

**Embargoed: Not for release or discussion before  
10:00 am (EST)  
Wednesday, January 17, 2018**

# HOW HIGH THE BAR?

How would other nations perform  
if their students were judged by  
Common Core or NAEP benchmarks?

EXECUTIVE SUMMARY • JANUARY 2018

A REPORT FROM THE NATIONAL SUPERINTENDENTS ROUNDTABLE AND THE HORACE MANN LEAGUE

## KEY FINDINGS

---

- **The pursuit of excellence requires rigorous standards backed by demanding assessments.** This report does not endorse an anti-testing agenda or seek to lower standards. The great value of credible, large-scale assessments is that they provide a window into the world of schools and solid estimates of student performance.
- **National judgments about student proficiency and many state Common Core judgments about “career and college readiness” are defective and misleading.** These judgments are based on benchmarks for NAEP and several of the tests associated with the Common Core. A form of assessment imperialism has come to define what a school is supposed to be.
- **NAEP misuses the term “Proficient.”** According to NAEP officials, Proficient does not mean grade level performance. The misuse of the term confuses the public. The effects of this misuse are reflected in most Common Core assessments.
- **NAEP’s term “Proficient” does not even mean proficient.** “Students who may be proficient in a subject, given the common usage of the term, might not satisfy the requirements for performance at the NAEP achievement level,” wrote NAEP officials.
- **Many independent analysts reject the standard-setting process used to develop the NAEP benchmarks.** From 1993 through 2016, analysts from such agencies as the U.S. General Accounting Office, the National Academy of Sciences, and the Brookings Institution sharply criticized the standard-setting process and questioned the results.
- **One motivation for establishing the NAEP benchmarks was the desire to demonstrate that “large numbers of students were failing,”** according to a former *New York Times* national education correspondent. A rushed process for developing the benchmarks was adopted by the policy body governing NAEP – despite experts’ objections – in part because a prominent member of the policy body acknowledged he was “fed up with technical experts.”
- **Advocates who push for school improvement on the grounds of questionable benchmarks** are not strengthening education and advancing American interests, but undermining public schools and weakening the United States.
- **The NAEP bar for Proficient and Common Core bars that track NAEP’s benchmark have been set so erroneously high they conceal student performance.** They also (a) defeat assessment’s purpose of providing valuable insights into student performance; and (b) establish performance bars that the vast majority of students in countries all over the world cannot clear.
- **The challenge of clearing these performance bars is global.** In no nation do a majority of students clear the proficiency/ career and college readiness bars in Grade 4 reading. In just three nations (Japan, Singapore, and South Korea) do a majority of students clear this bar in Grade 8 mathematics, while only one (Singapore) does so in Grade 8 science.

---

# In recent years, communities all over the United States have been faced with bleak headlines about the performance of their students and their schools on state Common Core and NAEP (National Assessment of Educational Progress) results.

A particular concern is that only roughly one-third of students meet key Common Core and NAEP benchmarks of “Proficient” or “career and college readiness.”

The reality is that communities all over the world would face identical bleak headlines if their students sat down to take the NAEP or Common Core assessments. When citizens read that “only one-third” or “less than half” of the students in their local schools are proficient in mathematics, science, or reading, they can rest assured that the same judgments can be applied to students throughout most of the world.

Why would anyone expect fourth-grade students to be “Proficient” when asked to interpret reading passages experts agree are appropriate for Grade 7? Is it reasonable to believe that 30 percent of 12th-grade students who completed Calculus are not Proficient in mathematics? Meanwhile, 69 percent of pre-Calculus students and 92 percent of those who completed trigonometry and Algebra

II are deemed to be failures by these faulty benchmarks.

The fault lies not in the students. Not in the schools. Not in the Common Core. Nor even in the assessments themselves. It lies in the flawed benchmarks that define acceptable performance on these assessments.

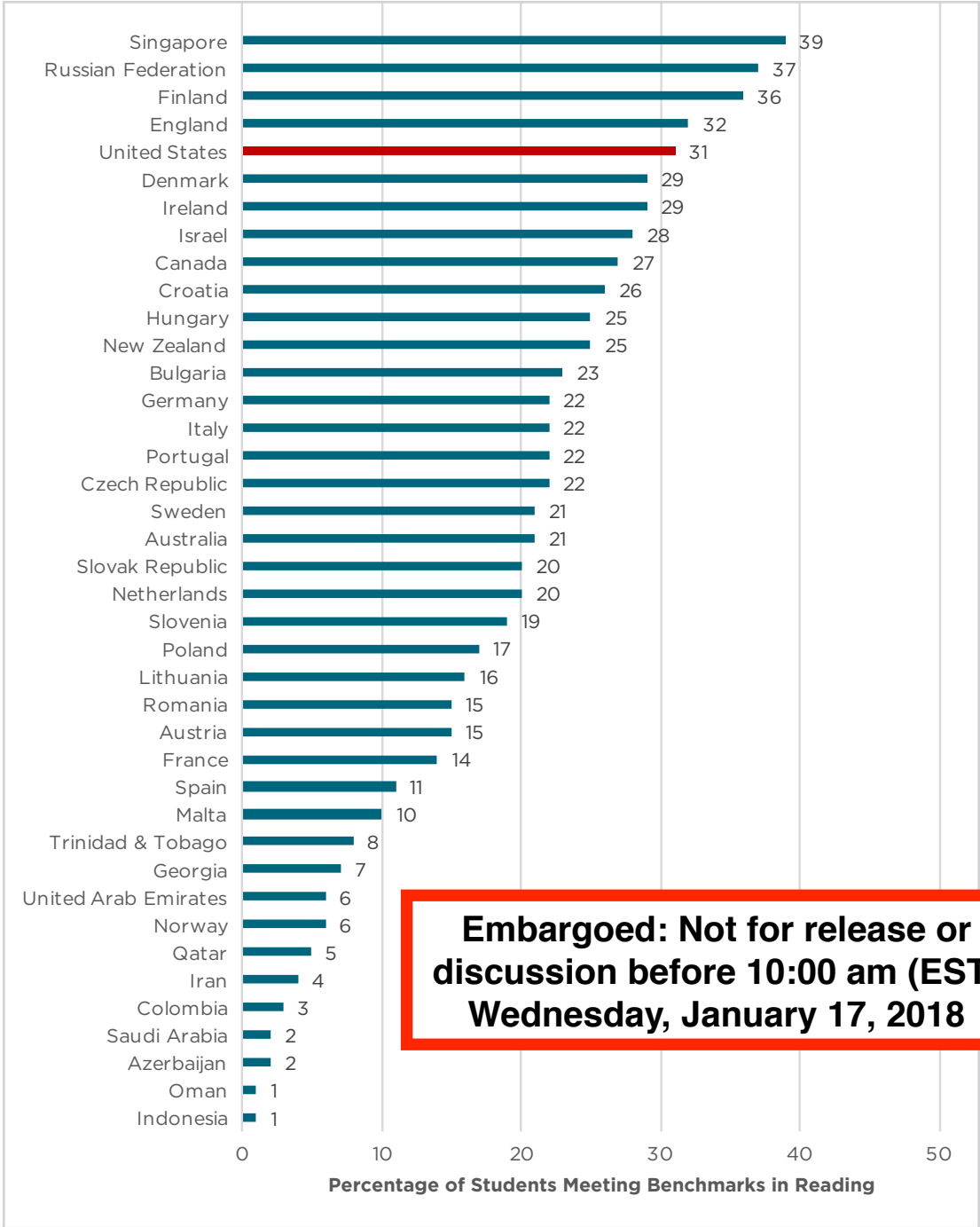
Globally, in just about every nation where it is possible to compare student performance with our national benchmarks, the vast majority of students cannot demonstrate their competence because the bars are set unreasonably high.

The conclusion in the prior paragraph is based on statistical analyses of existing international assessments, including 2011 results from a fourth-grade assessment of reading (PIRLS – Progress in International Literacy Study). The release of the 2016 PIRLS results as this report went to press in no way alters this conclusion.

**TABLE 1: Nations in which a Majority of Students Can be Expected to Clear the Common Core/NAEP Bars of Proficiency or Career and College Readiness, by Grade and Subject**

Grade and Subject	Number of Nations in which Majority of Students Clear the Bar
Grade 4 Reading	0
Grade 8 Mathematics	3
Grade 8 Science	1

**FIGURE 1: Percentage of Grade 4 Students by Nation Who Can Be Expected to Clear the NAEP Proficient Benchmark in Reading**

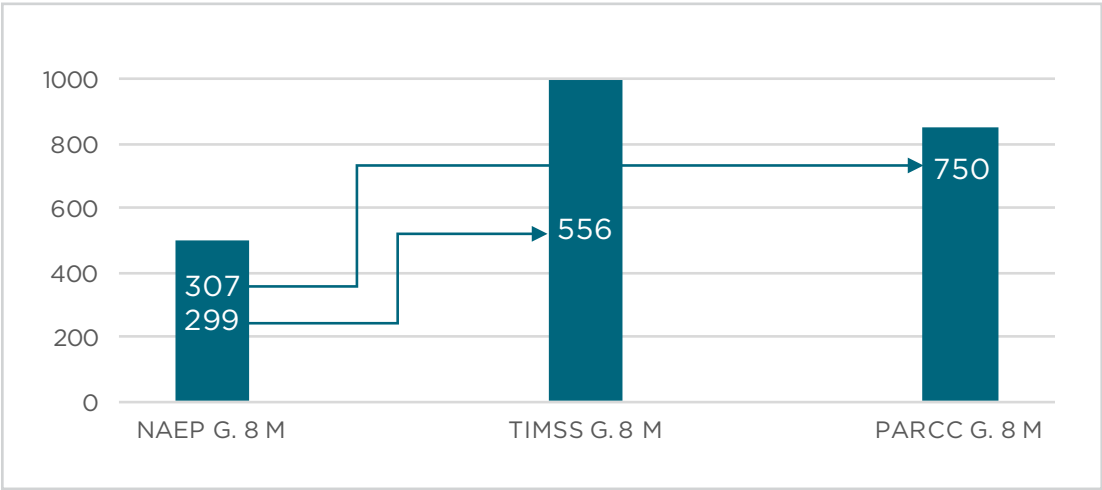


# MAPPING ONE TEST’S BENCHMARK ONTO ANOTHER

How do you understand what a score on one test means on another test with a different scale? For example, a score of 24 on an ACT test compared to a score of 520 on the SAT? Or a score of 299 on NAEP with a score of 620 on an international assessment? A number of procedures with different strengths and weaknesses can be employed. This study draws on two. The first, *statistical moderation*, uses complex statistical tools to map one test onto another. The second, *equipercentile ranking*, isolates the percentile at which NAEP’s

Proficient benchmark is determined, by grade and subject, and defines an equivalent score on companion assessments by isolating the same percentile. Both approaches are widely used by the U.S. Department of Education and its contractors to link NAEP scores to state assessments and to international assessments in reading, mathematics, and science. Figure 2 displays how a particular score on NAEP’s Grade 8 Mathematics assessment translates into scores on comparable assessments from TIMSS and one of the Common Core consortia.

**FIGURE 2: Grade 8 Math: Mapping NAEP Scores onto TIMSS & PARCC Scales**



**Read as:** A score of 299 on the NAEP 8th grade math scale converts into 556 on the comparable TIMSS scale. A NAEP score of 307 translates into a PARCC score of 750.

# THE COMMON CORE CONNECTION

When the principal Common Core assessments were developed by the Partnership for Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortium (SBAC), developers faced pressure to align their “career and college readiness” benchmarks with NAEP’s standard of Proficient. The pressure succeeded. The scores defining college and career readiness in many these assessments either match the NAEP proficiency benchmark, or are very close to it (see Table 2).

It can be expected that when the “career and college readiness” benchmarks of the Common Core assessments align with or approach the NAEP Proficient benchmark, most students around the world will also be found wanting. It is highly likely that states that abandoned PARCC or SBAC to develop their own assessments under the Every Student Succeeds Act (ESSA) will find similar challenges if their new benchmarks are aligned with those of NAEP, PARCC, or SBAC.

**TABLE 2: Relationship of Common Core “Career and College Ready” Benchmarks to NAEP Proficient Benchmark**

Grade and Subject	Assessment	NAEP Equivalent of “Career and College Ready”
Grade 4 English/Language Arts	PARCC	Approaches Proficient
	SBAC	Basic
	Florida	Proficient
	New York	Proficient
Grade 4 Mathematics	PARCC	Approaches Proficient
	SBAC	Basic
	Florida	Proficient
	New York	Proficient
Grade 8 Mathematics	PARCC	Proficient
	SBAC	Approaches Proficient
	Florida	Proficient
	New York	Proficient
Grade 8 English/Language Arts	Florida	Proficient
	New York	Proficient

## RECOMMENDATIONS

---

- I. **Redefine NAEP's basic terminology.** We recommend that the National Assessment Governing Board rename the NAEP benchmarks as Low, Intermediate, High, and Advanced.
- II. **Emphasize caution in interpreting these benchmarks.** We recommend that the U.S. Department of Education emphasize in every NAEP publication that the U.S. Congress insisted that NAEP benchmarks be understood as acceptable only on a "trial basis," and that results based on the benchmarks be interpreted "with caution."
- III. **Educate the public about the assessment findings outlined in this report.** We recommend that local school leaders – state chiefs, superintendents, board members, and teachers – vigilantly educate their local communities about the flaws embedded in the term Proficient and how school systems abroad would perform if held to that standard.
- IV. **Revisit the decision to tie state assessments' "College Readiness" standards to NAEP's Proficient benchmark.** We recommend extreme caution before acting on the assumption that state agencies or psychometricians understand who is "college ready" and who is not, especially in determining whether students in Grades 4 and 8 are "on track" to be "college ready."
- V. **Develop a national K-12 capacity to better analyze assessments.** We recommend that the major national organizations representing a variety of K-12 constituencies develop significant capacity to analyze and comment on developments in national and international assessments.

## PUBLIC EDUCATION: A LARGER PURPOSE

Educators have an obligation to produce graduates who are competent in reading, writing, and mathematics – and prepared to earn a living. Assessment and accountability are critical to that effort. But education is about more than earning a living. It's about living a life. Children are not standardized test scores. That is why the surface appeals of "Proficient" or "career and college ready" fail to capture the nature of the public good that is the public school. Education is this nation's greatest asset in pursuit of the American Dream. The values embedded in that dream are the real standards around which educators, citizens, and the assessment community should rally.

## ACKNOWLEDGMENTS

---

This summary of a much longer and more detailed report is a product of a two-year study completed by James Harvey (National Superintendents Roundtable) and Emre Gönülates (Michigan State University and Teachers College, Columbia University). We thank them for developing the research and framing the argument that is contained in the complete report, available on the Roundtable website at [www.superintendentsforum.org](http://www.superintendentsforum.org). We also thank the Roundtable for supporting Dr. Harvey's time to complete this project and for covering the expenses associated with Dr. Gönülates' invaluable contributions. All of the data documenting the findings in this executive summary can be found in the full report.

We thank the members of the Roundtable Steering Committee and the board of the Horace Mann League who endorsed this work from the outset.

Several individuals reviewed an earlier draft of the complete report and provided helpful guidance. We are grateful to the following for taking the time to give us the benefit of their views, many of which we incorporated:

- David Berliner, Regents Professor of Education Emeritus, Arizona State University, former president of the American Educational Research Association
- Eva Baker, Founding Director of the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), University of California, Los Angeles
- Henry Braun, Boisi Professor of Education, Boston College, and Director of the Center for the Study of Testing, Evaluation, and Education Policy
- Charles Fowler, President, School Leadership, LLC, New Hampshire
- Tom Loveless, Brown Center, Brookings Institution
- David Rutkowski, Center for Educational Measurement, University of Oslo.

We want to acknowledge the early support and encouragement of this research by John Chatten-McNichols and John Jacob Gardiner of Seattle University, along with the late dean of Seattle University's College of Education, Sue Schmitt. William Schmidt and Richard Houang of Michigan State University were also sources of inspiration.

Finally, we thank Rhenda Meiser, of Meiser Communications, Kai Hiatt of the National Superintendents Roundtable, and Anne Paxton from ProForum for their careful reading of the manuscript. Kathy Mathes of Mathes Design designed the report. Ms. Meiser counseled on the release and public announcement of the report. We are deeply in their debt.



The Horace Mann League

---

FOR MORE INFORMATION:

National Superintendents Roundtable

9425 35th Avenue NE, Suite E | Seattle, WA 98115 | 206-526-5336

[www.superintendentsforum.org](http://www.superintendentsforum.org) | [jamesharvey@superintendentsforum.org](mailto:jamesharvey@superintendentsforum.org)