

September 18, 2018

NAEP Achievement-Level-Setting Program
National Assessment Governing Board
800 North Capitol Street, NW
Suite 825
Washington, D.C. 20002

Response to request for comments on "Draft Policy Statement on Developing Student Achievement Levels for the National Assessment of Educational Progress," posted Monday, September 10, 2018 by the National Assessment Governing Board, U.S. Department of Education in Notices of the Federal Register, Vol. 83, No. 175

These comments are authorized by the Steering Committee of the National Superintendents Roundtable, representing 75 school superintendents responsible for the education of some five million K-12 students, and 25 former superintendents.

Summary

The Roundtable appreciates the opportunity to comment on the draft policy statement. Although initially inclined to commend the National Assessment Governing Board (NAGB) for this effort to respond to the 2016 report of the National Academy of Sciences,¹ the Roundtable concludes that the proposed statement is unlikely to attain the goal defined in its first paragraph, namely to produce achievement levels for the National Assessment of Educational Progress (NAEP) that are "reasonable, useful, and informative to the public."

The Roundtable supports high standards. While it believes NAEP misuses the term "Proficient," it does not ask for a lowering of standards but for a clearer definition of what they mean in terms parents and the public can understand. In pursuit of that objective, we make the following observations: The original achievement levels were developed in a rushed process. Those levels have produced results that have confused educators, citizens, and policymakers. Instead of being reasonable, the benchmarks represent "wishful thinking" and defy "reason" and "common sense," according to knowledgeable experts. The latest research linking NAEP's benchmarks to international assessments reveals that the majority of students in most nations cannot clear NAEP's proficiency bar. Finally, the proposed modifications to the policy definitions for NAEP achievement levels are unresponsive to the history of criticisms of the achievement levels and to the latest research.

We conclude with suggestions for rewriting the policy definitions and an appeal to extend the comment period from the currently contemplated 20 days to six months.

¹ Christopher Edley, Jr. & Judith A. Koenig (eds.). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Washington: National Academies Press, 2017.

Rushed Initial process

Public Law 107-279, the *Augustus F. Hawkins-Robert T. Stafford Elementary and Secondary School Improvement Amendments of 1988*, created the National Assessment Governing Board and charged it with identifying appropriate achievement goals, which were established as “Basic,” “Proficient,” and “Advanced.”¹ An advisory panel appointed by NAGB in June 1990 to agree on the levels, and the proportion of students at each level who could be expected to answer questions correctly, reached agreement on these issues in November,² a remarkably brief period, given the importance and complexity of the issues involved. NAGB rejected the advice of experts to slow down the benchmarking process.³ NAGB board member Chester E. Finn, Jr. later dismissed the value of technical expertise: “I get fed up with technical experts [who] . . . take an adversarial stance toward some of the things that are most important the views of those operating NAEP, such as setting standards.”⁴

The results have confused educators, the public, and policymakers

Understandably the public, many commentators, and policymakers are confused by these benchmarks (achievement goals).

It is clear, as NCES states emphatically that “Proficient is not synonymous with grade-level performance.” According to Dr. Peggy Carr of NCES if the public wants to understand how many US students are performing on grade level it is more accurate to look at Basic scores, than Proficient results.⁵

Yet, on 2016, former television commentator Campbell Brown, newly leading an education advocacy organization called *The 74* released a video on *Slate* arguing that “Two out of three eighth graders in this country cannot read or do math at grade level.” Asked for evidence of this claim she cited NAEP, apparently confusing the term Proficient with being at grade level.⁶

Major organizations such as Achieve Inc. and Students First misuse the term consistently in their publications, while Achieve and NCES issue analyses lamenting the fact that state definitions of proficiency fall far short of the NAEP Proficient benchmark.⁷

¹ Jack Jennings, *A Paper on the NAEP Long-Term Trend Assessments Prepared for the National Assessment Governing Board*, February 13, 2017.

² M.A. Vinovskis, *Overseeing the Nation’s Report Card: The Creation and Evolution of the National Assessment Governing Board*. Washington: NAGB, 1998.

³ Vinovskis, M.A., *op. cit.*

⁴ Chester E. Finn, Jr., “An Interview with Chester E. Finn, Jr.,” in Lyle V. Jones and Ingram Olkins (eds.), *The Nation’s Report Card: Evolution and Perspectives*. Bloomington: Phi Delta Kappa Educational Foundation.

⁵ Story Hinkley, “National Testing: What Does it Mean for a Student to be ‘Proficient?’”, *Christian Science Monitor*, April 12, 2018.

⁶ Tom Loveless, “The NAEP Proficiency Myth”. *The Brown Center Chalkboard*, the Brookings Institution, June 13, 2016.

⁷ See for example: Achieve, Inc: *Proficient vs. Prepared 2018: Disparities between State Tests and the 2017 National Assessment of Educational Progress*. Washington: Author, May 2018; also Bendeira de Mello, Rahman, T., and Park, B.J., *Mapping State Proficiency Standards onto NAEP Scales: Results from the 2015 NAEP Reading and Mathematics Assessments*. Washington, National Center for Education Statistics, 2018.

Earlier this month, *Education Week* released its latest “Quality Counts” report.⁸ In this effort to “grade the states” on school quality, the report included state-by-state data on poverty gaps, high school graduation rates, and Advanced Placement experience. But the foundation of its achievement index is the proportion of students meeting the NAEP proficiency definitions in fourth- and eighth-grade reading and mathematics.

Oddly, the term Proficient as used by NAGB does not even mean proficient. As former NAGB officials wrote in 2001:⁹

Nor is performance at the Proficient level synonymous with ‘proficiency’ in the subject. That is, students who may be considered proficient in the subject, given the common usage of the term, might not satisfy the requirements for performance at the NAEP achievement level.

In short, although most citizens are understandably interested in performance at grade level, many confuse performance at grade level with NAEP’s proficiency benchmark. Understandably, journalists and many members of the public ignore the benchmark of Basic, interpreting it as an indication that large numbers of students are barely scraping by. This interpretation does great damage to public perceptions of public schools and perceptions of the ability and skills of the next generation of young Americans. To add to the confusion, officials associated with NAGB have openly acknowledged that NAEP’s definition of proficiency is unrelated to common usage of the term.

Benchmarks represent “wishful thinking”

The rushed process produced a scientific debate that lasts to this day.

NAGB hired a team of evaluators in 1990 to study the process involved in developing the three levels. A year later the evaluators were fired after their draft report concluded that the process “must be viewed as insufficiently tested and validated, politically dominated, and of questionable credibility.”¹⁰

In 1993, the U.S. General Accounting Office labeled the standard-setting process as “procedurally flawed” producing results of “doubtful accuracy.”¹¹

In 1999, the National Academy of Sciences reported the achievement-level setting procedures were flawed: “difficult and confusing . . . internally inconsistent . . . validity evidence for the cut scores is lacking . . . and the process has produced unreasonable results.”¹²

Shortly after *No Child Left Behind* was signed into law in 2001, Robert Linn, past president of the American Educational Association and of the National Council on Measurement in

⁸ “Quality Counts 2018: K-12 Assessment & Chance for Success. Grading the States.”, *Education Week*, September 5, 2018.

⁹ Loomis, S.C. and Bourque, M.L. (eds.). *National Assessment of Educational Progress Achievement Levels, 1992-1998 for Reading*. Washington: National Assessment Governing Board.

¹⁰ Jack Jennings, *op. cit.*

¹¹ U.S. General Accounting Office. *Educational Achievement Standards: NAGB’s Approach Yields Misleading Interpretations*. GAO/PEMD-993-12. Washington, DC: Author. (ERIC database: ED359268)

¹² Pellegrino, J.W., Joners, L.R., and Mitchell, K.J. *Grading the Nation’s Report Card*. Washington, DC: National Academy Press.

Education, and former editor of the *Journal of Educational Measurement* called the “target of 100% proficient or above according to the NAEP standards is more like wishful thinking than a realistic possibility.”¹³

In 2007, researchers concluded that fully a third of high school seniors who completed calculus, the best students with the best teachers in the country, could not clear the proficiency bar. Moreover, they added, fully 50 percent of those who scored “basic” in twelfth grade math had achieved a bachelor’s degree¹⁴ (a proportion comparing favorably with four-year degree rates at public universities).

The Buros Institute, named after the father of *Mental Measurements Yearbook*, criticized the lack of a validity framework for NAEP assessment scores in 2009 and recommending continuing “to explore achievement level methodologies.”¹⁵

Fully 30 percent of 12th-graders who completed calculus were deemed to be less than proficient, said a Brookings Institution scholar in 2016, a figure that jumped to 69 percent for pre-calculus students and 92 percent for students who completed trigonometry and Algebra I. These data “defy reason” and “refute common sense,” he concluded.¹⁶

Finally, the NAS study to which the proposed rule responds took note in 2016 of the “controversy and disagreement around the achievement levels, noting that Congress has insisted since 1994 that the achievement levels are to be used on a trial basis until on objective evaluation determined them to be “reasonable, reliable, valid, and informative to the public.”¹⁷

In the Roundtable’s judgment, such an objective evaluation has yet to be completed and a determination that the achievement levels are “reasonable, reliable, valid, and informative to the public” has yet to be seen.

Linking studies conclude most students in most nations cannot clear “proficiency” bar

The Roundtable points also to research studies dating from 2007 to 2018 indicating NAEP’s proficiency bar is beyond the reach of most students in most nations. When Gary Phillips of the American Institutes of Research (and former Acting Commissioner of NCES) asked how students in other nations would perform if their international assessment results were expressed in terms of NAEP achievement levels, his results were sobering.¹⁸ The results demonstrated that just three nations (Singapore, the Republic of Korea, and Japan) would have a majority of their

¹³ Cited in Loveless, 2016.

¹⁴ Scott, L. A., Ingels, S. J., and Owings, J. A. (2007). *Interpreting 12th-graders’ NAEP-scaled mathematics performance and postsecondary outcomes from the National Education Longitudinal Study of 1988 (NELS:88)*. U.S. Department of Education, National Center for Education Statistics, (NCES 2007-328).

¹⁵ Buckendahl, C.W. et al, (2009). *Evaluation of the National Assessment of Educational Progress*. Lincoln, NE: Buros Center for Testing, University of Nebraska.

¹⁶ Loveless, T., “The NAEP Proficiency Myth,” (2016). Brookings Institution: Brown Center Chalkboard, June 13, 2016.

¹⁷ Edley & Koenig (eds). Op. cit.

¹⁸ Phillips, G. W. (2007). *Expressing international education achievement in terms of U.S. performance standards: Linking NAEP achievement levels to TIMSS*. Washington, DC: American Institutes for Research.

students clear the NAEP bar in 8th-grade mathematics, while Singapore alone could meet that standard (more than 50% of students clearing the bar) in science.¹⁹

Subsequently Hambleton, Sireci, and Smith (2007) and also Lim and Sireci (2017) reached conclusions similar to those of Phillips.

An analysis conducted by Gonulates and Harvey in 2017²⁰ and reported by the Horace Mann League and the National Superintendents Roundtable in 2018²¹ applied Phillips' methodology to international fourth-grade reading results as defined by the Progress in International Reading and Literacy Survey (PIRLS). PIRLS is conducted by the International Association for the Evaluation of Educational Achievement (IEA), the organization that administers the TIMSS mathematics and science assessments. Not a single nation participating in PIRLS can demonstrate that a majority of its students clear the NAEP reading proficiency bar in Grade 4. Singapore leads the way, with 39 percent of its students clearing the bar, followed by the Russian Federation (37 percent), Finland (36 percent), England (32 percent), and the United States (31 percent).

In light of these findings, it defies logic and common sense, as Dr. Loveless noted, to act on the belief that NAEP's definition of proficiency is a reasonable and reliable guide to policy action.

Proposed new policy definitions define anything less than "NAEP Proficient" as deficient

Against that backdrop, it is surprising to see that NAGB proposes not to modify its policy definitions to clarify the term "Proficient" but to tinker at the margins with the definitions and add the word "NAEP" in front of each of the benchmarks. The justification offered in the Federal Register is to "better differentiate the NAEP achievement levels from other common uses of Basic, Proficient, and Advanced." But the effect is to define anything less than NAEP Proficient as deficient.

The table on the following compares the existing policy definitions of the terms Basic, Proficient, and Advanced, with the policy definitions attached to the new terms "NAEP Basic," "NAEP Proficient," and "NAEP Advanced." Like the existing policy definitions, the proposed definitions point inexorably to Proficient as the ideal to which states, the nation, and major demographic groups should, on average, aspire.

In the Roundtable's judgment these minor modifications are in no way responsive to the major criticisms that have been leveled at the NAEP benchmarks over the years. To retreat behind the claim that the Proficient benchmark is an aspirational standard is deceptive and evasive. In light of the results of the international linking studies completed between 2007 and 2018, far from being aspirational, the Proficient benchmark is delusional. In Dr. Linn's phrase, the proficiency benchmark represents wishful thinking.

¹⁹ It should be noted that no nation (or city or province) could demonstrate that its students met the NAEP definition of Advanced. (Cities and provinces, such as Chinese Taipei, Hong Kong are generally ignored for purposes of the analysis in this section.)

²⁰ See Appendix C of *How High the Bar?: "Applying NAEP Benchmarks to PIRLS results."*

²¹ The National Superintendents Roundtable and the Horace Mann League: *How High the Bar? How Would Other Nations Perform if Their Students Were Judged by Common Core or NAEP Benchmarks?* Seattle: Authors, 2018.

Benchmarks	Current Policy Definitions	Proposed Policy Definitions
Basic	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade.	This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.
Proficient	This level represents solid academic performance for each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.	This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
Advanced	This higher level signifies superior performance.	This level signifies superior performance beyond NAEP Proficient.

Repairing the damage inflicted on public schools and their students by NAEP's misleading terminology is all the more critical today as groups such as Achieve, Inc. openly advocate for NAEP's definition of Proficient as the benchmark to be employed in state assessments under the *Every Student Succeeds Act*.

Recommendations

Against the backdrop of the discussion above, the Roundtable makes two recommendations:

1. Revise the proposed policy definitions so that they are more informative to the general public (one of the principal goals of the NAS analysis) We propose the following definitions:

NAEP Basic	This level is roughly analogous to performance at grade level in each NAEP assessment. It denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.
NAEP Proficient	This level represents extremely demanding academic performance, a level beyond the abilities of most students in most nations. Students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.
NAEP Advanced	This level signifies superior performance beyond NAEP Proficient.

2. Extend the comment period on the draft policy statement for six months, during which time NAGB should hold between four and six hearings around the nation to solicit the advice of educators, parents, and other stakeholders. George Santayana's oft-quoted aphorism is usefully

born in mind here: Those who cannot remember the past are condemned to repeat it. The rushed process that produced the NAEP benchmarks in 1990 provides a cautionary tale for us today. NAEP has existed for nearly half a century. The benchmarks have been in use for over 25 years. Taking a few months to get it right on this occasion will be time well spent.

We commend NAGB for inviting comments and appreciate the opportunity to offer our insights. We stand ready to answer questions or help in any way with this process as it moves forward.

Sincerely,

A handwritten signature in black ink, appearing to read "James J. Harvey". The signature is fluid and cursive, with the first name "James" being more prominent and stylized.

James J. Harvey, Executive Director
(On behalf of Roundtable Steering Committee)

c.c. Peggy Carr, NCES

by email to NAEPALSpolicy@ed.gov

&

Peggy.Carr@ed.gov

And by overnight mail to:

NAEP Achievement-Level-Setting Program
National Assessment Governing Board
800 North Capitol Street, NW
Suite 825
Washington, D.C. 20002