

1. On what basis and timeline does NCES expect the decision about the need for state-level estimates to be made?

The National Assessment Governing Board, in collaboration with NCES, discussed the technical and operational issues related to conducting a national writing assessment using computer-based assessment software. After evaluating the size, scope and technical challenges facing an assessment as large and technically-complex as NAEP, it was decided that the 2011 assessment model would be to:

- Collect state- and national-level estimates for grade 4 writing; it is a paper and pencil assessment, and
- Collect national-level estimates for writing grades 8 and 12; they are computer-based assessments.

There was a deliberate decision to not conduct the 2009 grade 8 writing assessment at the state-level for three primary reasons:

1. Size and complexity of the assessment. This would mean conducting the state-level writing assessment in over 5,000 schools to more than 160,000 students in a six week period. (Note: We expect that, while writing is voluntary, most states (more than 46-48) would have opted to participate.)

The sample is a minimum of 30 students selected at random from across all classrooms in the school. The 60-minute assessment is administered within a 90-minute block of time in a dedicated space within the school.

While NAEP has conducted earlier small studies and the science 2008 field test and 2009 science probe, NAEP has no experience fielding a computer-based assessment across all states and jurisdictions.

2. Quality and reliability of the assessment results are of paramount importance to NAEP. A major NAEP requirement, in order to produce fair comparisons across all 50 states, the Department of Defense Education Activity, and 18 large urban school districts, is that the administrations in all schools are conducted in the same, standardized way.

Consequently, it was decided that it would be better to start small and do a first-class job in this first fully-computer-based assessment than to risk a myriad of technical and operational problems that could be difficult, if not impossible, to ameliorate in the field with a full operational state-level assessment. The discussion below about 2002 writing special study and the 2008 science field test and the 2009 science operational probe refers to most of the factors considered in the decision.

3. Cost. (See next question.)

A computer-based Technological Literacy Assessment in 2012 is proposed for grades 4, 8 and 12. We expect that by 2015, both reading and mathematics assessments will be partially or wholly computer-based as well, leading to computer administrations for most NAEP assessments within the decade.

2. What are the cost implications of planning to produce national versus state-level estimates?

Our estimate for conducting a full state- and national-level writing assessment at grades 4, 8 and 12, as we do for reading and mathematics, is approximately \$34 M. Our projected cost differential with the

national-level only 8th grade writing assessment is \$4 M. Specific cost estimates for the computer component of the assessment are provided in the answer to Q4.

We estimate that we need approximately 800 units to conduct a national-level only assessment (about 12,000 8th grade students) and 8,000 units to conduct a state-level assessment (about 160,000 8th grade students) in the NAEP six-week assessment window. Our early estimate for the marginal computer cost for a national-level assessment (NOT including administrator cost) is approximately \$250,000 and about \$2,500,000 for a state-level assessment. There will be some variations in the estimates based on the complexities of getting the units to remote schools (e.g., Alaska or DoDEA) and the number of states who sign up for the writing assessment.

3. Why is the NAEP program proposing the particular technology approach for which this collection is geared? We had understood that NAEP was interested in pursuing the technology solution that HSLs is using.

The HSLs computer assessment is based on a technical model originally developed by NAEP and was carefully considered by NAEP for the 2011 writing assessment. NAEP has made several attempts to make the school-provided model work for NAEP assessments, including the 2009 science assessment. NAEP had to abandon the model, however, during the 2008 science field test, primarily because limitations of the school-based computer equipment posed several risks to the successful administration of an interactive computer-based assessment, which required a degree of technological sophistication that could not be supported by the original model.

NAEP assessments are designed to be interactive, incorporating simulations and video, and require correct versions of Flash and SVG applications to run on the computers. The form of the assessment, based on the NAEP writing framework, prevents NAEP from using the multiplicity of school-provided computers. The wide variability in computer hardware across schools and even within schools has implications that range from font and format displays to consistency and quality in more sophisticated applications that rely on sound, graphics, or video. This variability can result in different stimulation for and interpretations by students, and alter their responses. The NAEP program is also required to support multiple levels of accommodations, including visual and aural accommodations. In designing the custom solution, the NAEP program has created the assessment solution to provide an integrated environment that supports approved NCES accommodations. Standardizing the hardware will also allow us to standardize extremely frequent back-up, to minimize loss of content and time should a machine fail (due to hardware or student error). Special networking will permit the administrator to reestablish a student on the same or a parallel machine quickly.

This level of standardized administration procedures is critical for NAEP because of NAEP's legislative requirement to report trend over time (i.e., in this case for all assessments through 2023) and the requirement to compare states to the nation and to other states and to compare performance of various subgroups. Administering the assessment in the context of wide variability in school-based computer systems, would violate NAEP's psychometric assumption of standard assessment conditions, and would thus call into question inferences made about student performance. In anticipation of administering computer-based NAEP state-level assessment, the need for standardization is further emphasized by the often politically charged contexts in which states compare their performance on NAEP. In such contexts, NAEP must ensure the elimination of any factors associated with the quality of computer systems within states that might be perceived to account for states' performance on NAEP. Such factors range from the availability of basic functionality and tools such as spell check, bold and cut-and-paste, to the availability of local servers to ensure that a student can be moved to a different computer with no loss of data, in the

event of computer failure. The impact of variability among states is less significant for the HSLs survey, because the HSLs survey is more similar to a paper and pencil test administered on computer, is not interactive, and does not incorporate simulations or the types of video displays referred to above.

At this time, working with NAEP's Sampling and Data Collection Contractor, we believe that 350 8th grade and 350 12th grade schools nationwide will be in the sample and that 800 laptops (units) will be necessary to equip 50 assessment teams with 15 assessment computers each, plus one administrative computer for a total of 800 units.

4. Does NAEP also currently plan to pilot the HSLs approach?

As stated in the answer to question 3 above, NAEP has piloted the technical approach adapted by HSLs, but NAEP's different technical requirements prevented the approach from being utilizable by NAEP.

5. What are the cost implications of a full-scale study using laptops compared to the HSLs model? OMB is interested in seeing a side-by-side comparison.

Below are estimates for the *marginal* costs for NAEP using two different strategies, both of which we have experience implementing. By marginal, we mean that these are the additional costs beyond what would be necessary if we were not administering the assessment on computers. The costs are combined for 8th and 12th grades. For example, we would employ assessors to administer the assessment regardless of the mode of delivery; consequently, the cost for assessors is included only to the extent that we would need to hire them for additional time. These are approximations that we have developed with input from our contractors.

NAEP using school-provided computers:

Contractor charge for one extra school visit to test the school equipment = \$700,000
Contractor charge for technician (assume half the schools) = \$630,000
Contractor charge for additional assessment visits (assume 10% of the schools) = \$210,000
Contractor charge for availability of laptops and related equipment for 1/3 of schools = \$82,500
Contractor charge for Help Desk (6 months to accommodate pre-visits) = \$300,000
TOTAL: \$1,922,500

NAEP providing laptop computers:

Contractor charge for availability of laptops and related equipment = \$250,000
Contractor charge for additional training of field staff (assumes 700 schools) = \$87,500
Contractor charge for Help Desk (7 weeks during Assessor training and assessment window) = \$60,000

TOTAL: \$397,500

For the 2011 assessment, the laptop option is the only way to provide a standard assessment experience across all schools, is the most cost effective option considering the NAEP administration procedures for other assessments (reading and mathematics) being administered in the same window, and provides flexibility for the content and format of future assessments.

6. For this collection, we do not understand why a teacher incentive is necessary. Also, please explain how such an incentive fits within the terms laid out in the supporting statement of the Generic ICR under which this request is submitted, below:

"Respondents for activities conducted in the laboratory (that is, cognitive interviews and focus groups) under this clearance may receive a small stipend. This practice has proven necessary and effective in recruiting subjects to participate in this small-scale research, and is also employed by the other Federal cognitive laboratories. Respondents for methods that are generally administered as part of field test activities (that is, split sample tests, behavior coding of interviewer/respondent interaction, and respondent debriefing) will not receive payment unless there are extenuating circumstances that warrant it."

We have revised the submission to not include a teacher incentive.