

**Illustrative Language for an RFP  
To Build Tests To Support  
Instruction and Accountability**

**Prepared By Members of the  
Commission on Instructionally Supportive Assessment**

**Convened By**

**American Association of School Administrators  
National Association of Elementary School Principals  
National Association of Secondary School Principals  
National Education Association  
National Middle School Association**

**October 2001**

This report was prepared by the following members of the Commission on Instructionally Supportive Assessment. Commission members' affiliations do not denote institutional endorsement of this document.

David C. Berliner  
Regents' Professor of Education  
College of Education  
Arizona State University

Carol Camp Yeakey  
Professor of Urban Politics and Policy  
Curry School of Education  
University of Virginia

James W. Pellegrino  
Distinguished Professor of Cognitive Psychology  
Distinguished Professor of Education  
University of Illinois at Chicago

W. James Popham (Chair)  
Professor Emeritus  
Graduate School of Education and Information Studies  
University of California, Los Angeles

Rachel F. Quenemoen  
Senior Fellow for Technical Assistance and Research  
National Center on Educational Outcomes  
University of Minnesota

Flora V. Rodríguez-Brown  
Professor of Curriculum and Instruction/Reading, Writing and Literacy  
University of Illinois at Chicago

Paul D. Sandifer (Ret.)  
Consultant to the Office of Assessment  
South Carolina Department of Education

Stephen G. Sireci  
Associate Professor  
School of Education  
Center for Educational Assessment  
University of Massachusetts, Amherst

Martha L. Thurlow  
Senior Research Associate, Department of Educational Psychology  
College of Education and Human Development  
Director, National Center on Educational Achievement  
University of Minnesota

## Introduction

In a separate report, *Building Tests To Support Accountability and Instruction: A Guide for Policymakers*<sup>1</sup> the Commission on Instructionally Supportive Assessment identifies nine requirements that must be satisfied if statewide tests are to support both instruction and accountability:

- Requirement 1: A state's content standards must be prioritized to support effective instruction and assessment.
- Requirement 2: A state's high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident.
- Requirement 3: The results of a state's assessment of high-priority content standards should be reported standard-by-standard for each student, school, and district.
- Requirement 4: The state must provide educators with optional classroom assessment procedures that can measure students' progress in attaining content standards not assessed by state tests.
- Requirement 5: A state must monitor the breadth of the curriculum to ensure that instructional attention is given to all the content standards and subject areas, including those that are not assessed by state tests.
- Requirement 6: A state must ensure that all students have the opportunity to demonstrate their achievement of state standards; consequently, it must provide well-designed assessments appropriate for a broad range of students, with accommodations and alternate methods of assessment available for students who need them.
- Requirement 7: A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy *Standards for Educational and Psychological Testing* and similar test-quality guidelines.
- Requirement 8: A state must ensure that educators receive professional development focused on how to optimize children's learning based on the results of instructionally supportive assessments.

<sup>1</sup> *Building Tests To Support Accountability and Instruction: A Guide for Policymakers*, is available online from the associations that convened the Commission on Instructionally Supportive Assessment at: [www.aasa.org](http://www.aasa.org), [www.naep.org](http://www.naep.org), [www.principals.org](http://www.principals.org), [www.nea.org](http://www.nea.org), and [www.nmsa.org](http://www.nmsa.org)

Requirement 9: A state should secure evidence that supports the ongoing improvement of its state assessments to ensure those assessments are (a) appropriate for the accountability purposes for which they are used, (b) appropriate for determining whether students have attained state standards, (c) appropriate for enhancing instruction, and (d) not the cause of negative consequences.

State policymakers might adopt various approaches to address these requirements; for example, by directing a state agency to do so. Many states, however, will likely seek the assistance of external contractors to undertake work related to one or more of the Commission’s requirements. Typically, such support is solicited in the form of a request for proposals (RFP) issued to suitable agencies or individuals.

This report presents illustrative language that states might incorporate into RFPs they issue to satisfy the Commission’s requirements. These RFPs may be addressed to test contractors, curriculum and instruction consultants, independent researchers, or other agencies depending on the nature of the requirement.

State personnel who write RFPs are encouraged to adapt the illustrative language; hence the report is available in html as well as PDF format. The Commission does not endorse particular procedures for satisfying its nine requirements, nor does it recommend the particular RFP language found in this report. Instead, the language is provided to assist states as they develop RFPs.

The illustrative language in this report focuses on one aspect of the RFPs that states might issue—procedures to create, install, and evaluate assessments that meet both instructional and accountability purposes. Other than the tests themselves, states will not need to depart substantially from what is routinely included in RFPs that solicit bids for developing statewide tests (e.g., qualifications of the bidders, reporting requirements, payment schedules). Nor will states need to alter substantially many of the technical issues normally addressed in assessment-related RFPs (e.g., year-to-year equating of tests administered at certain grade-levels, establishing performance standards based on specific standard-setting procedures).

However, as a state moves toward assessments that fulfill both instructional and accountability purposes, and that are fully inclusive, some technical requirements will change to meet the expectation that all students are included in the assessments, and that reports of students’ performance support instructional improvement. These changes are not addressed in this report; therefore, states will need to work with their contractors and stakeholders to ensure that the technical requirements of their RFPs are consistent with the purpose and function of the assessments described here.

Finally, before turning to illustrative RFP-language related to each of the Commission’s nine requirements, states may wish to consider including introductory language in their RFP such as the following: “The procedures specified in this RFP represent an approach to accomplishing the required work. Bidders are invited, however, to propose alternative procedures provided those alternatives are cost-effective and supply the required services and deliverables.”

## **Illustrative RFP Language**

States may solicit proposals from a variety of contractors to conduct work related to the Commission's nine requirements. The illustrative language provided for Requirements 1, 2, 3, 4, and 6 is appropriate for soliciting proposals from test developers. Work related to Requirement 1 also might be solicited from a vendor other than a testing firm. Requirements 5, 7, 8, and 9 specify data collection and evaluation tasks best completed by contractors other than testing firms.

Because Requirements 1, 2, 6 and 7 entail procedures with which architects of state-issued RFPs may be unfamiliar, detailed language is offered in this document. Less detailed language is suggested for Requirements 3, 4, 5, 8, and 9. And in some instances, such as Requirement 8 that deals with professional development programs for educators, the report offers limited language. Experienced staff developers in most states can provide expert assistance in that aspect of the RFP's preparation.

Throughout the report segments of the illustrative language are bracketed to remind readers to modify this text for specific state contexts. For instance, the illustrative RFP-language related to an item-review committee reads: "An item-review committee of [20] members will be assembled at [a site to be recommended by the bidder]." The size of the item-review committee and the site of its meeting are the decision of the individuals preparing the RFP.

In addition, the report includes examples of products or procedures that are described in the RFP language. These examples are presented in italicized text.

### **Requirement 1**

**A state's content standards must be prioritized to support effective instruction and assessment.**

#### **Illustrative RFP-Language**

The tests that bidders develop in response to this RFP must accurately estimate the degree to which a student has attained state content standards<sup>2</sup> that represent essential or high-priority skills and/or knowledge.

Therefore, bidders will need to carry out a defensible procedure for prioritizing the state's content standards in each of the subject areas/grade levels for which a state test is being developed. This procedure must be based primarily on preferences registered by

<sup>2</sup> States will need to make sure that any illustrative language in this report meshes with the manner in which their content standards have been developed and approved.

committees of educators approved by the Department of Education (DOE).<sup>3</sup> Classroom teachers and curriculum specialists must be involved in prioritizing the content standards. Neither (DOE) staff nor the bidder's staff should take a substantive role in prioritizing content standards. It would, however, be appropriate for them to comment on any technical or practical difficulties likely to be encountered in assessing a given content standard.

Because prioritizing will reduce the number of content standards that will be assessed by state tests, it is important for each content standard assessed by the tests to represent an essential or significant skill and/or body of knowledge. It is unlikely that any of the content standards assessed by the test would deal exclusively with knowledge. Rather, it seems more likely that bodies of knowledge would be subsumed under content standards that focus on high-level cognitive skills.

Therefore, it may be necessary to reconfigure some current content standards so that they subsume other content standards. If reconfigurations of content standards result from prioritizing and are proposed, the bidder must provide a justification. In particular, bidders must provide evidence that any content standards subsumed under existing, or reconfigured, standards represent enabling subskills or knowledge for the essential skill/knowledge sought.

A useful illustration is represented by assessments of students' written composition skills in which students generate original writing samples. Written communication skill is a significant cognitive competency that embraces a number of subskills (e.g., written mechanics, content organization) as well as knowledge of the topic addressed in the composition. But these enabling skills are subsumed by a single, powerful content standard. These are the kinds of high-import content standards that should be assessed on the state test(s).

To clarify the DOE's intent, a prioritizing procedure is described below. Bidders are free to modify this process or to use another procedure.

### ***Example of A Procedure to Prioritize State Content Standards***

*The bidder will convene a panel [20-25 members], [identified in collaboration with DOE] to prioritize the state's content standard in [the subject areas/grade levels]. The panel [Curricular Prioritizing Panel (CPP)] shall be composed of equal numbers of classroom teachers, other educators, curriculum specialists from school districts, and representatives from institutions of higher education who specialize in the content area for which the test is being developed. Typically, different CPPs will be needed for tests in different content areas, although the same panel might prioritize content standards at different grade levels in the same content area.*

*The bidder will moderate at least one meeting among panel members. [DOE personnel will be present as observers.] Prior to the CPP meeting, the bidder will provide panel members with a*

<sup>3</sup> Department of Education (DOE) is used throughout the report to designate the state agency responsible for issuing RFPs. Some states use different designations for this agency, or issue RFPs through another state agency.

*set of the appropriate state content standards, a description of the procedures panel members will follow to prioritize the standards during the CPP meeting, and information on required preparation for the meeting. As important, bidders must tell CPP members about all the steps in the prioritization process so members understand how their ratings and rankings of content standards will be used to arrive at final sets of standards the bidder will provide DOE.*

*Each CPP member will be asked to review the content standards before the meeting. Panel members should be supplied with forms to rate the importance of each standard. The nature of this review will depend on the specific prioritizing process proposed by the bidder. To illustrate, one possible procedure would require panel members to rate each content standard using a rating system such as the following:*

- (a) essential (i.e., standards that are the most important for students to attain);*
- (b) very desirable, (i.e., standards that are extremely important for students to attain);*
- (c) desirable (i.e., standards that are somewhat important for students to attain;*
- (d) optional (i.e., standards that are least important for students to attain).*

*The bidder must describe the specific language and response categories they intend to use in securing per-standard ratings.*

*Bidders will compile CPP members' initial ratings into four groups (essential, very desirable, desirable, and optional) so that panel members can review and discuss their initial ratings at the panel meeting. The bidder must indicate decision-rules it employs to place the standards into these categories.*

*During the CPP meeting panel members will be asked to develop (1) a set of essential content standards ranked from most-to-least important and (2) a set of very desirable content standards ranked from most-to-least important. To expedite this process, the bidder could present panel members with the content standards they have rated individually as essential and very desirable, ask them to rank each set of standards from most to least important, discuss, and as needed, revise their ratings. The bidder may suggest more than one iteration of this process. During the meeting, panel members also might be asked to reformulate existing content standards to represent broad, significant outcomes. If DOE prefers, it could carry out this reformulation and provide these reconfigured standards to panel members for their consideration.*

*As a result of the CPP members' work, the bidder will provide DOE (1) a set of essential content standards ranked from most-to-least important and (2) a set of very desirable content standards ranked from most-to-least important.*

## Requirement 2

**A state’s high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident.**

### **Illustrative RFP-Language**

So that educators understand clearly the skills and knowledge in the content standards that are assessed by state tests developed in response to this RFP, bidders must supply a short, clear, and concise assessment description for each standard. These assessment descriptions, which should not exceed three paragraphs, must describe the cognitive domains and cognitive demand(s) students are expected to meet. In addition, each assessment description must be accompanied by a minimum of [three] different illustrative item-types that might be used to assess the standard. [These sample items must be labeled as “Illustrative Item-Types.”]

Each assessment description should focus on the nature of the cognitive demand(s) to which students are expected to respond. Educators who read the assessment description, and the content standard on which it is based, should be able to arrive at an accurate understanding of the intellectual operation(s) that the content standard requires of students. Moreover, if any enabling subskills or bodies of knowledge are identified in the assessment description, educators will understand the nature and importance of these subskills or bodies of knowledge.

Bidders must identify the kinds of personnel who will develop the assessment descriptions. The DOE strongly urges bidders to involve both instructional and measurement staff. In addition, educators, especially classroom teachers, must be involved in the generation and/or review of the assessment descriptions since they will be the important users of the descriptions.

All assessment descriptions should be pilot-tested with small groups of educators, and bidders should describe acceptable outcomes from these pilot tests as well as any decision-rules they will employ to determine a description’s adequacy. Bidders also should indicate the kinds and numbers of educators who participate in the pilot-tests.

To assist bidders, an illustrative assessment description that shows one way of satisfying this section of the RFP is provided below.

### ***Example of an Assessment Description***

*Overview: This assessment description and the illustrative tasks that follow it are intended for an eleventh- and twelfth-grade U.S. history course. If the skill in this assessment description is promoted at earlier grade levels in which U.S. history is taught, the language and cognitive complexity of the illustrative items will need to be simplified.*

*To measure this skill, the state must identify (and communicate to educators) eligible historical events for students to consider. As an example, the following historical events were identified as suitable by one district for its eleventh- and twelfth-grade U.S. history courses: Constitution,*

*Territorial Expansion, Civil War, Reconstruction, Industrial Revolution, Imperialism, World War I, Depression, New Deal, World War II, Cold War, Civil Rights, Viet Nam, Communication Revolution.*

Content Standard: *Students will be able to draw upon historical lessons to deal with today's societal problems.*

Assessment Description: *Given a prose account of a real or fictitious current problem, as well as a proposed solution to that problem, students will be able to respond appropriately to any one or any combination of the following subtasks:*

*Subtask 1: Event Selection. Identify at least one significant historical event (such as the industrial revolution) that is, at least in part, germane to the problem and its proposed solution.*

*Subtask 2: Event Justification. Justify the relevance of the identified historical event(s) to the problem and its proposed solution.*

*Subtask 3: History-based Prediction. Make a defensible history-based prediction regarding the proposed solution's likely consequences.*

*Subtask 4: Defense of a Prediction. Support that prediction on the basis of parallels between the identified historical event(s) and the proposed problem-solution.*

*Students will be presented, orally or in writing, with a real or fictitious current day problem-situation along with a proposed solution to that problem. (See illustrative items below.) After they have had an opportunity to consider the problem and the proposed solution, they will be asked to supply written or oral responses to one or more of the four subtasks listed above.*

*Although they may be asked to supply responses to individual subtasks so that their mastery of those particular types of subtasks can be determined, all students will ultimately be required to respond to a comprehensive task such as the one illustrated below. When they respond to a comprehensive task, of course, they will be given a problem and proposed solution that are different from the ones used for individual subtasks.*

Illustrative Tasks/Items. *Directions: Read the fictitious problem described below as well as the proposed solution to that problem, then respond to each of the tasks indicated. (Note: These illustrative items would not be presented on the same test form.)*

## WAR OR PEACE

*Nation A, a large, industrialized country whose population is almost 100,000,000, has ample resources, and is democratically governed. It also owns two groups of islands that, although distant, are rich in iron ore and petroleum.*

*Nation B, a country with far fewer natural resources and a population of only 40,000,000, is about one-third as large as Nation A. Although much less industrialized than Nation A, Nation*

*B is as technologically advanced as Nation A. A three-member council of generals governs Nation B.*

*Recently, without any advance warning, Nation B ruthlessly attacked Nation A. As a consequence of this attack, more than half of Nation A's military equipment was destroyed. After its highly successful surprise attack, Nation B's rulers have proposed a "peace agreement" calling for Nation A to turn over its two groups of islands to Nation B. If Nation A does not concede the islands, Nation B's rulers have threatened all-out war.*

*Nation A's elected leaders are fearful of the consequences of the threatened war because their military equipment is now much weaker than that of Nation B. Nation A's leaders are faced with a choice between (1) peace obtained by giving up the islands or (2) war with a militarily stronger nation.*

*Nation A's leaders decide to declare that a state of war exists with Nation B. They believe that even though Nation B is now stronger, in the long term Nation A will prevail because of its greater industrial capability and richer natural resources.*

*Task 1 In an essay, drawing on your knowledge of American history, select one or more important historical events that are especially relevant to the fictitious situation described above. Then justify the relevance of your selection(s). Next, make a reasonable history-based prediction about the likely consequences of the decision by Nation A's leaders to go to war. Finally, defend your prediction based on the historical event(s) you have identified.*

*Note: The evaluation of a student's response to this four-step, comprehensive task will be based on the quality with which each of the following have been carried out: [1] event(s) selection, [2] event(s) justification, [3] history-based prediction, and [4] defense of prediction. Individual subtasks would, apart from the "War or Peace" illustration, require a different problem and proposed solution.*

*Task 2 In the spaces below, name one or more important events in American history that are particularly relevant to the problem and proposed solution described above.*

*Task 3 In an oral presentation of one-to-two minutes' duration, identify at least one important historical event that is especially pertinent to the situation described above, then justify why you believe this to be so.*

*Task 4 In the situation given in the above box, two fictitious nations are described. From the four choices below, choose the one answer presenting the two nations and the armed conflict most comparable to those described above.*

- A. Nations: U.S. and Italy I; Conflict: World War I*
- B. Nations: U.S. and North Korea; Conflict: Korean "Police Action"*
- C. Nations: U.S. and Spain; Conflict: Spanish American War*
- D. Nations: U.S. and Japan; Conflict: World War II*

## **Requirement 3**

**The results of a state’s assessment of high-priority content standards should be reported standard-by-standard for each student, school, and district.**

### **Illustrative RFP-Language**

Tasks that the successful bidder completes under Requirement 1 will result in a list of prioritized content standards. To satisfy this section of the RFP, bidders must provide DOE with strongly supported recommendations regarding which of the highest-priority content standards (see Requirement 2) will be assessed during the time available for testing [which the DOE addresses elsewhere in this RFP].

In addition, the tests that are developed in response to this RFP must enable the DOE and educators to make valid inferences about students’ performance related to each content standard. Therefore, bidders must (a) specify the number of assessment items/tasks related to each standard, and (b) provide evidence that the tests they construct will yield accurate per-standard evidence of a student’s performance.

Bidders also must provide the DOE with information on the probable accuracy of reporting test results standard-by-standard at various levels of aggregation. That is, bidders must supply estimates of the likely accuracy of per-standard reports of student performance at the state, district, school, and student levels. Similarly, bidders must provide probable accuracy-estimates at different reporting levels not only for total-group results, but also for disaggregated student groups (e.g., racial, disability, limited English proficiency, and gender subgroups). To ensure greater accuracy, bidders may choose to report results aggregated across standards in addition to per-standard reporting. If this is the case, the proposed procedures for such aggregation must be described.

At the same time, bidders must describe how per-standard test results will be reported to relevant constituencies, for example, to educators, parents, and students.

The DOE recognizes the technical problems associated with providing accurate estimates of individual student’s performance on each content standard. Yet, it is our conviction that per-standard results, even results whose accuracy in certain instances will be somewhat reduced, will prove to be instructionally useful to educators who currently have no evidence about students’ performance related to individual content standards.

## **Requirement 4**

**The state must provide educators with optional classroom assessment procedures that can measure students' progress in attaining content standards state tests do not assess.**

### **Illustrative RFP-Language**

As a consequence of prioritizing the state's content standards, many important standards will not be assessed by state tests that will be developed in response to this RFP. Nevertheless, these standards should be assessed. The DOE believes this can be done via classroom assessments that can either be developed by, or supplied to, educators. To this end, the DOE intends to make available *optional* classroom assessments the state's educators can use if they choose.

Therefore, bidders must describe how they would create assessment descriptions and build optional classroom assessments for the content standards that are not assessed on state tests. Each bidder's proposal must indicate awareness of the need to create (a) valid classroom assessments that busy educators are likely to use; and (b) reliable classroom assessments whose results schools and districts can report to the DOE for accountability purposes.

Bidders must describe the procedure they will use to choose the content standards designated for the generation of optional classroom assessments. DOE must approve the content standards finally selected for this activity.

For each content standard chosen, and at each grade level for which an optional classroom assessment is to be produced, bidders must create two forms of each assessment. This will provide educators more flexibility in how they use the assessments. For example, the assessments could be employed as instructional pretests and posttests. [Although these tests do not need to be of equal difficulty psychometrically, bidders must create test forms that are reasonably similar in difficulty levels.]

Bidders must describe procedures they will employ to assess and provide the DOE with data on the optional classroom assessments' validity and reliability. To this end, bidders must describe how they have developed, piloted tested the assessments with students and educators, and revised and finalized the assessments. Each test form must be pilot-tested with students and educators. In addition, each test form should be pilot-tested on at least [six] students in a session in which students' reactions to the test are sought and documented after they complete the test form.

Because the costs associated with this activity will vary based on the number of tests produced, bidders should supply cost estimates for the assessments in [10-test] increments [(two forms per test)], that is, for [10] pairs of optional classroom assessments, [20] pairs of classroom assessments, etc.

The bidder must provide the DOE with recommendations for disseminating the optional classroom assessments to educators through online and print avenues. Any information educators need to administer, interpret, and/or score the assessments must be made available to the DOE at the time the classroom assessments are delivered. The assessments must be easy to use and easy to score.

## **Requirement 5**

**A state must monitor the breadth of the curriculum to ensure that instructional attention is given to all content standards and subject areas, including those that are not assessed by state tests.**

### **Illustrative RFP-Language**

State tests focused on high-priority content standards could inadvertently narrow curriculum coverage. Therefore, bidders must propose cost-effective methods that they will use to provide *annual* estimates of curricular coverage at the state, district, and school levels. These methods may reflect either qualitative or quantitative procedures, or a combination of both.

Bidders' proposed methods of monitoring curricular breadth should provide evidence regarding both the *intended* curriculum and the *enacted* curriculum. In other words, bidders must describe how they propose to monitor curricular aims as well as the curriculum that is actually implemented, which may or may not correspond closely with the intended curriculum. Bidders must make clear how their methods of monitoring curricular breadth would address this possibility.

In addition, bidders must propose a reasoned approach for reporting, and including in accountability processes, the results of local classroom assessments measuring standards that are not part of the state test (see Requirement 4). These local assessments must be an essential component of the monitoring of statewide curricular breadth.

## **Requirement 6**

**A state must ensure that all students have the opportunity to demonstrate their achievement of state standards; consequently, it must provide well-designed assessments appropriate for a broad range of students, with accommodations and alternate methods of assessment available for students who need them.**

### **Illustrative RFP-Language**

All students must have the opportunity to demonstrate their achievement of the same content standards. Therefore, to satisfy this section of the RFP, bidders must design state tests that allow the maximum number of students possible (and students with diverse characteristics) to take the same assessments without threat to the validity and comparability of the scores.

To this end, bidders must demonstrate how they will develop “universally designed assessments.” Designed from the beginning to allow participation of the widest range of students, these assessments result in valid inferences about the performance of all students, including students with disabilities, students with limited English proficiency, and students with other special needs.

While universally designed assessments will not eliminate the need for all accommodations (e.g., tests in Braille or large print, extended time to take the test, individual test administration), they can significantly reduce the need for them. As important, universally designed assessments increase the variety of accommodations that can be used without threat to the validity and comparability of the scores. Overall, these assessments result in inclusive accountability measurement, and they provide instructionally supportive information across the full range of students.

Universal design processes do not change the definition of the construct measured by the assessment but focus instead on the following assessment-design characteristics that bidders should integrate into their test development procedures.

Test conceptualization. Bidders must define the construct(s) to be measured precisely and explicitly so tests measure the construct while minimizing the effects of irrelevant factors. Bidders also must include the full range of students in the definition of the target population that will take the assessments.

Test construction. Bidders must develop items that minimize the effects of extraneous factors (e.g., avoid unnecessary use of graphics that cannot be presented in Braille, use font size and white space appropriate for clarity and focus, avoid unnecessary linguistic complexity when it is not being assessed). Bidders also must provide for a full range of test performance to avoid ceiling or floor effects, and they must develop an item pool of sufficient size to permit the elimination of items that are not found to be universally appropriate during the test tryout and item analysis.

Test tryout, analysis, and revision. Bidders must include a full range of students in the tryout sample (e.g., students with disabilities, students with limited English proficiency, other students with special needs). Because there may be constraints in sampling due to the low numbers of students with specific characteristics, bidders may need to identify over-sampling strategies, (e.g., select groups of items for which additional sampling will occur). Bidders should include the use of accommodations during the test tryout.

As part of required test item analysis, bidders must analyze item characteristics to determine which items can be used with the full range of students and with accommodations. This includes examining items for evidence of disability bias and eliminating such items during test revision. Bidders must (a) include the full range of students and (b) use accommodations in any test administration conducted during tryout and revision.

Identification of other assessment methods. Bidders must describe how they will assist educators in devising alternative methods of assessing student progress (e.g., panel review, performance or portfolio assessment) for students who cannot take the universally designed state tests. Alternative methods must assess progress toward the same content standards and the same levels of performance that are assessed in the state tests. This requirement is in addition to any responsibility bidders must assume for developing federally required alternate assessments.

## **Requirement 7**

**A state must generally allow test developers a minimum of three years to produce statewide tests that satisfy *Standards for Educational and Psychological Testing* and similar test-quality guidelines.**

### **Illustrative RFP-Language**

This RFP provides sufficient time (i.e., three years for a contractor to create new state tests) for bidders to meet the assessment profession's standards related to test construction. [The major timeline milestones set forth elsewhere in the RFP indicate the major deadlines with which bidders must comply.]

As described elsewhere in this RFP, bidders must describe how they intend to use the time provided to create state tests that satisfy the DOE's need for accountability evidence *and* also support educators' instructional efforts (see Requirements 2, 3, and 4). The DOE regards the precepts embodied in the *Standards for Educational and Psychological Testing*<sup>4</sup> as key to ensuring that state tests will be suitable for these purposes. Therefore, bidders must explain how they will comply with the *Standards'* stipulations (e.g., assembling compelling validity evidence).

Bidders also must describe any evaluative activities in which they will engage to ensure that the state tests and the optional classroom assessments (see Requirement 4) they produce, and any related deliverables or services, are of high quality. These evaluative activities might involve the use of “cognitive laboratories” in which small groups of students respond to test items and describe to members of a bidder’s staff the nature of the intellectual operations that they have employed to complete the items. Bidders who choose to use such evaluative activities must describe the decision-rules they will employ to revise test items using evidence they have obtained from the activities.

If bidders plan to employ judgmental reviews of test items, the key elements of those reviews must be explicit. At a minimum, these reviews must address the item-quality issues embodied in the following four questions:

Standards congruence: Will students’ responses to a test item help educators accurately determine whether students have demonstrated the knowledge and/or skill embodied in the designated content standard the item is intended to assess?

Out-of-school factors: Is the test item free of content that would make a student’s likelihood of answering it correctly be dominantly influenced by factors other than what has been taught in school?

Instructional sensitivity: If a teacher has provided effective instruction related to this item’s content standard, is it likely that most of the teacher’s students will answer the item correctly?

Absence of bias: Is a test item essentially free of content that might offend or unfairly penalize students because of personal characteristics such as race, gender, religion, primary language, or disability?

Bidders who plan to use judgmental reviews of test items must describe the key elements of those reviews. To clarify the DOE’s intent, a review process is described below. Bidders are free to modify this process or use another procedure.

### ***Example of a Process to Review Assessment Descriptions and Test Items***

*Because well-formed assessment descriptions will play a pivotal role in the subsequent review of potential test items, it is important to refine all assessment descriptions as soon as possible. Therefore, the bidder will convene committees of experienced educators to review both the assessment descriptions (by mail) and the pool of proposed test items (in person).] These committees, one for each content area and grade level for which a test is being constructed and consisting of [20-30] members, can be described as [Materials Review Committees (MRCs)]. DOE personnel will designate the membership of each MRC, but most members will be classroom teachers or those who work closely with classroom teachers. The bidder will be responsible for recruiting members for the MRCs, although DOE will supply a list of potential invitees [plus suitable cover letters inviting participation in an MRC]. The bidder also will prepare an invitational letter for potential MRC participants. That letter, approved by DOE,*

*will describe the state tests the bidder is developing and how the MRC's review of assessment descriptions and test items fits into test development.*

*As a first step in the review, the bidder will provide each MRC member with the prioritized content standards that the state test is designed to measure and will draft assessment descriptions for each standard. Each content standard and assessment description is intended to provide educators with clear descriptions of the cognitive skills assessed by the state test.*

*Bidders will ask reviewers to (a) read each content standard and its accompanying draft assessment description and (b) evaluate each description's usefulness in instructional planning and ease of use. Reviewers' judgments should be made in response to the following questions or to improved versions of these questions suggested by the bidder. [Abbreviated versions of these questions also could be provided to reviewers as they rate assessment descriptions.]*

*Usefulness for instructional planning: 1. After reading the content standard on which this assessment description is based, and the description itself, would educators have a sufficiently clear idea of the cognitive demands required of students to plan effective instruction for students related to the content being measured? (Reviewers would respond: definitely yes, probably yes, probably not, or definitely not.) 2. If you have suggestions for improving this assessment description's usefulness for instructional planning, please supply them below.*

*Ease of Use: Is the assessment description sufficiently understandable and concise so that most educators would find it easy to read and use? (Reviewers would respond: definitely yes, probably yes, probably not, or definitely not.) 2. If you have suggestions for making the assessment more useful to educators, please supply them below.*

*Bidders will supply MRC members with pre-stamped envelopes to return their reactions to the draft assessment descriptions. After those reactions have been considered, it is likely that some of the descriptions will need to be modified, thereby leading to another round of by-mail MRC reactions to any modified assessment descriptions. Appropriate DOE staff must approve all modifications. It is imperative that when the MRCs meet to review test items, they have assessment descriptions to work with that the vast majority of MRC members have judged to be easy for educators to use and useful in instructional planning.*

*As a second step in the review, bidders will convene members of the MRC to review test items under conditions that are carefully monitored by members of the bidder's staff. The bidder must describe in detail the security procedures it will employ during these item-review sessions. Moreover, the main elements of the item-review procedure should be described when bidders respond to this RFP. For example, the nature of any planned orientation the bidder will provide MRC members must be well described.*

*The bidder may choose to use an item-review process in which reviewers:*

- 1. Individually answer specified questions for each test item in a subset of items.*
- 2. As a group consider the subset of items to see if any need to be modified.*
- 3. Individually examine any test items that have been altered.*
- 4. As a group reconsider any altered items.*

*Bidders may propose other item-review procedures, which they must describe in detail. In addition to the test items, reviewers need the final assessment description for each content standard as well as the content standard. The review process suggested here is particularly useful if MRCs review items in subsets of [10-15] items. The individual/ group procedure continues until all items have been reviewed.*

*The item-review questions the MRC members answer must address, at least, the item-quality issues embodied in the following four questions:*

*Standards congruence: Will students' responses to a test item help educators accurately determine whether students have demonstrated the knowledge and/or skill embodied in the designated content standard the item is intended to assess?*

*Out-of-school factors: Is the test item free of content that would make a student's likelihood of answering it correctly dominantly influenced by factors other than what has been taught in school?*

*Instructional sensitivity: If a teacher has provided effective instruction related to this item's content standard, is it likely that most of the teacher's students will answer the item correctly?*

*Absence of bias: Is a test item essentially free of content that might offend or unfairly penalize students because of personal characteristics such as race, gender, religion, primary language, or disability?*

*Bidders may propose the use of additional item review questions. It should be noted that this suggested review process requires bidders to have indicated (at item-review time) the chief content standard that each item is intended to assess.*

*The length of time the MRC meets depends on the number of items to be reviewed. Based on experience in conducting item-review meetings, bidders should propose and defend the probable length of such meetings. [Bidders will be responsible for all travel-related expenses associated with the MRC meetings and for any stipends paid MRC members.]*

*At the conclusion of each MRC session, the bidder will summarize the review data on all evaluative dimensions employed (e.g., for each of four questions cited above). Then the bidder and the DOE will designate test items deemed suitable for field-testing. The vast majority of the items selected for field-testing will have received positive evaluations on all evaluative dimensions employed during the MRC sessions.*

*Bidders are to describe, in sufficient detail, the procedures they intend to employ to ensure item security. For example, if particular types of item-monitoring procedures are to be used in order to maintain security of the item pool, those procedures should be explicated in a bidder's proposal. Any type of confidentiality form that bidders will ask MRC members to sign should also be described.*

## **Requirement 8**

**A state must ensure that educators receive professional development focused on how to optimize children’s learning based on the results of instructionally supportive assessments.**

### **Illustrative RFP-Language**

This requirement, which ties professional development to the state’s testing program, is intended to remind bidders that test information needs to be acted on, not left simply as a report to schools, school districts, and the state. The DOE intends for professional development to be an integral part of the state’s testing program so that assessment results inform instructional improvement.

Suitable professional development involves educators in the design, implementation, and continuous improvement of a variety of strategies that result in building knowledge and skills to improve instruction in schools and classrooms. Therefore, the DOE anticipates development of an assessment-related professional development system that includes components such as: [e.g., linkages among practitioners within content or subject areas; best practice networks; collaborative analyses by educators of exemplary videotaped lessons; availability of training and support on a variety of effective instructional and assessment strategies.]

Although it will be the DOE’s responsibility to coordinate these assessment-related professional development activities, the DOE fully expects bidders to cooperate in its conceptualization of these professional development activities. Moreover, the DOE expects the contractor to provide assessment-relevant insights about how to make the professional development program most effective.

Therefore, bidders must describe the kinds of contributions dealing with assessment and/or instruction that they plan to make to the development and, possibly, delivery of professional development activities to support the state and optional classroom assessments they are developing in response to this RFP.

Among the bidders’ suggestions should be a demonstration of how educators can use disaggregated student performance data to design differentiated instructional strategies so that all subgroups of students are achieving at high levels. For example, bidders may demonstrate how reports of assessment results for districts, schools, and individual students can be used in instructional planning. Similarly, bidders may describe how they will develop a training package to assist educators in using assessment data for instructional decision making.

## **Requirement 9**

**A state should secure evidence that supports the ongoing improvement of its state assessments to ensure those assessments are (a) appropriate for the accountability purposes for which they are used, (b) appropriate for determining whether students have attained state standards, (c) appropriate for enhancing instruction, and (d) not the cause of negative consequences.**

### **Illustrative RFP-Language**

Because the requirements in this RFP specify state tests and optional classroom assessments that will change the nature of the state's assessment program, the DOE is interested in receiving bids from contractors to evaluate the assessment program.

Therefore, successful bidders must function independently from any test contractor(s) involved in the development of state tests and optional classroom assessments, and in the implementation of the assessment system. In addition, these independent evaluators must have expertise and experience in designing and conducting program evaluations, reporting results, and using results for program improvement. Bidders must describe the formative and summative evaluation procedures they will use to identify strategies for continually improving the state's assessment system.

The evaluative procedures bidders employ must provide the DOE with information about whether state assessments are:

- (a) appropriate for the accountability purposes for which they are used;
- (b) appropriate for determining students' attainment of the state's content standards;
- (c) sensitive to instructional quality; and
- (d) not the cause of negative consequences for students, educators, schools, and school districts.

To satisfy this section of the RFP, bidders must include in their proposals:

- (a) initial evaluation questions linked to the primary four areas of interest above;
- (b) initial data gathering strategies, including qualitative and quantitative strategies as appropriate;
- (c) initial analysis strategies in each of the primary four areas of interest, including a description of stakeholder involvement as meaning is derived from the data; and
- (d) proposed interim and final report formats and strategies for building consensus for recommendations to change the system.

In addition, because the evaluation activities will be conducted in collaboration with the DOE, bidders must describe the nature of any expectations regarding the DOE's participation in evaluation design and implementation.

## **A Final Reminder**

This report is designed to assist states that may choose to issue RFPs that solicit contractors to assist in the development of state tests and optional classroom systems advocated by the Commission on Instructionally Supportive Assessment. The RFP-language provided here illustrates ways in which an RFP might be worded. It should not be concluded that the Commission's members, either individually or collectively, are recommending one way for states to proceed.