

EDUCATION WEEK

Published Online: October 25, 2011

Published in Print: October 26, 2011, as **NAEP's Odd Definition of Proficiency**

NAEP's Odd Definition of Proficiency

By James Harvey

Released in August, the U.S. Department of Education [study](#) mapping state proficiency standards [Back to Story](#)

onto the National Assessment of Educational Progress scales produced a remarkable statement from Joanne Weiss, the chief of staff to U.S. Secretary of Education Arne Duncan. According to an [article](#) in the Aug. 24 issue of *Education Week*, Weiss said the practice of permitting each state to set its own proficiency standards amounts to “lying to parents, lying to children, lying to teachers and principals about the work they’re doing.” Her intemperate outburst crosses the line, not only by the standards of what passes for civil discourse in Washington these days, but also for what it says about the assessment itself.



Indeed, a plausible case can be made that when it comes to telling fibs about proficiency, NAEP has a nose that annually grows longer, for its definition of proficiency is seriously flawed.

As an assessment, let us be clear, NAEP is highly regarded. It is thought of as “the nation’s report card.” Yet a controversy has surrounded NAEP’s achievement levels of basic, proficient, and advanced since they were developed in the 1990s. Congress still insists that every NAEP report include [this disclaimer](#): “[The] National Center for Education Statistics has determined that NAEP achievement levels should continue to be used on a trial basis and should be interpreted with caution.”

Far from interpreting the NAEP achievement levels

“with caution,” Ms. Weiss threw caution to the winds in questioning those state educational leaders who followed the law and common sense in defining proficiency around performance at grade level.

Since definition is crucial in any discussion of standards, let’s define the terms of the discussion. The No Child Left Behind Act, passed by Congress in 2001 as the latest reauthorization of the Elementary and Secondary Education Act of 1965, permitted states to develop their own assessments and set their own proficiency standards to measure student



—iStockphoto/Olga Axyutina

achievement. Most states, for their purposes, quite sensibly defined proficiency as performance at grade level.

What about NAEP? Oddly, NAEP's proficient standard has little to do with grade-level performance or even proficiency as most people understand the term. NAEP officials like to think of the assessment standard as "aspirational." In 2001, long before the current contretemps around state assessments, two experts associated with the National Assessment Governing Board—Mary Lynne Bourque, staff member to the governing board, and Susan Loomis, a member of the board—made it clear that "the proficient achievement level does not refer to 'at grade' performance. Nor is performance at the proficient level synonymous with 'proficiency' in the subject. That is, students who may be considered proficient in a subject, given the common usage of the term, might not satisfy the requirements for performance at the NAEP achievement level."

It is hardly surprising, then, that most state assessments aimed at establishing proficiency as "at grade" produce results different from a NAEP standard in which proficiency does not refer to "at grade" performance or even describe students that most would think of as proficient. Far from supporting the NAEP proficient level as an appropriate benchmark for state assessments, many analysts endorse the NAEP basic level as the more appropriate standard because NAEP's current standard sets an unreasonably high bar.

What is striking in reviewing the history of NAEP is how easily its governing board has shrugged off criticisms about the board's standards-setting processes.

In 1993, the National Academy of Education argued that NAEP's achievement-setting processes were "fundamentally flawed" and "indefensible." That same year, the General Accounting Office concluded that "the standard-setting approach was procedurally flawed, and that the interpretations of the resulting NAEP scores were of doubtful validity." The National Assessment Governing Board, or NAGB, which oversees NAEP, was so incensed by an **unfavorable report**  it received from Western Michigan University in 1991 that it looked into firing the contractor before hiring other experts to take issue with the university researchers' conclusions that counseled against releasing NAEP scores without warning about NAEP's "conceptual and technical shortcomings."

In addition, NAGB absorbed savage criticism from the National Academy of Sciences, which concluded in 1999 that "NAEP's current achievement-level-setting procedures remain fundamentally flawed. The judgment tasks are difficult and confusing; raters' judgments of different item types are internally inconsistent; appropriate validity evidence for the cut scores is lacking; and the process has produced unreasonable results. ... The results are not believable."

For the most part, such pointed criticism has rolled off the governing board like so much water off a duck's back.

As recently as 2009, the U.S. Department of Education received **a report**  on NAEP from the University of Nebraska's Buros Institute. This latest document expressed worries about NAEP's "validity framework" and asked for a "transparent, organized validity framework, beginning with a clear definition of the intended and unintended uses of the NAEP assessment scores. We

Most state assessments aimed at establishing proficiency as 'at grade' produce results different from a NAEP standard.

recommend that NAGB continue to explore achievement-level methodologies.” In short, for the last 20 years, it has been hard to find any expert not on the Education Department’s payroll who will accept the NAEP benchmarks uncritically.

Those benchmarks might be more convincing if most students outside the United States could meet them. That’s a hard case to make, judging by a 2007 analysis from Gary Phillips, a former acting commissioner of the National Center for Education Statistics. Phillips set out to map NAEP benchmarks onto international assessments in science and mathematics and **found**  that only Taipei (or Taiwan) and Singapore have a significantly higher percentage of proficient students in 8th grade science than the United States does. In math, the average performance of 8th grade students in six jurisdictions could be classified as proficient: Singapore, South Korea, Taipei, Hong Kong, Japan, and Flemish Belgium. Judging by Phillips’ results, it seems that when average results, by jurisdiction, place typical students at the NAEP proficient level, the jurisdictions involved are typically wealthy—many with “tiger mothers” or histories of excluding low-income students or those with disabilities.

None of this is to say that the method of determining the NAEP achievement levels is entirely indefensible. Like other large-scale assessments—the International Mathematics and Science Study, the Progress on International Reading Literacy Survey, and the Program on International Student Assessment—NAEP is an extremely complex endeavor, depending on procedures in which experts make judgments about what students should know and construct assessment items to distinguish between student responses. Panels then make judgments about specific items, and trained scorers, in turn, bring judgment to bear on constructed-response items, which typically make up about 40 percent of the assessment.

That said, it is hard to avoid some obvious conclusions. First, NAEP’s achievement levels, far from being engraved on stone tablets, are administered, as Congress has insisted, on a “trial basis.” Second, NAEP achievement levels are based on judgment and educated guesses, not science. Third, the proficiency benchmark seems reachable by most students in only a handful of wealthy or Asian jurisdictions.

Finally, enough questions exist about these achievement levels that Congress should commission an independent exploration to make sense of the many diverse definitions of proficiency found in state, NAEP, and international assessments. A national assessment that suggests proficiency is beyond the reach of students throughout the Western world promises to confuse our educational challenges, not clarify them.

James Harvey is the executive director of the National Superintendents Roundtable, in Seattle. He has been examining the history of the National Assessment of Educational Progress for his dissertation work at Seattle University.