

Scope of Work for Model Science Assessment Systems Design Teams

Overview

The National Research Council's Committee on Test Design for K-12 Science Achievement requests that each design team prepare a paper that lays out its conception of a model for a state system of science assessments. At a minimum, the model should meet the requirements of the No Child Left Behind Act of 2002 (NCLB). Accordingly, the assessment system should adhere to the following terms specified in the legislation:

1. States must have challenging academic content standards in science. Science content standards may be grade-specific, cover more than one grade, or may be course-specific at the high school level.
2. States must administer science assessments, which are to be aligned with the state's science standards and *involve multiple up-to-date measures of student academic achievement*, including measures that assess higher-order thinking skills and understanding, at least once each in grades 3-5, 6-9, and 10-12
3. Assessments may include either (or both) criterion-referenced assessments or augmented norm-referenced assessment. The assessments may be comprised of a uniform set of assessments statewide or a combination of state and local assessments.
4. At least 3 achievement levels should be specified (e.g., basic, proficient, and advanced).
5. The same assessment system should be used to measure the achievement of all children, and the system should provide for participation of all students. Reasonable adaptations and accommodations should be made for students with disabilities and limited English proficient students.
6. Assessment results should be reported in aggregate for the full group of test takers, disaggregated for specified population groups, and at the individual level. Reports should include both descriptive and diagnostic information.

The committee encourages design teams to move beyond these specific requirements in proposing a model for building a system of high quality science assessments that is standards based and strives to improve science learning among the nation's students.

Specific Workplan

Each design team will have approximately six months to prepare a 50 to 75-page paper laying out its conception of a model for a state system of science assessments. In addition, design team leaders will be asked to attend the second meeting of the committee, currently scheduled for October 22 - 24, to discuss their approach to this charge. Team leaders and selected members of the teams will be asked to attend the third meeting of the committee and present a mid-course progress report. This meeting is scheduled for late January or early February 2004. Final designs will be presented to the committee in late April or early May. Final versions of the papers will be due by June 1, 2004. Expenses for travel to committee meetings will be covered by the National

Research Council and are in addition to the funds provided for conducting the work of the teams. In addition, members of the design teams may be asked to participate in dissemination activities as a follow up to the committee's report

Key Components of Model Assessment Systems

The committee will lay out a conceptual frame of questions and issues that each design team will need to consider in creating its model. Each team will have some latitude in developing the specific details for its model, however all models should be standards-based and should focus on promoting science learning. The committee's conceptualization of each team's charge will likely include the components described in the following sections. Each design team will be asked to focus on a specific model for designing a system of science assessments and may be asked to emphasize certain aspects of the model. However, it is important that none of the key components of a system be ignored. For all aspects of the model, the team should keep costs as well as states' limited resources in mind and should propose ways to develop systems in an efficient and cost effective manner. In addition, the team should provide estimates of the timeline required for developing and implementing the various components of the proposed science assessment system.

In developing the model, the design team should consider that states are in various stages with regard to their systems of science assessments. Some may have an established system, and their efforts may involve moving to a new system that meets the requirements of NCLB. Others may be in the earliest stages of developing a system. The design team should therefore describe procedures by which a state might adapt its current system to move toward the proposed system as well as procedures for implementing the system from the ground up.

Design teams should lay out an explicit theory of action about how the system would work and how the pieces (state, local, school/classroom – assuming levels in addition to the state would be involved) would be expected to fit together to achieve alignment with state science standards and to support student learning. There should be explicit examples of "pieces" at various levels and how they fit together.

Instructional, Curricular, and Content Issues

The design team should lay out a strategy by which the state can develop a system of science assessments in which curriculum, instruction, and assessments across grade levels and topics are aligned with each other and with state science standards. For the purposes of this report and to provide a common basis for describing this process, the design team should use the National Science Education Standards to exemplify how the strategy might be implemented. The paper should describe the process for identifying the competencies to be covered on the assessment and should detail the steps to be taken to ensure consistency among material covered by the assessments and curriculum and

instruction. The system should include mechanisms by which results from large-scale assessments can inform instruction and classroom practice with the ultimate objective of improving science learning. As part of this discussion, the design team should also specify the process for developing and setting performance standards. In addition, the design team should consider the potential negative consequences associated with the system (e.g., narrowing of science curriculum to teach to the test) and describe ways to circumvent these potential unintended consequences.

Concrete examples should be included in the description of the model assessment system. To assist with this, the committee will negotiate with each team the selection of a conceptually related cluster of standards (e.g., conservation of matter, science and technology, personal and social relation of science) to develop examples at each of the grade levels. The design team should include exemplar items for the cluster of standard(s) at each grade span, and describe how evidence from that cluster could be combined with evidence from other clusters to classify students into one of NCLB achievement levels. Exemplar items must be scientifically accurate, age appropriate, and measure students' understanding of important concepts. In addition, for exemplar open-ended items or performance assessment tasks, the team should provide exemplar scoring rubrics.

Development of the Assessment System

The design team should specify the process for developing the assessment system used at each grade level as well as the strategies they use to assure alignment between levels and across topics. Design teams should be as specific as possible about test specifications and the rationale for their test blueprints.

In addition to procedures for identifying the skills, content, and competencies to be evaluated, there should be discussion of the procedures for determining the item format(s) to be used on the assessments. Consideration should be given to a variety of available item formats, such as, multiple-choice, constructed response, performance assessments, portfolios, etc. In addition, consideration should be given to assessment tasks that rely on teachers' ongoing appraisals of performance in the classroom. Further, the design team should specify the process used for determining the developmental appropriateness and scientific accuracy of the tasks to be included at each grade span. The design team should discuss how the various formats might be incorporated into a comprehensive system of assessments targeted at measuring a wide range of cognitive skills across grade levels. Design teams should suggest ways that district-wide or local classroom assessments that are aligned to state standards, curriculum, instruction, and the large-scale state assessment can be used in conjunction with the large-scale assessment to inform instruction.

Design teams should include a description of processes and procedures to be used to conduct bias, sensitivity, and technical reviews of items and tasks. The proposed

processes and procedures should include the methods for reviewing items and for involving teachers, other educators, and science experts in the review process. The review process should pay special attention to ways to ensure that items and tasks are accessible to students with disabilities and English language learners. In addition, technical reviews should include a plan for ensuring that items and tasks are scientifically accurate and age appropriate.

The model should also include a plan for developing scoring procedures. For example, if the system includes open-ended items, the plan should include discussion of ways to develop a scoring rubric for open-ended items and the mechanism for training scorers and conducting the scoring process.

In developing the model, the design team should consider that one potential objective of science assessments under future legislation will likely be to track performance trends over time. Thus, the proposed model should include discussion of ways to implement appropriate scaling and equating procedures that will enable maintenance of performance trends.

Given that NCLB calls for reporting results according to performance standards, the model should include discussion about processes for determining and setting performance levels.

Involving Teachers

The model should include a variety of mechanisms for involving teachers in the design and development of the assessment system, both to help the assessment system function well, and to help the teachers learn how to use the assessment system to improve their instructional practices. The design team should consider ways in which teachers can participate in item writing, item review, scoring, and other assessment development activities. In describing these plans, the design team should outline the ways in which teachers would be trained to participate in these activities.

Professional Development

The model should include plans for ensuring that teachers and administrators are fully informed about the assessment system – the content and skills evaluated, the means for evaluating mastery of these skills, and the ways results are reported. Details should be included for ways to provide professional development activities that educate teachers and administrators about how best to prepare students for the assessment, understanding assessment results, and using them to make instructional decisions.

Including and Accommodating Students with Special Needs

In developing its model, the design team should keep in mind that a primary objective of NCLB is to include all students in the assessment system and should propose ways for accomplishing this objective. The discussion should include procedures for developing assessments so as to reduce the need for accommodations, e.g., making sure time limits are reasonable, using plain language. In addition, the plan should include discussion of procedures for specifying the kinds of accommodations that should be offered to students with disabilities and English language learners.

Reporting Assessment Results

The design team should develop a plan for reporting assessment results to wide variety of audiences, including students, parents, teachers, schools, school districts, and states. The proposed model should include examples of reports that are appropriate for each of these audiences. In developing samples of reports, the design team should consider the ways the reported information might be used and develop sample reports that are appropriate, given these uses. The committee is particularly interested in examples of reports that would be useful for teachers and school administrators in planning instructional programs.

Use of Assessment Results

Although science assessment does not currently fall under the accountability measures of NCLB, design teams should consider the ways that the reported assessment information might be used in an accountability system. In addition, design teams should discuss and provide examples of the ways in which reported assessment results can be used by teachers and principals to evaluate students' achievement and inform instructional practice.

Meeting Standards for Technical Quality

Design teams should consider and incorporate professional technical standards for content and testing as detailed in the *National Science Education Standards* (National Research Council, 1996) and in *Standards for Educational and Psychological Testing* (AERA, APA, and NCME, 1999).

Use of Technology

In designing its model, the team should consider ways in which technology can be used in the assessment system to make the system more efficient. In particular, the design team should outline ways technology could be used to enhance evaluation of skills, to utilize innovative item formats, to provide accommodations to students with special needs, to score open-ended responses, and/or to enhance score reporting.