

### ***Who watches the watchmen?***

## ***Transparency might guard the integrity of the tests given by the National Assessment of Educational Progress***

**Sandra Stotsky**

***Quis custodiet ipsos custodes?***, a [Latin phrase](#) found in the work of the Roman poet [Juvenal](#), is commonly used to refer to the problem of how one monitors the actions of persons in positions of power. We don't know if we have a problem with the validity of the results of the "Nation's Report Cards" or the extent of the problem (if there is one) because there is little transparency on the test development process.

Tests given by the National Assessment of Educational Progress (NAEP) have been funded by Congress since their inception in the late 1960s and early 1970s. Called the nation's "Report Cards," all states have been required since 2001 as part of No Child Left Behind to give these tests at least every two years. By law, they can be given only to a stratified random sample of students across each state in each subject, and since the early 1990s they have been given in two forms (Long-term and Main). NAEP tests are given at three specific grade levels (4, 8, and upper high school) according to a schedule worked out in advance. We are told that the 2017 results, to be released in mid-April 2018, may reflect computer-based testing, together with an explanation of how the results of paper-and-pencil tests differ from those generated from student use of computers. We won't know more until the release.

Given the growing dissatisfaction with the statewide tests aligned to Common Core's standards, which are built into all four-year education plans submitted by state departments of education without state legislative or local school board approval to the U.S. Department of Education in 2016/2017, it is not surprising that many parents are concerned about the independence and integrity of these "Report Cards." Do NAEP tests reflect the knowledge and skills in each subject area that subject experts in each area agree should be tested at the tested grade levels to the extent that they are tested? Concerns have been expressed about mathematics in particular because it is the language of science and the foundation of most technical areas of study today and because of the decades-long controversies over how it should be taught, and what should be taught, in K-12.

Two questions need to be answered by the new commissioner of the National Center for Education Statistics (NCES), the organization/agency that creates NAEP tests. Policies influencing the tests must be approved by the National Assessment Governing Board (NAGB), the group appointed on a staggered basis by Congress to shape NAEP policies.

The first question is whether all NAEP mathematics test items are reviewed by a small group of mathematicians as part of the process of test development. The second question is whether new mathematics test items were added for the 2017 tests (at grades 4, 8 and upper high school) that are aligned to Common Core's standards. NCES adds some new test items in most if not all testing cycles, but by law, NAEP test items are not supposed to reflect any particular set of standards.

Why these two questions now? The reason lies in the content of the 2007 National Validity Study, an account of the examination by five mathematicians representing a range of viewpoints on pedagogy who had been asked by NCEES to examine math test items used on NAEP tests. The mathematicians found many test items to be “marginal” or “flawed” and wondered how valid NAEP math tests results could be if they included deficient items. <https://files.eric.ed.gov/fulltext/ED499213.pdf>

Understandably, NCEES staff were upset with this implication. In a lengthy response, NCEES then-Commissioner Mark Schneider raised the possibility that NCEES may not have included enough subject matter expertise in the test development process for NAEP math tests.

[https://nces.ed.gov/whatsnew/commissioner/remarks2007/11\\_23\\_2007.asp](https://nces.ed.gov/whatsnew/commissioner/remarks2007/11_23_2007.asp)

He wrote as follows:

“The most prominent finding of the NVS validity study related to the mathematical quality of the items.... Although the current standing committee has always included at least one mathematician, and there are many mathematicians available at ETS, we may not have achieved the needed representation of mathematicians during item development.”

In other words, he questioned whether sufficient academic expertise was present in the development of the mathematics test items. Five mathematicians with a range of views on mathematics education in K-12 achieved a remarkable consensus on the problems they saw. His remarks further imply that NCEES may not have included sufficiently knowledgeable reviewers of test items in other subject areas as well.

Is it too much to hope that Congress may ask the NAGB

- (1) to ensure that at least a handful of subject matter experts review the test items to be used in each round of NAEP testing in each subject tested, and
- (2) to make public the names of the expert reviewers?