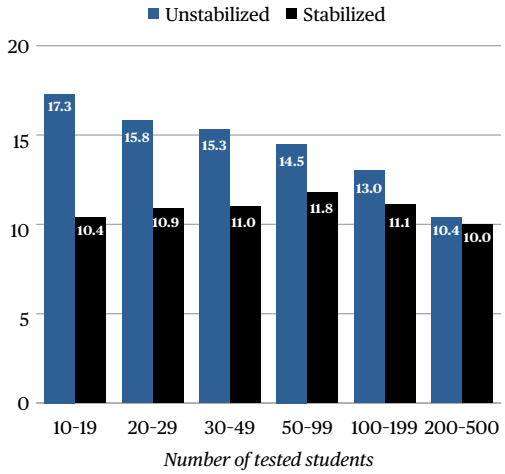
This study used Bayesian stabilization to improve the reliability (long-term stability) of subgroup proficiency measures that the Pennsylvania Department of Education (PDE) uses to identify schools for TSI and ATSI. The study team applied two statistical models to subgroup-specific proficiency rates, one aligned with PDE's accountability rules for ATSI and the other with rules for TSI. The results of the stabilization models were then compared with the unstabilized proficiency rates currently used in accountability calculations to assess whether stabilization increased statistical reliability.

Key findings

- **Stabilization improved statistical reliability, especially for small subgroups.** In unstabilized data, the variation in proficiency rates and the number of students in the subgroup were strongly correlated: proficiency rates were more variable and more likely to be extreme due to measurement error in smaller groups (see figure). Stabilized proficiency rates, in contrast, showed similar variation across subgroup sizes, indicating that they are more reliable because they reflect less measurement error—less random good or bad luck.
- **Stabilization could allow for inclusion of subgroup sizes under 20 in subgroup proficiency measures by improving the reliability of proficiency rates.** In unstabilized data, variation in proficiency rates is markedly higher for subgroups of 10-19 students than for subgroups of 20 or more students. With stabilization, proficiency rates for subgroups with 10-19 students vary less than unstabilized proficiency rates for subgroups with 20-29 students, suggesting that stabilized data meet the reliability requirements currently in use.

Stabilization decreases the prominence of measurement error in subgroup proficiency rates

Median standard deviation of proficiency rates



Note: Values were calculated using two-year averages of academic proficiency rates for the 2016/17 and 2017/18 academic years.
Source: Pennsylvania Department of Education data.

For the full report with technical details, see https://ies.ed.gov/ncee/rel/Products/Publication/106926.