

**ILLUSTRATIVE LANGUAGE  
FOR A STATE-LEVEL REQUEST FOR PROPOSALS  
TO DEVELOP INSTRUCTIONALLY SUPPORTIVE ACCOUNTABILITY TESTS  
FOR THE NO CHILD LEFT BEHIND ACT**

*(This illustrative language, regarding hypothetical "State X," is provided for those who are faced with the need to produce or evaluate state-developed accountability tests. Permission is hereby granted to copy, adopt or adapt any materials contained herein. W. James Popham, April 2004.)*

State X wishes to secure assistance to develop a statewide student assessment system that will not only satisfy federal and state requirements for educational accountability, but will also serve as a potent stimulus to improve educational quality within the State X public schools. This request for proposals (RFP) is focused on the creation of a statewide assessment system, the State X Testing System (SXTS), that will supply much, but not all, of the data needed in order to make a statewide educational accountability system function effectively. Because the purpose of the new SXTS tests will be to both contribute to a newly established accountability system and, at the same time, to support instructional improvement, the SXTS assessments can accurately be characterized as *instructionally supportive accountability tests*.

**I. Instructionally Supportive Accountability Tests**

*A modest number of high-import curricular aims.* There are three dominant elements that will distinguish the new SXTS tests from many of the accountability tests found currently in other states. First, in each subject tested and at each grade level where the tests are to be administered, the tests will assess students' mastery of only a relatively modest number of curricular aims (that is, cognitive skills or bodies of knowledge), but these assessed curricular aims must be particularly significant. In this way the state's educators can focus their instructional energies on promoting students' mastery of those powerful curricular aims. The SXTS assessment targets will be demonstrably derivative from the State X Content Standards, but will coalesce a number of those content standards so that State X educators will not be overwhelmed by the instructional need to pursue too many curricular aims. If smaller-than-usual numbers of curricular aims are to be assessed by SXTS, then it is apparent that those curricular aims must be potent indeed. The curricular aims to be assessed by our state's new accountability tests must represent the truly pivotal skills and knowledge that our state's children need in order to succeed in school and in life.

*Teacher-friendly assessment descriptions.* Second, instructionally supportive SXTS tests must be accompanied by clear, teacher-palatable descriptions of each curricular aim to be assessed so that the state's teachers can truly understand the nature of the powerful curricular aims their students are to achieve. Great care must be devoted to the generation of these assessment descriptions so that busy teachers will be willing to

read them. The assessment descriptions must be as brief as possible, yet sufficiently detailed to provide a reasonable picture of the nature of the particular skill or body of knowledge being assessed. Usually, one or more examples of appropriate test items will accompany each description. The purpose of these descriptions is to clarify satisfactorily the nature of each curricular aim assessed by a given SXTS test so teachers can direct instruction toward their students' generalizable mastery of that important curricular aim rather than toward students' answering correctly the items on a particular test. Great care will need to be given to the fashioning of these assessment descriptions, for their function in making SXTS an instructionally supportive intervention will be crucial.

*Instructionally useful reporting of results.* Finally, an instructionally supportive SXTS test will supply results of students' test performances in such a manner that the system's accountability *and* instructional needs will both be satisfied. This means that the kinds of state-level, district-level, and school-level results called for by the No Child Left Behind Act (NCLB) will be made available in a timely manner, both for total groups of students and for NCLB-designated disaggregation of results for student sub-groups. That is, the proportion of students whose scores reflect proficient-or-above levels of achievement will be reported according to performance levels established for the new assessment system. Care must be taken to assure federal officials that these new performance expectations will not be watered-down achievement levels and will, in fact, be based heavily—but surely not exclusively—on the levels of proficiency already approved for State X by the USDOE. The selected contractor must provide a documented analysis attesting to the relationship between the newly established performance levels and the earlier performance levels set for the NCLB exams previously used in State X.

In addition to the data-reports needed for accountability purposes, however, an instructionally supportive SXTS must provide each teacher, each student, and each student's parents with a report indicating whether *each curricular aim assessed by the test has or hasn't been mastered*. For example, suppose at grade six it had been determined that there were five powerful, well-described curricular aims in mathematics to be tested by a spring SXTS test. Well, when test results were made available in July of that year, an indication of a student's attainment of *each* of those five curricular aims would be provided to a school so that the school's officials could, in return, relay this information (if they chose to do so) to the student and to the student's parents.

This curricular-aim reporting of results would also be provided to teachers. However, to make it easier for teachers to determine which parts of their own instruction had been effective, a separate per-class summary must also be provided describing the proportions of a particular teacher's students who had mastered each assessed curricular aim. These class-level reports of students' mastery of every curricular aim assessed, made available to the state's teachers during the summer of each year, will permit teachers to decide whether to alter certain parts of their instructional activities during the upcoming school year. Although such outcome-mastery differentiated reports will surely satisfy the NCLB call for diagnostic achievement assessments, these reports

will also prove particularly helpful to State X teachers as they systematically attempt to strengthen the caliber of their own instructional activities.

Care must be taken to ensure that this kind of per-class reporting will be used exclusively for instructional improvement on the part of individual teachers—and will not be used as data for any kind of teacher appraisal system. Clearly, differences in the composition of a teacher's students from year to year would render such teacher-focused evaluations unsound. However, to make it more likely that teachers will use these reports to benefit their students, there will be an explicitly stated prohibition issued by the State X Department of Education (DOE) forbidding the use of such data to evaluate individual teachers. Teachers need to feel comfortable in discovering what segments of their instruction aren't working well—so that those segments can be improved. The specter of using class-level SXTS results for teacher-evaluation purposes is likely to reduce the degree to which many teachers would rely on such data for purposes of instructional improvement.

*A distinctive assessment system.* This, then, is the purpose of this RFP. We wish to solicit the support-services of a contractor capable of devising a distinctive genre of instructionally beneficial state-level assessments. The kinds of accountability tests called for in this instance are definitely not “same-old, same-old” statewide accountability tests. Rather, from genesis, they are intended to satisfy the demands of educational accountability as well as a mandate to improve State X education. We are searching for an innovative test-development contractor capable of assisting us in this exciting move toward the creation of accountability tests that will demonstrably improve instruction in State X schools. Although we have outlined much of our current thinking in this RFP, we look forward to a productive, problem-solving collaboration with members of a test-development firm who recognizes that the long-term viability of large-scale assessment in this nation depends on the creation of tests that will fulfill an accountability function while, at the same time, will nurture educational improvement.

## **II. Identification of the Curricular Aims to be Assessed by SXTS**

Perhaps the most important task of this entire test-development undertaking arises early in the project, namely, the isolation of the skills and bodies of knowledge to be assessed at each grade level and in each subject tested. In order that sufficient attention will be given to this critical task, the work on the isolation of the curricular aims that will serve as SXTS assessment foci must be initiated immediately upon the approval of a selected contractor. It is foreseen that committees of State X educators must reach closure on these key curricular choices during Month X 2004.

*The need for coalescence.* The essential process to be followed in deciding on the curricular aims to be tested in each subject and at each grade level is (1) to review the curricular aims currently embodied in the existing State X Content Standards, then (2) attempt to coalesce the most important of these extant aims into a smaller number of significant, instructionally addressable, broad yet measurable targets. For each curricular aim to be assessed, it must be possible to determine with reasonable

accuracy (*reasonable* rather than unflawed accuracy) whether a given student has mastered that curricular aim. Moreover, because more than a few test items will typically be needed to arrive at such student-specific inferences about students' curricular-aim mastery, it is apparent that *fewer* rather than *more* curricular aims must be chosen as assessment foci for SXTS.

However, given the extensive curricular thinking that has already gone into the isolation of the State X Content Standards, it is clearly appropriate to use those curricular aims as the point of departure for the delineation of the new set of curricular aims serving as the basis for SXTS. Thus, the selected contractor will first engage the advisory/analytic services of two or three nationally recognized curricular specialists in reading, in writing, and in mathematics from contractor-submitted (but DOE approved) nominees. These national experts will, then, collaborate with a small number of State X curricular experts in each of the three subject areas to be assessed. It would be the task of these half-dozen or so individuals (per subject) possibly working together electronically, by conference-call, video-conference, or in-person, to come up with a set of potentially assessable curricular aims that then could be reviewed and, as necessary, revised by committees of 25-30 State X educators (the three SXTS Content Committees) during a two- or three-day meeting (for each subject) in Month X 2004. Emerging from these meetings must be the limited number of significant curricular aims that will serve as the SXTS assessment targets at each grade where students are assessed by NCLB tests and in each assessed subject, that is, in reading, writing, and mathematics.

It is imperative for each curricular aim chosen to represent a *teachable* skill or body of knowledge. This in no way suggests that the curricular aims selected for SXTS should not be challenging. Indeed, quite the contrary should be the case. The modest number of curricular aims measured by the SXTS tests should be truly worthwhile. However, the individuals who are serving as the architects of those curricular/assessment targets should always be "thinking instructionally." The curricular aims sought must be conceptualized in a manner that will permit teachers to do an effective job in getting their students to achieve those aims.

*Building a solid case for content-related evidence of validity.* When the SXTS Content Committees have finally arrived at a set of curricular aims that they wish to be assessed, it is important to collect evidence from committee members related to the degree to which the newly developed tests will satisfactorily measure the skills and knowledge that those tests ought to measure. Therefore, committee members should review carefully the State X Content Standards and then, considering the committee-identified curricular aims that are to be assessed by SXTS, respond individually to a contractor-developed inventory that, most likely at each grade-level assessed, would ask a question somewhat along these lines: "In reviewing the curricular aims that your committee has just identified for SXTS tests in relationship to the State X Content Standards, please circle the percent that best answers the following question: *What percent of the most important curricular aims present in the relevant State X Content Standards do you believe will be assessed by the SXTS?* (100%, 90%, etc.) These

responses should be summarized and included in a contractor-created report describing content-related validity evidence for the SXTS tests.

On that same point, because the original identification of the State X Content Standards is so important in building a content-related evidence trail, the contractor must include a detailed description in such a report regarding how and when the State X Content Standards were identified and by whom.

The contractor is to be responsible for convening the three SXTS Content Committees (reading, writing, and mathematics) and moderating all meetings of these committees—although DOE personnel and, if possible, a few of the national consultants who had previously identified potential curricular targets would play active roles in the conduct of these sessions. The contractor will need to prepare and distribute copies of the proposed curricular aims (the ones generated earlier by the half-dozen curricular consultants) for committee members—ideally in advance of the Month X meetings. The contractor must also propose a sensible way of securing committee agreement (e.g., relying on ranked curricular choices) and documenting these choices in a manner suitable for the contractor’s preparation of a report describing content-related evidence of validity underlying SXTS.

From the earliest initiation of the test-development work, the contractor must carefully document each step in the creation of the SXTS tests so that a technically sound report describing content-related evidence of validity for the tests can be submitted to DOE. This report should be delivered to DOE no later than Month X 2005.

If a contractor wishes to propose and defend an alternative strategy for isolating a small number of unarguably significant curricular aims to be assessed by SXTS, we will be pleased to consider such alternatives.

### **III. Generation of Assessment Descriptions**

Immediately upon determination of the SXTS assessment targets, the contractor will prepare descriptions of the skills and/or bodies of knowledge to be measured. These descriptions, although aimed chiefly at the teachers who will be promoting students’ mastery of the curricular targets being described, will most likely also be suitable for communicating with parents and other concerned citizens regarding what is going to be measured by SXTS in a given grade and subject. These will be particularly important descriptive documents, hence the contractor must develop them with consummate care, including a small-scale tryout of each assessment description that is being developed.

*An assessment description’s quality.* Two criteria must be satisfied by each assessment description. First, the description must communicate with sufficient clarity so that different teachers (and, possibly, others) will arrive at essentially homogeneous interpretations of the nature of the skill or body of knowledge to be assessed. Second, the description must be sufficiently concise and well written so that busy teachers will be willing to read it. These two criteria, clarity and conciseness, will sometimes be in

conflict. Clearly, one could produce a lengthy, remarkably circumscribing assessment description of a to-be-assessed curricular aim that, because of the description's inordinate detail, would most likely yield similar interpretations by all teachers who actually had the patience to read through that hyper-extensive document. But such a lengthy assessment description, of course, would rarely be read by real-world teachers. Similarly, a remarkably terse assessment description of only a few sentences and a single sample item might be read by most teachers, but those teachers' subsequent interpretations regarding what was to be assessed would, in all likelihood, range wildly.

The task of the contractor in this instance is to strike the most defensible balance between satisfying these two criteria. Considerable judgment and verbal artistry will be required to complete this task satisfactorily. The SXTS assessment descriptions, each depicting one of the limited number of important skills or bodies of knowledge being assessed in a particular subject at a given grade level, will function as the communication cornerstones of an assessment system intended to improve instructional quality in State X.

*Shake-down tryouts.* In order to make more certain that the resulting assessment descriptions are suitable, the contractor should conduct one or more shake-down tryouts of each near-final description with a minimum of 10 teachers reviewing each description. The teachers should be asked to read an assessment description silently, then individually describe in writing (insofar as possible, "in your own words") what the teacher believes the essence of the to-be-assessed skill or body of knowledge actually is. (These independently authored interpretations should subsequently be evaluated to ascertain the degree of homogeneous interpretations they represent.)

Shake-down teachers should also be asked to complete an anonymous appraisal of each assessment description according to the likelihood that real-world teachers (their colleagues) would actually be willing to read it, along with being obliged to read several other similarly written assessment descriptions. The structure of these brief, anonymous questionnaires would be approved by DOE, but the contractor should devise a brief, carefully worded questionnaire that gets at the teacher-palatability of each description. Obviously if a given assessment description fails to yield reasonably homogeneous interpretations from most teachers who take part in the shake-down trial, or is regarded as unlikely to be read by busy teachers, then that assessment description must be revised in order to remedy its shortcomings. A brief report describing the results of this shake-down activity should be prepared by the contractor and submitted to DOE soon after the completion of the assessment descriptions.

#### **IV. Development of SXTS Test Items**

*Assessment descriptions as test-item specifications.* The assessment descriptions that set forth what is to be assessed each year by SXTS tests will play a significant role in the contractor's generation of test items. Therefore, the creation of the SXTS items must, in the main, await the completion and DOE-approval of the assessment descriptions. In a very real sense, the assessment descriptions will function as *test-item*

*specifications* that will guide the contractor's item-writers. The contractor, in addition, must prepare a set of *test specifications* for every test—indicating the overall number of items of different types per assessed curricular aim. These test specifications are, of course, distinguishable from the test-item specifications represented by the assessment descriptions.

In essence, the assessment descriptions will communicate the nature of the particular cognitive skill or body of knowledge that is to be assessed by specific types of test items. Thus, when the contractor's item-generation staff produces such items, it must be recognized that students' responses to a set of like-purpose items should yield a reasonably accurate inference about whether the student has or hasn't mastered the skill or body of knowledge being described. In fact, as will be seen shortly, a major factor in the appraisal of each SXTS test items will be judgments regarding its contribution to such mastery/non-mastery inferences.

The test specifications themselves need to be developed in collaboration with the SXTS Content Committees and others who will determine which curricular aims are actually to be assessed on a given year's test. For instance, if the time-limitations for assessment literally preclude the assessment of as many curricular aims as the Content Committees would like, then this needs to be known *early on* so that, if necessary, the curricular aims to be assessed can be reconsidered.

*The tests needed.* SXTS items are needed for tests to be administered in grades 3-8 and once in grades 10-12 in reading, writing, and mathematics. It is foreseen that, for the reading and mathematics tests, the items will be dominantly selected-response in nature (chiefly of the multiple-choice variety) or short-answer constructed-response items. For the writing tests, students will be asked to generate an original writing sample.

We are eager to carry out this important SXTS assessment program in a technically sound and educationally defensible, but cost-efficient manner. This translates into a desire to create SXTS tests that, at the most affordable cost, can yield the important sorts of inferences about students' mastery that we seek.

Two instruction-related measurement considerations should be mentioned at this time. First, we wish to install an assessment system that will encourage the state's teachers to promote students' *generalizable* mastery of the skills and knowledge being assessed. A test consisting exclusively of multiple-choice items would surely not serve this purpose well. We want our students taught so that they can apply what they have learned when responding to a variety of stimuli, not only to one type of test-item. Such generalized mastery of skills and knowledge is more apt to transfer to subsequent applications in both school and non-school settings. Therefore, even though the scoring of short-answer items is clearly more costly than is the scoring of selected-response items, we wish to include at least a modest number of items that require human rather than electronic scoring. However, a second consideration is that we wish to take up as little *instructional* time with SXTS testing as possible. Ideally, the amount of testing time

would be reduced to an absolute minimum so that the state's teachers would have more classroom time to teach, not test. Yet, because students may be asked on the SXTS tests to display their skills and/or knowledge in response to *either* selected-response or constructed-response items, teachers will be more likely to use their instructional time to promote students' generalizable mastery of SXTS-measured curricular aims.

Accordingly, there are to be SXTS tests that, at maximum, take no more than two consecutive 50-minute administrations per year in mathematics, no more than two consecutive 50-minute administrations per year in reading, and no more than two consecutive 50-minute administrations in writing (as will be seen later in the RFP, we foresee that there will be two writing samples collected from students each year at times other than when the statewide reading and mathematics tests are administered). If it is possible for the contractor to demonstrate that the skills and knowledge to be assessed in a given year can be measured with reasonable accuracy in less time than suggested by these maxima, we will be delighted to consider the adoption of shorter testing periods.

When the actual item-production begins, exclusively under the auspices of the contractor (This does not preclude the hiring of State X educators to assist in item-generation, although there will be no special evaluative advantage given to proposers who intend to do so.), those individuals producing the items should be guided chiefly by the assessment descriptions previously indicated *and by the nature of the upcoming item reviews* to be described later in the RFP. The contractor must create a sufficient number of items so that, subsequent to item reviews by committees of State X educators, there will be an adequate number of items to permit (1) a full-scale statewide field-test of the entire SXTS system in 2005 which would also permit the inclusion of subsets of embedded "field-test" items for use in subsequent test administrations [We anticipate employing routine embedded field-tests of a limited number of new items each year in future SXTS administrations.]; (2) an item-coverage of at least 25 percent so that, if there are deficits in certain categories of items, the extra items can be used in their stead; and (3) enough items to produce two sets of *mini-tests* that will be provided to teachers as optional classroom assessments employable as progress-monitoring assessments during teachers' ongoing classroom instruction.

*Pilot-testing sample items.* We will view with favor the submission of proposals that include some modest pilot-testing of at least a sample of the items intended to measure each curricular aim to be assessed by SXTS. These early-on pilot tests might involve 5-10 students per item-type. The students would first be asked to complete a small number of items reflecting mastery of a particular skill or body of knowledge. Then, having completed the items, the students will be directed to consider their responses while a member of the contractor's staff goes through all items (aloud) with the students in a discussion-based attempt to discern the viability of the assessment tactics being employed in those items. Although not intended to be a significantly expensive undertaking, we are looking for contractors who are constantly willing to determine that their assessment approaches are making sense to children and, therefore, are more apt to yield valid inferences about children's mastery of a particular skill or body of

knowledge. We suggest that this pilot testing be priced separately as a contract option. State X may or may not choose to adopt it. However, if the additional cost and the description of the proposed activity seem suitable, we will definitely be inclined to adopt such a procedure.

*Mini-tests.* Some clarification is necessary regarding our expectations for the previously mentioned mini-tests. We wish SXTS to improve schooling in our state. For teachers to determine if their students are moving satisfactorily toward mastery of the modest number of significant curricular aims assessed by SXTS, it will be beneficial if DOE provides *optional* classroom assessments that can be embedded in the teacher's ongoing instruction in a manner not unlike teachers' traditional use of instructional "units" and, at the conclusion of a unit, the administration of a "unit test" well in advance of a "final" examination for the entire course or subject. Thus, the contractor is to prepare two complete sets of these mini-tests, each separate mini-test focused on a particular SXTS-assessed curricular aim. These mini-tests will be provided by DOE to teachers at each grade level and each subject for which SXTS tests are to be administered. Teachers, then, can choose if and when these "formative assessments" would be administered. Perhaps in the future, this sort of embedded assessment can provide supplemental accountability evidence that, along with the data from the summative statewide SXTS assessments, can contribute to State X's overall accountability plan. However, at this moment the exclusive role of these embedded reading and mathematics tests will be to provide the state's teachers with assessment tools whereby those teachers can more effectively (and more efficiently) monitor their students' progress toward the mastery of SXTS-assessed skills and knowledge.

*Writing-assessment considerations.* Turning to the assessment of students' writing prowess, it is our intention to require students to create an original writing-sample on two occasions during the school year. We currently contemplate that the first of these two samples will be collected via local administration (not dramatically different, in many ways, from how a teacher might use the optional DOE-provided mini-tests in reading and mathematics) in January and April of a school year. Each writing-sample (focused on different sorts of writing at different grade levels as determined by the SXTS Writing Content Committees and as explicated in each grade-specific assessment description for writing) would be composed by students in no more than 50 minutes and would either be scored (1) at the district-level—with subsequent audit-scoring of 10 percent of these writing samples by a state-level audit committee during scoring sessions conducted by the contractor or (2) at the state level by State X educators who will be scoring all responses under the overall direction of the contractor.

Proposers should describe their scoring preference and defend the merits of their preferred approach for the scoring of students' writing-samples. Although the writing-samples are to be scored holistically, based on rubrics whose evaluative criteria must be instructionally addressable, proposers should suggest procedures that they believe would make this twice-a-year formal assessment of students' writing most supportive of instructional improvement. Proposers, of course, must indicate the procedures they

intend to employ in order to create a defensible degree of equi-difficulty in the writing prompts (items) that are to be used in SXTS.

As was true with the test items in reading and mathematics, the contractor must supply two additional items (prompts) per grade level so that, again optionally, the state's teachers could formatively assess their students' achievement in writing during the school year with those mini-tests. Such embedded assessments, ideally seen by students as part of the teacher's regular instruction and classroom assessment activities, will allow teachers to adjust their instruction in order to address assessment-identified deficits in students' writing skills. These mini-tests in writing, of course, must assess the grade-specific genre of writing that had been established by the SXTS Writing Content Committees.

As indicated previously, the contractor must develop items in full anticipation that those items will be rigorously reviewed by State X educators who have been directed to scrutinize the items from a content and potential bias perspective. We turn, then, to these item reviews.

## **V. Review of SXTS Items**

*Two evaluative perspectives.* All items will be reviewed by an appropriate SXTS Content Committee as well as the SXTS Bias-Review Committee during in-person item reviews. (On-line item-reviews will not be acceptable.) These review sessions will be coordinated and moderated by the contractor at sites and times to be determined by the DOE in consultation with the contractor. Each SXTS Content Committee will consist of 25-30 members and the SXTS Bias-Review Committee will also consist of 25-30 members. These review committees will be comparable in composition to similar committees used in the development of statewide tests throughout the U.S. However, the nature of the item-review criteria that will be employed when these committee members scrutinize items will be more stringent than those usually encountered in other settings. The membership of all SXTS committees will be determined by DOE.

*Content Committees.* During a 2-3 day meeting of each of the three Content Committees (reading, writing, and mathematics), every contractor-submitted item will be judged according to each committee member's independent answers to the following questions:

1. Contribution to a mastery/non-mastery inference. Will a student's response to this item contribute to an accurate inference about the student's mastery of the assessment description's curricular aim which, according to the test's developers, this item measures? (Yes/No/Not Sure)
2. Instructional sensitivity. If a State X teacher attempts, with reasonable effectiveness, to promote students' mastery of the curricular aim being assessed by this item, is it likely that a substantial majority of the teacher's students will be able to answer the item correctly? (Yes/ No/Not Sure)

Some clarification is probably required regarding the nature of “reasonable effectiveness” and how it should function in this sort of item-review question. We are thinking here of a teacher who does an average-or-better instructional job. We are looking for instructionally sensitive items so that, after having been taught by such an average-or-better teacher who has focused on the curricular aim that the item is intended to help measure, a substantial majority of the teacher’s students will then answer the item correctly. If proposers wish to specify a particular percent of the class who should answer correctly, e.g., 75%, they may do so. (The contractor, of course, would need to make sure all item-reviewers understand what is meant by “reasonable effectiveness.”)

Then, after all of the items for a particular assessment description have been reviewed by committee members, each member should be asked to consider the *entire* set of the items linked to a particular assessment description, then answer the following question for that set of items:

3. Curricular coverage. Assume that a random selection of (Here, the contractor and the DOE will have decided how many items of this type are needed to assess a student’s mastery of this particular curricular aim. So, let’s assume for illustrative purposes in this RFP, that the number is 10 items.) 10 of this particular type of items will be used on a given SXTS test to assess a student’s mastery of the curricular aim set forth in the item’s assessment description, how satisfactorily will the coverage of those 10 items reflect the student’s mastery of this curricular aim? (Very Well/Satisfactorily/ Not Very Well)

The contractor should assemble and distribute sets of 10 DOE-approved illustrative items for each curricular aim to help reviewers arrive at the stipulated curricular coverage judgments for each assessed curricular aim.

The particular language and response-options employed in these review questions can be modified with DOE approval, so the contractor may wish to propose alterations in the wording. However, the thrust of the item reviews should be maintained.

Results of all item reviews by the SXTS Content Committees must be summarized and well-documented in a report submitted to DOE by the contractor. This report will play a prominent role regarding content-related evidence of validity for the SXTS tests.

*The Bias-Review Committee.* The SXTS Bias-Review Committee will scrutinize each item on the basis of its members’ individual responses (on a contractor-supplied form) to the following question:

1. Item bias. Might this item offend or unfairly penalize any group of students on the basis of personal characteristics such as race, gender, or ethnicity? (Yes/No/Not Sure)

Then, when an entire set of items for a particular subject and grade level have been reviewed, members of the SXTS Bias-Review Committee should be asked to respond to a question pertaining to the entire set of items:

2. Item-group bias. Might this entire set of items, as a collective group, offend or unfairly penalize any group of students on the basis of personal characteristics such as race, gender, or ethnicity? (Yes/No/Not Sure)

As with the Content Committees, results of all Bias-Review Committee judgments should be carefully documented and supplied to DOE in a timely report.

It is assumed that the contractor will moderate all meetings of these in-person committees and that, at the beginning of each set of item-review meetings, there will be an appropriate orientation supplied to committee members regarding the specifics of the item-review questions to which they will soon be responding. Practice items and discussions of committee members' judgments regarding these practice items should be included in the orientation activities—along with ample opportunities for committee members to seek clarification. We want our item-reviewers to understand their upcoming judgmental tasks thoroughly and to feel comfortable when carrying them out. Results of the Bias-Review Committee's judgments must be summarized and reported to DOE in a timely fashion.

## **VI. Field Testing SXTS Items**

The contractor will arrange and carry out a separate field-test of all committee-approved items in a set of DOE-designated schools during Month X 2005. Although, because of the installation of a brand new SXTS, in this instance a separate field-test is required, we intend in the future to employ embedded field-testing wherein the final section on six sub-forms of each year's administered tests contains form-embedded field-test items. Accordingly, we wish the contractor to incorporate that same field-test approach during this initial, separate field-test. The contractor must, however, propose how many sub-forms and how many students must be involved to provide suitable evidence of item quality and, at the same time, supply at least some additional items for the SXTS pools (via the sub-form embedding).

For this field-test, because we wish to make it as cost-effective as possible, we intend for the items administered to all students (as opposed to those administered as part of the embedded sub-form items) to be administered to sufficient numbers of students so that differential item functioning (DIF) statistics can be computed on each item. These DIF statistics are to be determined on the basis of gender differences as well as differences among the performances of African-American, Hispanic-American, and white student groups.

## VII. Reporting SXTS Results

There are two major purposes for SXTS: (1) providing reliable information from which valid inferences may be drawn for the purposes of school and district accountability and (2) providing teachers, parents, and students with information that is supportive of the instructional process. In service of those purposes, contractors must develop a reporting system that includes, as a minimum, the types of reports described below. The proposed reports must be designed in a manner that facilitates their understanding and use by students, parents, teachers, and school administrators in: (a) providing students with instruction that promotes mastery of the curricular aims; and (b) improving instructional practice and in evaluating the overall effectiveness of the SXTS educational program in its schools and districts. The reports are not intended, nor should they be used, for the evaluation of individual teachers.

Although the following listing of reports contains an indication of the type of information that will be included in each report, the final determination of content and format will be made by DOE staff in consultation with teachers, other school district personnel, and the contractor. As noted earlier, contractors must provide for the active inclusion of teachers in the review of all proposed reports of assessment results.

1. *Parent/Home*. This report should include numerical, graphic, and narrative information that provides an indication of the student's overall performance and clearly explains the student's achievement level on each of the curricular aims that are included in the SXTS tests. The report should indicate relative strengths and weaknesses in relation to *each* assessed curricular aim and should be presented in non-technical language that will make the content of the report accessible to most parents/guardians. A graphic indicating the overall score, with a 95 percent confidence band, must be included. The report must also include definitions of the "below basic," "basic," "proficient," and "advanced" achievement levels.

Because the curricular aims that are included in the assessment might, depending on the results of the work of the SXTS Content Committees, vary somewhat from year-to-year, the design of the report must be flexible enough to accommodate changes in that content.

2. *Student*. This report will contain basically the same information as the Parent/Home report, but is intended for use by teachers when working with individual students.
3. *Class Roster*. This report will be a roster of all students in a class and will include each student's overall score and performance on each of the curricular aims that are included in the assessment.
4. *Below-Basic/Basic Roster*. This report will identify all students whose overall achievement level is indicated as either below-basic or basic, but less than

proficient, and will identify, for each student, the specific curricular aim(s) on which the student needs additional instruction. The below-basic and basic students shall be listed in separate sections of the report.

4. *Class Summary.* This report will include summary information, e.g., the numbers and percentages of students in each of the achievement categories, e.g., basic, and the numbers and percentages of students demonstrating proficiency for each of the assessed curricular aims, for the class as a whole, and for each sub-group that is identified for the purpose of NCLB accountability and reporting.

This class-summary report is seen by DOE as the pivotal document by which teachers learn whether their instructional activities have been successful. Accordingly, substantial tryout activity (with State X teachers) is foreseen so that a truly appropriate level of detail per assessed curricular aim will be presented. We dare not overwhelm teachers with too much information on these reports.

6. *Item Response Summary.* If DOE decides to release test items or test forms following each administration, the contractor will develop item-response summaries that provide information such as the following: (a) numbers and percentages of students selecting each option for selected-response items and (b) numbers and percentages of students at each score-point for constructed-response items. This report, if developed, will be provided for each class within a school and as a grade-by-grade-summary for the school as a whole.
7. *School Summary.* This report will provide class-by-class summary information for all classes within a school at each of the grade level(s) included in the assessment. Information will include the number and percentages of students classified in each of the four achievement levels and the number and percentages of students demonstrating proficiency on each of the curricular aims included in the assessment. In addition to the class-by-class summary, the same information shall be provided grade-by-grade for the school as a whole and for each sub-group identified for the purposes on NCLB reporting and accountability.
8. *School Fall Assignment.* As a result of the various grade-level organizational patterns of schools, some students who are tested in the spring will be enrolling in a different school within the district the following fall. To assist teachers (in the receiving school) meet the instructional needs of these students, the contractor will develop Fall Assignment reports for each of the receiving schools. The reports must provide, for each student, the student's name, overall achievement level, and an indication of mastery/non-mastery for each curricular aim included in the SXTS tests. The information

necessary for identifying these students and generating the reports must be collected at the time of test administration.

9. *District Summary.* This report will provide school-by-school summary information for all schools within a district at each of the grade levels included in the SXTS assessment. Information will include the number and percentages of students classified in each of the four achievement levels for each of the curricular aims measured in SXTS. In addition to the school-by-school summary, the same information shall be provided by grade for the district as a whole and for each sub-group identified for the purposes on NCLB reporting and accountability.
10. *State Summary.* This report will provide the same information called for in the District Summary, but will do so for all State districts and schools.

### **VIII. Instructional Suggestions**

The dual-mission of SXTS is (1) to provide accountability evidence by which State X citizens, and particularly parents, can judge the effectiveness of particular schools and school districts and (2) improve the nature of classroom instruction provided to the state's students. In connection with this second mission, it will be helpful to many State X teachers if a set of suggested instructional tactics could be provided for promoting students' mastery of the SXTS-assessed curricular aims. Accordingly, when SXTS Content Committees meet, and have sufficient time, a set of suggestions regarding promising instructional approaches should be collected, refined, and then disseminated to the state's educators. The instructional ideas proffered in these relatively brief *Teaching Possibilities Booklets* would clearly be suggestions, not mandates. Different teachers can employ markedly different instructional approaches, yet promote similar achievement levels for their students. However, the release of a series of DOE-collected instructional ideas would be greeted with enthusiasm by many State X teachers (especially by those teachers in most need of such instructional suggestions).

The contractor will be responsible for producing these booklets so that, no later than Month X of 2005—after the field-test of SXTS items, the state's teachers can garner some insights about instructional approaches related to the skills and knowledge recently assessed in their classrooms. Although many initial instructional suggestions can be secured from the SXTS Content Committees, additional insights should be collected via a short series of small-group sessions with groups of State X teachers—intended to expand and refine early versions of the instructional suggestions booklets. A formal series of tryouts of early version *Teaching Possibilities Booklets* is absolutely essential.

## **IX. Professional Development**

We recognize that the nature of the NCLB tests we are calling for in this RFP represent a marked departure from the types of accountability tests with which most State X teachers are familiar. Thus, in order for those teachers to derive optimal instructional dividends from these instructionally supportive accountability tests, a meaningful statewide professional development program for State X educators will be needed.

Although the bulk of these professional development activities will be provided by DOE personnel, we wish a contractor's proposal to include a description of what the contractor would regard as a cost-effective but potent staff-development program related to SXTS.

As we see SXTS professional development, it must be conceptualized at the very outset of the project and, as deemed suitable by DOE and the contractor, carried out from the test-development project's earliest days until the project's conclusion. We are not looking for last-minute tokenism here. Rather, we want SXTS professional development to be an enterprise that allow State X educators, and State X children, to obtain the greatest possible benefits from an instructionally supportive assessment program.