COMPARABILITY

Comparability Without the Harms of Standardization

Education Accountability Remade paper series

WHAT DOES THE CURRENT SYSTEM ASSUME ABOUT COMPARABILITY?

That scores only mean the same thing if the test is identical everywhere: the same items, time limits, conditions, and scoring. The paper argues comparability can come from the knowledge or skill being assessed, not from identical tests.

THE KEY POINTS IN BRIEF

- **A big investment for limited return.** Since No Child Left Behind, states have pursued data comparability through identical tests, a multi-billion dollar build-out that still costs up to $40M per state to administer. Twenty years on, the inequities the scores highlight remain.
- **Real harms.** Curriculum narrows toward tested subjects, test prep crowds out learning, and students report anxiety or disengagement. Campbell's Law holds that high-stakes indicators get distorted, and the approach clashes with how students learn: through engaged, contextual work and revision.
- **Construct comparability is the alternative.** Different tasks can be compared if they measure the same construct against common definitions of mastery. That takes clear rubrics at each level of performance and a validity argument backed by calibrated, shared scoring.
- **It already works at scale.** The New York Performance Standards Consortium has run this way for 25+ years across 12,000+ students. Networks in New Hampshire, Kentucky, and Massachusetts show that moderation raises scoring reliability and educators' assessment literacy.
- **New system elements are needed.** Regional calibration hubs, open performance task banks, and peer-review networks can be built over time, keeping rigor and equity visible while giving local communities room to define authentic work.

QUESTIONS TO ASK YOURSELF

If you're a parent or caregiver

- When I see my child's score, what do I really know about what they can do?
- What skills that my child needs to thrive in the world aren't measured by the state tests?
- How does my school make sure expectations are the same for every student, even when assignments differ?

If you're a student

- Does one test show how I think, or would a project show it better?
- Do I get the chance to revise my work and improve it?
- If my project differs from a student's in another school, how would we know the bar is the same?
- How would it feel to be able to make choices about how I demonstrate what I've learned?

If you're an educator

- Do I know what proficient looks like in my subject, in a rubric I helped shape?
- Do I get time to score student work with colleagues and compare our judgments?
- Could I show how my tasks measure the standard if someone asked?
- Would focusing on projects and performance assessments instead of tests make learning more interesting for me and my students?

If you're a policy maker

- What does our testing infrastructure cost, and who benefits from it?
- Could we define common constructs and mastery levels so local tasks differ but expectations match?
- What support and oversight would keep every student group visible?