

MODEL CONTRACTOR STANDARDS & STATE RESPONSIBILITIES

for State Testing Programs

Arizona, Colorado, Florida,
Georgia, Maryland, Massachusetts,
Michigan, Ohio, Pennsylvania,
Tennessee and Texas

Developed by
AccountabilityWorks, Inc.



Contributors:
S. E. Phillips and Theodor Rebarber.
Review and input by **Richard W. Cross**
and **John H. Oswald.** Editorial assistance
by **Thomson W. McFarland.**

Introduction

State reforms to develop K-12 academic standards and to assess the performance of all students on these standards have resulted in a substantial increase in the number and scope of contracts with testing companies for statewide assessment programs. The *No Child Left Behind Act of 2001* (Public Law 107-110) requirement that all states test reading and mathematics in grades 3-8 will further increase the number and scope of statewide testing programs. Many of these assessments are high-stakes for students (e.g., graduation or grade promotion tests) and/or educators (e.g., accountability programs). With high school diplomas, monetary awards or federal funding for schools and school systems dependent on test results, it is imperative that state assessments be of high quality, meet professional standards for best practice, be delivered in a timely manner, and be scored accurately. With increasingly tight budgets, it is similarly imperative that assessment programs be developed and implemented in an efficient and cost effective manner without sacrificing quality.

Creating a high quality state testing program requires both cooperation and accountability. It recalls the arms control motto, “trust but verify.” The main participants in this relationship are state agency staff and test vendor staff. To support these efforts, the Education Leaders Council and AccountabilityWorks have assisted participating states in developing these

Model Contractor Standards and State Responsibilities for State Testing Programs to communicate more clearly today’s expectations for the development and administration of high-quality, efficient, and defensible, high-stakes state testing programs. This effort included an intensive Testing Summit held on February 20-21, 2002 in Austin, Texas, attended by state education and testing industry leaders, where participants provided input and suggestions regarding this document. (A list of attendees is included at the end of this document.) The ELC Summit was made possible by a grant from the U.S. Department of Education to the State Education Policy Network (SEPN). For vendors, commitment to following the “vendor standards” described herein can be cited as evidence of self-regulation and adherence to best practices. Of course, such standards are also designed for use by states in designing contractual relationships with vendors and in managing and overseeing those relationships. For states, the outlined “state responsibilities” are intended to provide a model for what is necessary to create a high quality testing program and to serve as guidelines for policymakers enacting reforms in state testing programs. Some of the state responsibilities also describe important requirements for legal defensibility of high-stakes components within a state testing program. In an environment that encourages litigation of disagreements, adherence to professional standards and best practices will substantially increase pre-

paredness and the likelihood of success when a high-stakes testing program faces legal challenge.

This document is divided into four sections: Preplanning, Development, Administration, and Uses of Data. The Preplanning Standards address antecedent activities and decisions critical to the production of a comprehensive Request for Proposal (RFP) describing the products and services the state wants a vendor to supply to its testing program. Assuming the criteria specified in the Preplanning responsibilities have been met, the Development and Administration sections provide guidelines for the specification of contract activities and execution of the contracted work. The Uses of Data section deals activities subsequent to the administration of the test, but directly connected with the interpretation of test data and score reports. Each section begins with a brief introduction that provides background and explanatory information.

Many of the standards herein have both a vendor and state corollary. Within each section, standards that are “linked” between state and vendor share a number, but are differentiated by the letter that follows the number, with “S” designating a state responsibility and “V” a standard for a vendor. Standards which are specific solely to the state or vendor come with the same designators, to show which of the parties bears responsibility.

The purpose of this document is to clarify ideal roles and obligations in a typical relationship between a state and vendor, either with respect to test development or scoring and administration. The “typical” relationship is assumed to be that the state, in response to action by its legislature, has developed academic standards and is contracting with one vendor to purchase a custom-built assessment based on its own academic standards and

with another vendor to handle test administration and scoring. Assumptions of some kind are clearly necessary for the design of a “model” document of this type. Of course, there are few “typical” states that precisely, or even nearly, mirror all of the arrangements assumed here. Among the many possible variations are:

- A state agency buys access to a commercially developed and published test.
- A state agency hires a vendor to develop a test that is then owned by the agency. The state requires the vendor to provide materials, scoring services, and reporting services.
- A state agency hires a vendor to develop tests that will be owned by the state and then hires a second vendor to do the test administration, scoring, and reporting activities.
- A state agency buys access to a test publisher by one vendor but then hires another vendor to administer, score, and report the results.
- A state agency develops the test with the assistance of local districts and state universities. It then hires a vendor to administer, score, and report the results.

Clearly, this document will apply differently in each of these different scenarios. It is hoped, however, that the model is sufficiently clear and well-defined that a state using a different arrangement would be able to determine the necessary adjustments in those sections that require adjustment. Even in cases where the relationship or tasks in a given state appear to fit perfectly with the model described here, this document is only intended to provide a framework to ensure that relevant issues are addressed. Important issues need to be resolved in a way that is consistent with a state’s political process. Evaluating the acceptability of a process or contract between a state and vendor should not rely on the lit-

eral satisfaction of every standard or responsibility in this document, nor can acceptability be determined through the use of a checklist. Further, while retaining decision-making authority, a state may benefit from seeking the advice of the vendor regarding alternative methods for satisfying a particular guideline. Similarly, a responsible vendor will seek the state's advice or feedback at every point along the way where important decisions must be made. Regardless of how roles are defined and tasks are delegated, states retain ultimate responsibility and authority for state testing programs.

Throughout this document, the terms “testing companies” and “industry” apply generically to refer to all providers of test content, printing, scoring, validation, and other testing-related services, whether for-profit or not-for-profit, public or private. The term “state” applies to the educational enterprise of the fifty states, territories, and other appropriate jurisdictions (such as the Department of Defense Dependent Schools) and includes state education agencies (SEAs), state boards of education, and other official educational entities.

1

Preplanning

For a state testing program to follow standards of best practice, several preconditions should be met before an RFP is developed, a contract is signed with a vendor, and test development begins. These preconditions are important in enabling a vendor to produce a quality test that will satisfy the state's expectations. These preconditions include actions to be taken by the state legislature (or other responsible entity) as well as the state agency with authority to implement the testing program. The purpose of these preconditions is to support production of an RFP that specifies in detail the services and products to be provided by the vendor given reasonable timelines and resources, to ensure that the state has knowledgeable and adequately trained staff competent to assign all required activities to either itself or the vendor, to ensure adequate planning and funding to produce a quality testing program, and to ensure that staff is in place to competently supervise the vendor relationship going forward.

1.1.S. The state legislature should enact reasonable timelines and provide adequate funding for the testing program. The legislation should also include: statements of purpose; designation of authority for important tasks (e.g., standard setting); responsibilities of the state agency and the local districts; types of reports of results; uses of the data (e.g., school or district accountability); contracting authority.

Comment: In general, the lead time for developing a new, high-stakes assessment includes a minimum of 38 months: 6 months planning, preparation, & development of test blueprint; 6 months item writing & editing; 6 months item tryouts & analyses; 6 months preparation of field test forms & supporting materials; 6 months field testing, research studies (e.g., opportunity to learn surveys in the case of high-stakes tests) & analyses; 6 months development of final test forms, edit supplementary materials, set passing standards, finalize security, accommodations, & reporting policies; 2 months administer final tests, score, equate, and report results. This timeline begins with the signing of a contract with a vendor and assumes that state content standards for the subjects being tested have already been adopted.

Preplanning activities leading to the development of a comprehensive RFP, vendor bid time, proposal evaluation, and negotiations for the award of a final contract will often add at least another 6 months. Unanticipated complications that often accompany implementation of a new testing program will also add additional time. Thus, legislation that creates a new testing program should allow approximately 4 years from the time of passage until the first live tests are administered. States with well-established testing programs and experienced staffs in place may

be able to reduce the time required to develop additional tests (i.e., development limited to certain grades not previously covered).

Three to four years of lead time is also consistent with legal requirements for an adequate notice period and opportunity to learn (OTL) for tests with high stakes for individual students. OTL requires sufficient time for implementation of curricula and instruction that provides an adequate opportunity for students to have been taught the tested content before their initial attempt to pass high-stakes tests. If a state has developed sufficiently clear academic standards, notice requirements for OTL may be triggered by the publication date of the standards if a strong communication effort is undertaken and the state can demonstrate that schools have aligned instruction with the standards. When offered opportunities for input, potential vendors should alert states to unreasonable timelines and propose alternatives reasonably calculated to meet professional standards for best practice.

In addition to providing sufficient development time, legislation for a new testing program must provide adequate funding for agency preplanning activities, test development, test administration, and other program activities necessary to produce a quality test for each of the mandated grades and subjects. The required activities may be conducted by the state or an outside vendor, but funding must be sufficient so that no important steps are left out.

Development costs do not depend on the number of students to be tested. Therefore, a small testing program with limited funds and a relatively small number of students over which to spread the cost may only be able to develop tests for fewer grades and/or subjects than larger testing programs. Options for state cooperation to improve efficiency and lower costs are described in the accompanying innovation priorities document.

1.2.S. Where custom tests are to be designed on the basis of state academic standards, special care should be taken to develop high quality standards that are rigorous, clear and specific, consistent with sound research on curriculum and instruction, and well-organized to ensure that the lower levels serve as a sound foundation for the upper levels.

Comment: While sound standards-based tests must be well aligned with state standards, the standards should be designed so that they serve that purpose well. Consensus-building within a state is important in order to develop support for broad implementation, but consensus does not always or necessarily lead to quality. Where, for example, consensus is reached as a result of agreement on broad or vague standards statements, schools may focus excessively on the test itself for guidance on what to teach; tests are typically not designed to carry such a heavy load, leading to criticisms that teachers are “narrowing the curriculum” to what is being tested. More detailed criteria and models for high-quality standards exist and have been identified by organizations such as the Thomas B. Fordham Foundation and others.

1.3.S. The state agency responsible for testing should include staff with adequate knowledge and training in psychometrics, curriculum, communication, policy, special populations, and contracting to develop a comprehensive RFP, to complete all necessary activities not assigned to the vendor, and—especially—to monitor vendor performance throughout the contract period.

Comment: Agency staff must play an active role in the development of a quality testing program. In order to decide which services and products should be included in an RFP, agency staff must thoroughly understand the test development process and the requirements for a psychometrically and legally defensible test. Where knowledge and training is inadequate or lacking, staff should seek training and assistance from outside experts. Agency staff must be prepared to complete all required steps and activities not specifically delegated by the state to a vendor and to competently monitor all contract activities. It is possible for a state agency to outsource some of the steps and activities required prior to the selection of the main test development vendor, though adequate expertise should always exist on staff to monitor the performance of all such additional vendors.

The staffing needs of the state agency to support a statewide assessment program are significant. This is going to be one of the more serious tasks confronting states as they implement P.L. 107-110. At the minimalist end of the continuum, a state could theoretically have one person be the assessment coordinator and simply allow the contractor to do everything. At the other end, a state agency can hire sufficient number of staff members to coordinate the work of multiple contractors, assume primary responsibility for quality control work, and provide data analysis and dissemination/training activities within the state. Of course it would be impractical, if not impossible, for a state to meet the requirements of P.L. 107-110 with a “minimalist” staff.

State legislatures do not have to create a lot of permanent positions for the state bureaucracy if the agency molds together an office with a critical mass of permanent employees and out-sources tasks requiring more personnel (e.g., test development, editing, and test form assembly).

1.4.S. The state agency responsible for testing should develop a comprehensive RFP that describes clearly and in detail both the end products to be provided by the vendor and the development process.

Comment: The RFP is the roadmap for the creation of a quality testing program. It should contain detailed specifications for all key areas, including, but not necessarily limited to, the totality of services and products to be provided, timelines for performance, quality criteria, responsiveness criteria, mid-course correction opportunities, and the process for evaluating proposals. When developing the RFP, agency staff should be aware of all required activities for a defensible testing program and should specifically assign responsibility for each activity to itself or to the vendor. It is occasionally the case that the state is not sure how to accomplish a certain goal of the testing program and wants vendors to propose a solution. In this case, the RFP should clearly separate those requirements that are firm, those that are aspirational, and those that are simply unknown. It should be very clear to vendors whether they are responding to specific requirements or proposals for

implementing general requirements. The state should also clearly spell out timelines in the development process, both for test development and for scoring/administration.

1.5.S. To provide a fair comparison of proposals received in response to an RFP, states should require all vendors to either: (a) provide costs for a fixed set of services and products specified in detail in the RFP; or (b) specify in detail what services and products they could provide for fixed incremental costs (e.g., \$3, \$5, \$10 per student). The method should be chosen in advance by the state and clearly specified in the RFP.

Comment: When vendors bid on different combinations of services and products, their proposals are not comparable and it is difficult for the state to evaluate cost effectiveness. The lowest bid may have to be accepted when essential activities are missing or incomplete. When all vendors use the same method, a fairer evaluation of their proposals is possible.

States with precise knowledge of the test products and services they will need are more likely to benefit from option (a), while states with less precise knowledge, or whose plans may change, are more likely to benefit from option (b).

2

Test Development

The standards and responsibilities in this section describe test development activities necessary to produce psychometrically and legally defensible high-stakes tests efficiently and to a high standard of quality. These activities may be completed by either the state or the vendor. If responsibility for any portion of an activity is assigned to the vendor, the RFP and resulting contract should so provide. If the state decides to retain responsibility for the activity, the state may seek advice from the vendor but should clearly indicate that expectation in the RFP and resulting contract.

For each activity or portion of an activity assigned to a vendor, the RFP and resulting contract should describe in detail what is expected of the vendor, any special conditions or limitations, and the compensation to be paid. If a state requests changes or delegates additional responsibilities to the vendor after the contract has been signed, the state will likely need to renegotiate the price.

2.1.V. A policy for quality assurance for all test instruments shall be developed and implemented. The policy should include specific activities, timelines, and locus of responsibility for evaluating the quality of each assessment instrument, including but not limited to, item quality, graphics quality, print quality, forms quality, equating and scaling accuracy, and quality of ancillary materials (e.g., measuring instruments, lab equipment). It shall be described in the proposal and incorporated into the contract.

Comment: In test development, quality assurance is a split responsibility between the state and the vendor. The vendor maintains primary responsibility for the quality of the work it produces, but the state ensures that the vendor's product matches the state's intent. The vendor must have detailed quality control procedures developed prior to contract initiation and be prepared to staff and implement them efficiently and effectively. The vendor should also clearly articulate the expectations for interaction with state staff during this process so the state can adequately schedule and staff required reviews and signoffs to avoid unnecessary delays. All deliverables should be thoroughly checked by knowledgeable staff before being sent in draft form to the state. Following established quality control procedures will decrease the likelihood of errors and increase the vendor's reputation for delivering a quality product. The vendor may consider seeking outside review and/or certification of its quality control procedures to enhance their usefulness and credibility. It is also advisable for the vendor to formally document all reviews and signoffs and to have multiple checks where feasible and appropriate.

2.1.S. The state shall develop and implement a quality assurance policy that serves to monitor the vendor. This policy may incorporate techniques such as signing off on all page proofs, checking final copies of test booklets before distribution to schools, and sampling score reports to ensure typographic quality.

Comment: It is important at all stages of development to check work that has been completed to ensure quality, accuracy, adherence to professional standards, and satisfaction of all program policies, administrative rules and legislative statutes. When tests will be used to make high-stakes decisions, multiple checks and rechecks are imperative to detect and correct any errors. While, vendors maintain primary responsibility for the quality of their work, vendor checks do not entirely replace the need for quality reviews by states, especially in high-stakes situations. Such quality reviews should be systematic, including proper training for staff and adequate enforcement mechanisms, and must be implemented in an effective and timely way.

Equating and scaling accuracy are especially important under the ESEA reauthorization in PL 107-110, which require additional uses of data that depend on sound vertical scaling and annual equatings of forms to measure Annual Yearly Progress. It is especially important that equating and scaling studies be conducted on the appropriate populations to which the data will be generalized. In most cases, this means within-state samples for the equating and scaling studies. If states are using publishers' off-the-shelf standardized tests intact, it is appropriate to use the publisher's national equating and scaling groups. If the states are augmenting publisher's off-the-shelf tests with state-specific items, thus creating new scales that measure state standards more accurately, it is important that the samples for the equating and scaling studies be taken from inside the state, or from a source that can be proven sufficiently similar for statistical purposes.

2.2.V. If requested by states, vendors shall be prepared to explain why the test is an appropriate instrument for its intended uses to educational and political constituencies, the press or in court. In cases where such activities are extensive, provisions defining compensation should be included in the contract.

Comment: The state has primary responsibility for defending its programs and decisions but can be aided substantially by vendor expertise and knowledge. Vendors may be most knowledgeable about their products and may have consultants who can provide a national perspective on the issues of concern to various constituencies. The vendor may also have key information required for defending testing activities in a legal forum. All such involvement should occur only when specifically requested by authorized state staff. Costs for such vendor activities should be explicitly included in the RFP and contract or should be dealt with in a separate addendum as the need arises. Vendors should have adequate and appropriate staff to handle such activities.

2.2.S. The state retains primary responsibility for responding to questions regarding the testing system and instruments. The state must be prepared to answer such questions from its legislature, the public, and/or the media. To this end, states must employ some staff with substantive knowledge of psychometric issues and a familiarity with the particular state testing system.

Comment: States may request assistance from developers in explaining tests to important political and educational constituencies, such as state legislators, the press, education organizations, and parent groups. In cases of legal challenge, states may need assistance in court explaining how the test development process satisfied all necessary psychometric and legal standards. Contracts should indicate whether, or under which circumstances, such assistance would require additional compensation.

State agencies must also be proactive when it comes to communicating with the legislature. Longstanding relationships with legislators and legislative staffs will enable the agency to help shape policy that is sound and easily justifiable, rather than having to react to legislation that is thrust upon them.

2.3.V. The following documents shall be provided by vendors to states: a) an annual management plan, including a schedule for all tasks required to carry out the plan; b) test development and construction specifications; c) written status reports at regular, agreed-upon intervals as provided for in the contract. In between, updates should be provided as needed through a medium determined by the state (e.g., via phone or e-mail). Depending on the complexity of the program and contract specifications, monthly or quarterly planning meetings with the vendor should be held to discuss current progress, upcoming tasks, problems, and mid-course corrections.

Comment: Regular communication between state and vendor staff is a key component of a successful testing program. Detailed documentation of program activities is also important for creating a historical record, informing policymakers, and providing supporting evidence in the event of a legal challenge. Communication and documentation should not be left to chance but rather should be part of the project plan from the outset.

Statewide assessments are typically moving targets in the early years of development because priorities change, laws change, the learning curve for state staff is steep, and timelines may be ambitious. Therefore, regular communication between agency staff and vendor staff is essential as the details of the program are worked out. Planning meetings are useful for problem-solving and consensus building while written reports provide useful documentation of decisions made and work completed.

2.3.S. The state has responsibility for devising its own plan for test development, which should include those areas that the vendor cannot feasibly control. These may include the timely identification and convening of educator review panels; item selection or review; and collecting and providing all necessary information to the vendor for the development of supplementary test materials (e.g., test aids and administrative manuals).

Comment: Whether or not a state chooses to delegate test development activities to a vendor, there are some key functions that state staff must perform. For example, the state is in the best position to seek nominations and to constitute educator committees for such functions as reviewing items, selecting items, recommending passing standards, or scoring open-ended responses. States are also responsible for setting testing dates and procedures and for communicating this information to the vendor in time for inclusion in test administration manuals and other related testing materials. These decisions must be communicated to the vendor with sufficient lead time, as defined in the contract, to allow for typesetting, proofing, printing and distribution of these materials to the schools.

2.4.V. Assessments developed for state testing programs must be consistent with all relevant professional standards contained in the 1999 AERA/APA/NCME *Standards for Educational and Psychological Testing*.

Comment: The test *Standards* reflect a consensus among the membership of the three sponsoring organizations, the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. They cover major aspects of testing such as validity, reliability, setting passing standards, opportunity to learn, item development, bias reviews, equating, accommodations, English language learners (ELL), scoring, reporting, and documentation. Professional judgment is required in applying relevant standards and should reflect the goals of state policymakers to the maximum extent consistent with best professional practice.

2.4.S. The state should develop and implement a policy for monitoring the vendor's work for consistency with all relevant professional standards contained in the 1999 AERA/APA/NCME *Standards for Educational and Psychological Testing*. Options include the use of the state Technical Advisory Committee, employing external psychometric consultants hired directly by the state, or hiring an independent evaluator to provide periodic critical reviews of testing program activities.

Comment: While it is primarily the responsibility of the vendor to ensure that assessments developed for state testing programs are consistent with the test *Standards*, states also have a responsibility to be familiar with relevant standards and to institute procedures for systematically monitoring the degree to which state assessments are consistent with those standards. In particular, states are in the best position to investigate acceptable applications of the standards that are most consistent with the goals of state policymakers and to advocate for changes by the vendor when warranted. Closely monitoring consistency with the test *Standards* also allows the state to anticipate and correct any problems early in the process, and to be in a stronger position in the event of a legal challenge to the testing program.

2.5.V. A technical manual which includes all relevant psychometric information for each assessment in the testing program shall be developed. The technical manual shall be completed within 6 months of the first live test administration and shall be revised annually thereafter, a copy of which shall be available within 6 months of each successive test administration. The content and timeliness of the technical manual shall be described in the proposal and incorporated into the contract.

Comment: A technical manual is essential in documenting that each assessment instrument meets all professional standards for psychometric and legal defensibility. The technical manual also is an organized repository of psychometric information that should be available to users of test data, including researchers and the public. The technical manual should include, but is not limited to, information on: purpose, test blueprint, test development, validity, reliability, accommodations & testing ELLs, security, administration, scoring, equating & scaling, setting performance standards, opportunity to learn, reporting, and appropriate use & interpretation of test data. Appendices should include related materials such as relevant state statutes, administrative regulations, state standards, sample items, committee rating forms, state & district performance summaries by ethnic group, and other relevant information.

2.5.S. The state bears two responsibilities with respect to the technical manual: First, to provide the vendor with all information required for the creation of the technical manual, and second, to review and approve the final version of the technical manual.

Comment: It is important for the state and the vendor to work together on the Technical Manual. While the vendor takes major responsibility for the initial draft, the state has an important responsibility for furnishing necessary information to the vendor. For example, if the state has identified and coordinated educator review committees, the state must provide the vendor with a written description of the selection procedures and relevant information about the participants for inclusion in the appropriate chapter of the technical manual. The state should also undertake a thorough review of the draft technical manual with an eye to accuracy and usability by its intended audiences. The state should also provide the final signoff prior to publication and should assist the vendor in revising it as appropriate for subsequent testing cycles.

2.6.S. For high-stakes testing programs, a technical advisory committee (TAC) should be established to guide program activities. States may wish to establish additional committees for special purposes, as needed, such as for hand scoring or item development issues.

Comment: Guidance from both in-state and out-of-state experts on a TAC can provide troubleshooting assistance, a check on adherence to professional psychometric standards, support for psychometrically necessary but politically unpopular decisions, and information on research and options from other programs. TAC members can also be available for questions and comment as issues arise between

regularly scheduled meetings. In addition to the permanent TAC, states may find it useful to further establish similar ad hoc advisory committees in areas that present special challenges or are otherwise of special interest to their assessment program, such as hand scoring for a state that makes extensive use of open-response items on its assessment.

Responsibility for meeting arrangements may be delegated to the vendor, provided that authority for appointment or removal of TAC members remains with the state. However, the state should realize the expertise that the vendor possesses in terms of creating the TAC. The vendor can, in many cases, arrange meetings with much more ease administratively than can states. In addition, the vendors are knowledgeable about the preeminent experts that would be available to serve on the TAC.

2.7.S. Tasks, timelines, and the party responsible for each should be clearly delineated in contract documents for each assessment to be developed. Provisions should be included in the contract establishing conditions under which the state may add, delete, or modify contract requirements through contract amendments or change orders.

Comment: To facilitate communication and coordination between agency and vendor staff, it is important for both parties to know ahead of time what tasks must be completed by what deadline and by whom, for the assessments being developed to be ready for implementation by the intended date. The tasks should be described in enough detail so that contract managers can monitor performance on a regular basis and take necessary steps to correct any deficiencies before major problems develop. There should be a clear delineation of authority. The vendor should know who in the state is authorized to accept items, forms, policies, reports etc. There needs to be a clear system of sign-offs and chain of command, including allowance for emergencies (e.g., what to do if the testing director is out sick during the most crucial time of approving proofs of test booklets or score reports or rubrics.)

New testing programs should expect some unanticipated delays and difficulties and should include specific contract provisions for handling such situations. The goal must be to ensure that all tests are sound meet professional standards for psychometric and legal defensibility, even if that means revising timelines or task assignments to correct deficiencies. Unanticipated problems with item development, field testing, scoring or analyses can jeopardize the defensibility of a test if extra time is not taken to repeat steps or take appropriate corrective action. In the long run, failure to do so may adversely affect the viability of the entire state testing program.

The contract should apportion the financial responsibility for such delays based on the actions of the parties, including but not limited to, whether the delay resulted from vendor nonperformance, the state failing to meet agreed upon deadlines, or the state having made changes.

2.8.S. Written policies should be developed and implemented for: (a) test security, (b) test specifications, (c) item reviews (including but not limited to: sensitivity, differential performance, opportunity to learn, and psychometric quality), (d) accommodations for students with disabilities, (e) testing English language learners, (f) maintenance of confidential testing information, and (g) appeals.

Comment: Written state policies serve the dual function of communicating procedures and ensuring that no critical steps are left out. They also ensure that important decisions, such as whether a reader or calculator will be provided to students with disabilities, are made early in the process when decisions about test purpose, use, and content can be modified accordingly. This is also an ideal time, in consultation with agency counsel and a representative of the state attorney general's office, to ensure that proposed test instruments will satisfy all applicable federal and state laws. The state may request that the vendor provide consultation, but final policy decisions should be the responsibility of the state.

This is an area in which the state can both receive assistance from the vendor as well as provide assistance to the vendor. Vendors or independent consultants can help the state craft sound policies, based on their experience in the business. The state must be wary, however, of relying exclusively on advocacy groups to help them craft their policies.

The state can, and should, aid the vendors in the area of test security. To this end, the state should: monitor schools or districts; spot-check schools and districts to ensure that policies are being followed; communicate policy expectations to schools and districts; implement policy for the reporting and investigation of test security breaches; and work to enact rules and legislation that adequately deal with test security breaches. Sanctions for security breaches should include civil penalties such as the loss of license or credentials for teachers and administrators as well as criminal penalties for the most egregious cases. In addition, test scores in schools or districts where there is solid evidence of cheating should be invalidated; this is particularly true in accountability systems where consequences are tied to school performance.

2.8.V. Detailed policies and procedures for ensuring the security and integrity of each assessment instrument and the secure destruction of unneeded documents shall be developed and implemented. In addition, the vendor must bear some of the responsibility for preventing cheating and for catching cheaters. The vendor must follow state test security policy in situations where one exists.

Comment: A variety of procedures may be appropriate depending on the purpose(s) and use(s) of each assessment. Such procedures might include, but are not limited to, confidentiality agreements, signoff and storage requirements for test materials, numbering and sealing test booklets, procedures for returning test materials, directions for administration, training for test administrators, test preparation guidance, analyses for monitoring potential cheating and for investigating irregularities, ethics standards for educators, and sanctions for those who violate security policies. Contracts should also consider what is done with answer sheets at the end of a contract, how and at what cost a contractor will fulfill an agency request for answer sheets, the state's legal responsibilities for records storage, and

the states policies and laws related to parental access to such things as test booklets and answer sheets. Many states have found it useful to include test security violations and sanctions in legislation or administrative regulations.

2.9.S. In cases where the state retains substantial responsibility for test item drafting or review through educator committees, those assigned to such committees should reflect the goals and philosophy of state policymakers and be knowledgeable about scientific (i.e., evidence-based) research on curriculum and instruction. States may also wish to involve external content experts in reviewing test items.

Comment: States often retain substantial control over the initial drafting or, at least, review of test items through panels of local educators and curriculum staff or consultants. In some cases, states have discovered that such committees do not always reflect the academic goals or philosophy of policymakers, nor do they necessarily apply an accurate knowledge of rigorous research on curriculum and instruction. Given their critical function, education policymakers should ensure that such committees reflect their philosophy and possess a sound grasp of the best research on curriculum and teaching. Alternatively, states may wish to establish item review procedures to ensure that the work of item writing committees is reflective of policymakers' academic philosophy as well as scientific research. Policymakers may also find it useful to involve external content experts with national perspectives and expertise to review items.

2.9.V. The vendor has an affirmative responsibility to communicate to the state when items or item types do not reflect the best educational research. In addition, vendors must make efforts to correct situations where inappropriate standards are likely to lead to assessments that do not reflect the best educational research.

Comment: Vendors' national consultants can provide states with comprehensive information about appropriate use of items types and applications being used in other states. Vendors' consultants also follow educational research closely and can provide states with guidance on the latest scientific research. Where states are able to revise their standards prior to assessment implementation, vendors can provide professional advice regarding options consistent with best practice.

3

Test Administration

The standards and responsibilities in this section describe activities necessary to administer psychometrically and legally defensible high-stakes tests efficiently and to a high standard of quality. These activities are typically assigned to the vendor but responsibility may be shared with the agency. If the agency decides to retain responsibility for an activity, the agency may seek advice from the vendor but should clearly indicate that expectation in the RFP and resulting contract.

For each activity or portion of an activity assigned to a vendor, the RFP and resulting contract should describe in detail what is expected of the vendor, any special conditions or limitations, and the compensation to be paid. If a state requests changes or delegates additional responsibilities to the vendor after the contract has been signed, the state may have to renegotiate the price.

3.1.V. Where the state has delegated such responsibility to the vendor, a plan for developing and maintaining a database of student and school testing information shall be created. The plan should provide mechanisms for tracking student movement, keeping track of retests, collecting demographic information needed for data analyses and reporting, ensuring confidentiality of individually identifiable student data, correcting student identification numbers as needed, and updating files when errors are uncovered.

Comment: With multiple subjects, multiple grades, and retests, it is essential that test data be organized in a format that is accessible, accurate, provides all data needed for state and federally-mandated analyses, and tracks the testing history of students, items and test forms. Because most of the data collected will involve confidential or secure information, detailed policies for protecting the confidentiality of data collected and retained must be developed.

The RFP and resulting contract should clearly specify vendor expectations in this area. Creation and maintenance of electronic databases is expensive and the cost may be prohibitive for some small testing programs. If the state chooses to maintain or collect its own data, the contract should clearly specify the form and content of data files the vendor is expected to provide to the agency.

3.1.S. The state has the responsibility to collect and report useful data to a variety of constituencies, including satisfying federal requirements under the No Child Left Behind Act. Where permitted by state law, a database of student and school information can be highly useful. The state is ultimately responsible for ensuring that such a database of student and school information is maintained properly; where it has elected to delegate this responsibility to the vendor, the state is responsible for monitoring the work. States choosing not to use a state-level database possess other means for carrying out this function that a vendor does not, such as requiring school districts to provide the data.

Comment: When a state chooses not to contract with a vendor to maintain a state database, the state must assume the responsibility for collecting and maintaining assessment data in a form that will produce usable information for various constituencies and that satisfies applicable law. Appropriate procedures must be implemented to satisfy confidentiality requirements and to ensure proper use and access to all data. While a state has options other than creation of a statewide database, such options limit the usefulness of the available data.

3.2.V. The vendor proposal and resulting contract shall specify procedures for determining quantities of materials to be sent to districts (or schools), tracking test materials that have been sent, and resolving any discrepancies. A mechanism shall be developed for ensuring the accuracy of enrollment data supplied to the vendor and for updating school requests for additional or replacement materials. Instructions for handling test materials and for test administration (e.g., Administrator's Manual) shall be shipped to districts at least one month prior to testing to allow time for planning and staff training.

Comment: Valid and fair test results require adherence to all standard test administration conditions and security procedures by all test administrators. Test administrators are best prepared for this task when sufficient quantities of materials are received prior to testing and training has been provided using the actual instructions to be employed during testing. In order for districts to receive sufficient quantities of materials, accurate and timely enrollment information must be supplied to the vendor and a mechanism must be established for efficiently responding to requests for additional or replacement materials. Administrator's manuals and instructions for handling test materials are important communications for district planning and test administrator training and should be available for study prior to the receipt of test materials. By making such procedures and communications responsive to the needs and concerns of districts and schools, greater cooperation should be achieved.

3.3.V. Timelines and procedures for receipt and return of test booklets and answer sheets shall be consistent with an agreed upon test security policy and specifications in the RFP and resulting contract. Generally, test materials should arrive in sealed containers no earlier than one week prior to testing, should remain in a secure, locked storage area while in district/schools, and should be repackaged and picked up within two days after test administration has been completed.

Comment: The security of test materials and the accuracy of state test data depend on the timely receipt and return of test materials by schools. The RFP and resulting contract should provide detailed descriptions of all security procedures to be followed by the vendor, including procedures for distributing, tracking, and returning test materials.

The state may wish to delegate the responsibility for training of school and district personnel to the vendor.

3.3.S. States must develop and implement a policy for ensuring that schools and districts comply with the policies enumerated in the RFP and contract. When non-compliance is an issue, the state must be able to impose sanctions or otherwise compel action on the part of the local education agency . In addition, the state is responsible for the training of school and district personnel in the security policies.

Comment: The state retains responsibility for training, monitoring, and investigating local education agencies' compliance with established test security procedures. Administrative rules or statute should enumerate educators' responsibilities, proscribed activities and sanctions for violators. The state also has a duty to monitor contractor activities and to assist in the resolution of unforeseen circumstances (e.g., school closing on test week due to a major flood or storm damage).

3.4.V. Reliability for any high-stakes exam should be at the highest levels. Where open-ended response items or essays are included in an assessment, two raters shall score each response with at least 70% agreement on initial scoring. When raters disagree on initial scoring, resolution (re-scoring) by a senior or supervisory rater is required.

Comment: Tests that are to be used for high stakes for either educators or students should attain high standards of reliability, as may be exemplified by an overall internal consistency rating of at least 0.85 to 0.90 on a 0-1 scale. Such overall reliability will not be attained unless hand scored items, typically essays or other open-ended items, also attain adequate levels of inter-rater reliability. Trained raters using detailed scoring rubrics who are periodically rechecked for accuracy should be able to score responses with a high degree of agreement. When two raters disagree and the test is being used for high-stakes decisions about individual students, fairness dictates that an experienced third rater resolve the discrepancy. (In cases of items with a large number of score points, "agreement" may consist of

adjacent scores.) Alternative procedures for computerized scoring of open response items can include one trained rater serving as the second rater, with similar procedures for resolving discrepancies. For assessments that do not include high-stakes for students, a single rater may be sufficient as long as proper procedures are in place for checking samples for rater drift.

3.5.V. Quality control procedures for checking the accuracy of all item information, student scores and identification, and summary data produced by the testing program shall be developed and implemented. The standard for the error rate of data reports provided by a vendor to an agency for review is zero.

Comment: The vendor has a duty to formulate and implement quality control procedures for data generation that have as their goal the production of error-free reports and summary data. All data operations should be subject to multiple checks for accuracy before being released to the state. The vendor should document its quality control procedures for state review and create detail logs that trace the application of those procedures to the state data reports.

3.5.S. Data reports released by state agencies must also be error free. The state must develop its own quality assurance policy to monitor the work of the vendor. Data reports should be examined before general release. Effective techniques prior to release include: running score and summary reports on “dummy” data to ensure that the output is correct; close examination of a sample of the reports; sending preliminary data to select schools or districts for review; or having the state TAC or an outside consultant examine a sample of the reports.

Comment: When erroneous data is released publicly, the testing program loses credibility and incorrect decisions may be made. It is imperative that all reasonable procedures be used to check the accuracy of all testing program data before report distribution or public release. The vendor has primary responsibility to find and correct errors, with agency staff acting as a final check. The expectation of zero errors is contingent upon the state providing all necessary information. Non-trivial vendor errors may trigger financial penalties in states that include such provisions in their contracts.

3.6.V. When an item error, scoring error, or reporting error is discovered, the vendor shall notify state staff immediately. Vendor staff should then work closely with agency staff, and technical advisory committee members or outside consultants where appropriate, to develop a comprehensive plan for correcting the error. The plan should include the provision of timely and truthful information to the affected stakeholders.

Comment: The way in which an error becomes public and the actions taken to correct it can have a major impact on public perceptions. Straightforward communication of information as it becomes available and immediate corrective action can help restore public confidence in the vendor and the state testing program. Error does not include reasonable differences of opinion.

3.7.V. Testing report forms shall be received by the district or other responsible entity (e.g., charter school) no later than the end of the semester in which testing occurred. Individual student reports for multiple-choice tests should be received within 2 weeks of the date on which answer documents were received by the vendor. School, district, and state reports should be produced within 2 weeks of the cutoff date for return of answer documents. For tests containing open-ended items or essays requiring ratings, individual student reports should be received within 6 weeks of the date on which answer documents were received by the vendor. School, district, and state reports should be produced within 6 weeks of the cutoff date for return of answer documents. Where an assessment is composed entirely, or almost entirely, of essays or other open-ended items, more time is likely to be necessary for scoring. The contract should specify any antecedent conditions that must be met by the agency for reports to be delivered on time.

Comment: For data to be useful for instructional improvement and for making decisions about enrollment in remedial classes or summer school, it must be received prior to the beginning of the next instructional semester following the date of testing. Turnaround time will vary depending on program complexity but should be kept as short as possible while maintaining accuracy. If state staff with expertise believe that these timelines do not reflect their needs, they can elect to deviate from them; however, a rationale should be provided. It is understood that there are tradeoffs inherent in the timeline process, and state policymakers should be able to explain their reasoning for allowing vendors to go beyond these timelines, if they elect to do so.

Plans should include rules for scoring of late arriving papers, particularly with regard to calculating summary statistics. (E.g., how long should one school be allowed to hold up the state summary statistics?) Clear guidelines in this area are especially important for tests that include open-response items; in such cases, a contractor will typically have only a limited window of time to implement the work of the human raters. The beginning date of the 2-week or 6-week scoring window should be clearly defined in the contract. Further, the scoring timeline for the contractor should be defined to include all activities that the contractor needs to perform (i.e., including all of those required to ensure the integrity of the data, not just the scoring itself once these activities have been completed).

3.8.S. When the RFP and resulting contract provide reasonable timelines for scoring and reporting, and the agency has met its obligations, states may wish to include contractually agreed upon incentives for performance by the vendor. Incentives may include a bonus for early completion or a penalty for late performance or errors. Administration activity timelines may well exceed typical annual state appropriations; states may benefit from multi-year funding plans and contracts across fiscal years (which may be cancelled if the budget must be reduced or the program is eliminated). States must, of course, stay within statutory constraints imposed by their respective legislatures.

Comment: The RFP and resulting contract should contain workable timelines that allow sufficient time for scoring and quality control. When delays occur, timely communication is vital for resolving the problem expeditiously and dealing effectively with those affected. If bonus or penalty clauses are included in contracts, timelines for agency staff to complete prerequisite tasks should also be specified. States may want to consider contract payment schedules to vendors based upon the delivery of specified products and services rather than on the basis of calendar dates alone.

The majority of state testing programs choose a spring test administration that results in demands on vendors to produce reports for multiple programs during the same narrow time frame at the end of the school year. States able to schedule scoring during nonpeak periods may have greater flexibility in turnaround time and may gain a cost savings. Programs with bonus or penalty contract provisions may likely be given priority in such circumstances (though other considerations are also likely to come into play). The contract should contain the same scoring deadlines contained in the RFP. States may wish to attach to these deadlines specific liquidated damages for each day of non-delivery. In such cases, the contract should include provision for performance bonds against which the agency can claim the damages.

Funding is not a simple issue of obtaining annual appropriations. Activities for any given assessment administration from start to finish require approximately 18 months. This means that the typical fiscal year of 12 months and the assessment “year” of 18 months will conflict unless special provisions are made in the funding. One would not want to be in the position of having to write a contract for the first 12 months of activities and then another contract for the last 6 months of work. Furthermore, there is the likelihood that the fiscal year will not coincide with the RFP/contract/implementation cycle. For example, if the legislature appropriates funds that become available on July 1, 2002, it is impossible to create a program that will be implemented in March 2003. The solution is to create multi-year funding plans and permit the agency to contract across fiscal years. Contracts can be cancelled if budgets must be reduced or the program is eliminated. Contracts should allow for necessary audits if required by the state comptroller.

3.8.V. When a delay is likely, the vendor should notify agency staff immediately and provide a good faith estimate of its extent.

Comment: Immediate notification of the state when a delay is likely is always best practice for the vendor. Quick notification allows all parties involved to assess the scope of the problem, its impact, and any necessary actions.

4

Uses of Data

The standards and responsibilities in this section deal with activities subsequent to the administration of the test, but directly connected with the interpretation of test data and score reports. Because the scope of these standards is limited to testing and not to instructional practice and school administration, there are many activities that are related to testing and test results that are not covered.

Generally, use of the data is the responsibility of the state and the LEAs. Some of these activities might be delegated to vendors, however. It is important that the RFP and the resulting contract make it clear what is expected of the vendor. If the state requests changes or delegates additional responsibilities to the vendor after the contract has been signed, the state may have to renegotiate the price.

4.1.V. Clear and understandable reports must be developed for communicating test results to educators, students, parents, and the general public.

Comment: Clear communication and guidelines for interpretation are essential to appropriate use of test data. Interpretative guidelines should be reported for both individual and school level reports. Cautions on over-interpretation, such as using tests for diagnostic purposes for which they have not been validated, should be made clear.

4.1.S. The state is responsible for communicating the test results to educators, students, parents, and the general public. An important part of this responsibility is the design of reports of test data. The state might choose to do this itself, or delegate it to the vendor (see 4.1V.) If the state delegates the design of reports to the vendor, the state shall be responsible for clearly sharing with the vendor its expectations about the audience for the reports, the purpose of the testing program and the uses to which the data will be put. The state shall also make clear, in writing, its requirements for the languages of reports to parents and the community and whether the reports should be graphic, numerical or narrative. The state shall be responsible for approving report formats in a timely manner as described in the contract.

Comment: The state is in the best position to determine how the test results will be used and what data will best communicate relevant and important information to the various audiences. It is also the prerogative of the state to determine report

formats, types of scores to be reported and appropriate narrative information to accompany each report. Final report formats should be approved by the state before actual reports are printed. The state may also choose to provide access to data on a website designed by the state or its vendor.

4.2.V. If specific responsibility for monitoring the use of the test data is a part of the vendor's contract, the vendor shall develop detailed policies and procedures for promoting and monitoring the proper interpretation of test data and implement those plans. Regardless of delegation of responsibility in this area, the vendor shall have a system for compiling any information of which it becomes aware regarding the improper and/or incorrect uses of data and relaying that information to the state.

Comment: The vendor, just like the state, bears responsibility for supporting and encouraging the ethical and proper implementation of the assessment system. Where the vendor has become aware of inappropriate practices in the course of its work on the assessment system, these should be reported to the state.

4.2.S. The state shall determine how the test data are to be used, and develop detailed policies and procedures for the proper use of the data. The state shall use the resources of the vendor or other qualified individuals (such as the Technical Advisory Committee) as needed to ensure the proper use of the test data for the purposes for which the test is intended, and make all reasonable attempts to prevent the improper use and interpretation of the data.

Comment: The only purpose of the testing program is to provide data that meets the goals of the program. Improper interpretation and use of the data negate all of the activities that led to the creation of that data, wasting money and time and perhaps causing serious disservice to students in the state. Since the vendor knows the test well and often has the capabilities to assist in interpretation and dissemination, the state may want to include in the contract the use of the vendor's resources in conducting workshops around the state for teachers and administrators, joining and assisting the state personnel in presenting the data to stakeholders, such as legislative committees and the press, or assisting in the dissemination of the data. The state should use its greater knowledge of schools and districts in the state and their needs to help the vendor in these functions. The complementary expertise of the vendor and state should be utilized to ensure that the data is use in an appropriate manner.

Acknowledgements

We thank the following individuals for their advice and input regarding this document, primarily through their participation at the Education Leaders Council Testing Summit, held February 20-21, 2002 in Austin, Texas. By including their names here, we acknowledge their contributions; this list should not be interpreted as necessarily implying any endorsement of the ideas contained in this document by the individuals or organizations listed herein. This document is a product of the participating states listed on the cover, the Education Leaders Council, and AccountabilityWorks, Inc.

Michaele Alcaro

Policy Specialist
Pennsylvania Department of Education

Dr. Arturo Almendarez

Deputy Commissioner for Programs
and Instruction
Texas Education Agency

Leslye Arsht

President
StandardsWork

Virginia Belland

Vice-President, Sales
Eastern Division
EdVision

Sondra Bisig

Director of Curriculum
Implementation
Lightspan, Inc.

Larry Bosley

Vice-President Territory Sales
Vantage Learning

Stacey Boyd

President & CEO
Project Achieve, Inc.

Merri Brantley

Research and Policy Advisor
Office of the State Superintendent of
Schools
Georgia Department of Education

Benjamin Brown, Ph.D.

Executive Director
Evaluation and Assessment
Tennessee Department of Education

Ron Carriveau

State Assessment Director
Arizona Department of Education

David Chayer

Vice President, Research and Test
Development
Data Recognition Corporation

Maye Chen

Vice President of Operations
Project Achieve, Inc.

Mitchell D. Chester

Assistant Superintendent
Office of Assessment
Ohio Department of Education

Dr. Criss Cloudt

Associate Commissioner for
Accountability
Texas Education Agency

Wilmer “Bill” Cody
NCES/Westat
Education Policy Studies

William Cox
Standard and Poors

Richard Cross
Research Director
AccountabilityWorks

Keith Cruse
Managing Director
Texas Education Agency

Joel Dando
Manager Public Affairs
Renaissance Learning Inc.

Jill Dannemiller
Associate Director of School
Accountability
Ohio Department of Education

Alicia Diaz
Education Consultant
Ohio Department of Education
Policy Research and Analysis

Richard Dobbs
CTB McGraw Hill

Chrys Dougherty
Director of Research
National Center for Educational
Accountability/Just 4 Kids

Dennis Doyle
Co-founder and Chief Academic
Officer
Schoolnet, Inc.

Brad Duggan
Interim President and Executive
Director
National Center for Educational
Accountability/Just For Kids

Susan Engeleiter
President and Chief Operating Officer
Data Recognition Corporation

Thomas H. Fisher, Ed. D
Educational Testing & Assessment
Administrator
Florida Department of Education

Steve Fleischman
Executive Director
Education Quality Institute

Andrea Fuller
Vice President, Sales & Marketing
SMARTHINKING, Inc.

Matthew Gandal
Executive Vice President
Achieve, Inc.

Mark Gedlinske
Chief Technology Officer
Data Recognition Corporation

Christopher Gergen
President/Chairman and Co-founder
SMARTHINKING, Inc.

Dennis Gormley
ETS K-12 Works

Maureen E. Grazioli
Vice President, Sales and Marketing
Riverside Publishing

Ruth Grimes-Crump
Office of Elementary and Secondary
Education
U.S. Department of Education

Dr. David J. Harmon
Director of Research, Evaluation, and
Testing
Georgia Department of Education

Carolyn Haug
Policy Assessment
Colorado Department of Education

Dr. Hugh Hayes
Deputy Commissioner for Initiatives
and Administration
Texas Education Agency

Mark Heidorn, Ph.D.

Director of Program Operations-
Northern Region
CTB/McGraw-Hill

James M. Hill, III

Director, Computer-based Testing
Marketing
Harcourt Educational Measurement

Steve Hodas

The Princeton Review

James “Jim” Horne

Secretary
Florida Board of Education

Shannon Housson

Director of Analysis and Reporting
Texas Education Agency

Steven Hoy

Vice President
Lightspan, Inc.

Gary Huggins

Executive Director
ELAC

Jeremy Hughes

Director, Office of Educational
Assessment and Michigan Merit
Award
Michigan Department of Treasury

Brian Jones

Vice President for Communication and
Policy
Education Leaders Council

Debra Jones

Senior Intelligence Advisor
SAS inSchool

Dr. Margie Jorgensen

Vice President
Harcourt

John Katzman

CEO
The Princeton Review

Lisa Graham Keegan

CEO
Education Leaders Council

Richard Kohr

Measurement and Evaluation
Supervisor
Pennsylvania Department of Education
Division of Evaluation and Reports

Carolyn Kostelecky

Assistant Vice President of Educational
Services
ACT, Inc.

Steve Kromer

Vice President
NCS Pearson

Stephen Kutno

The Princeton Review

Jackie Lain

Associate Director
Standard and Poors

John E. Laramy, Ph.D.

President
Riverside Publishing

Cait Ngo Lee

Evaluation and Grants Coordinator
Lightspan, Inc.

Lenny Lock

Education Associate
Pennsylvania Department of Education
Division of Evaluation and Reports

Dr. Kathleen Madigan

Executive Director
National Council on Teacher Quality

Gary Mainor

President
NCS Pearson

Dewayne Matthews

Vice President
Education Commission of the States

Patricia McDivitt

Vice President, Educational
Measurement & Test Development
ETS K-12 Works

Thomson W. McFarland

Research Associate
AccountabilityWorks

Brian McGee

Regional Vice President, Southern
Region
Harcourt Educational Measurement

Georgia McKeown

Chief of Staff - Jim Horne
Florida Board of Education

Dr. Ron McMichael

Deputy Commissioner for Finance and
Accountability
Texas Education Agency

Michael McNally

Vice-President of Business
Development
Vantage Learning

Chad A. Miller

Policy Analyst
Education Leaders Council

Libbie L. Miller, Ph.D.

Statistical / Research Analyst
Lightspan

Dr. William J. Moloney

Commissioner of Education
Colorado Department of Education

Dr. Mark Moody

Assistant State Superintendent for
Planning, Results, and Information
Management
Maryland State Department of
Education

Judith Morgan

Executive Assistant/Press for Faye
Taylor
Tennessee Department of Education

Ina Mullis

Professor and Co-Director of the
International Study Center
Lynch School of Education, Boston
College

Dean H. Nafziger, Ph.D.

President
Harcourt Educational Measurement

Dr. Beverly Nash

Assistant Division Director
Research, Evaluation, and Testing
Georgia Department of Education

Jeff Nellhaus

Associate Commissioner for Student
Assessment
Massachusetts Department of
Education

Jim Nelson

Texas Commissioner of Education
Texas Education Agency

Jo O'Brien

Assistant to the Commissioner
Colorado Department of Education

Dr. Billie Orr

President
Education Leaders Council

John Oswald

Consultant
AccountabilityWorks

Shaundra Overmyer

Manager of State Assessment Programs
Data Recognition Corporation

Terrance (Terry) D. Paul

Co-Chairman
Renaissance Learning Inc.

Dr. Ron Peiffer

Assistant State Superintendent of
Schools
Maryland State Department of
Education

Heidi Perlman

Massachusetts Department of
Education

Robert B. Peterson

President
C4SI, Inc.

Linda Pfister

President and CEO
ETS K-12 Works

Susan Phillips

Adviser and Consultant
AccountabilityWorks

Susan Pimentel

Standardswork

Kelly Powell

ELC Consultant

Debbie Ratcliff

Senior Director of Communications
Texas Education Agency

Theodor Rebarber

President
AccountabilityWorks

Dr. June Rivers

Assistant Manager, Value-Added
Assessment and Research
SAS in School

Marilyn Roberts

Michigan Department of Education

Ed Roeber

Vice President for External Relations
Measured Progress

Dr. Bill Sanders

Research Fellow, University of North
Carolina
Manager, Value-Added Assessment and
Research
SAS in School

Gary A. Schaeffer, Ph.D.

Director of Research Projects
CTB/McGraw-Hill

John Schilling

Chief of Staff
Education Leaders Council

Linda C. Schrenko

State Superintendent of Schools
Georgia Department of Education

Amanda Seals

Director of Media Relations
Office of the State Superintendent of
Schools
Georgia Department of Education

Ramsay W. Selden

Vice President & Director, Assessment
Program
American Institutes for Research

Anne Smisko

Associate Commissioner for
Curriculum, Assessment, and
Technology
Texas Education Agency

Malbert Smith

President
MetaMetrics, Inc

Russell R. Smith

Managing Director
Rothary Landis Group

Larry Snowwhite

Vice President, Government Relations
Houghton Mifflin Company

Dr. Lew Solmon

Senior Vice President
Milken Family Foundation

Patricia Spence

Director of Psychometric Services
Data Recognition Corporation

Bernia Stafford

Vice President of School Marketing
and Evaluation
Lightspan, Inc.

Jack Stenner

Chairman and CEO
MetaMetrics, Inc

John Stephens

Executive Director
National Assessment Governing Board

Phyllis Stolp

Director of Development and
Administration
Texas Education Agency

Karen Stroup

Office of Management Services
Colorado Department of Education

Circe Stumbo

President
West Wind Enterprises

Leonard Swanson

Vice President, Technology
ETS K-12 Works

Robert W. Sweet, Jr.

Professional Staff Member
Committee on Education and the
Workforce

Faye Taylor

Commissioner of Education
Tennessee Department of Education

Blake Thompson

Data Analysis Manager
GreatSchools.net

Cassandra Thompson

ETS K-12 Works

William “Bill” Tudor

Chairman & CEO
EdVision

Cynthia Ward

Vice President, Contract Management
and Proposals
ETS K-12 Works

John Winn

Assistant Secretary
Florida Board of Education

Wendy Yen, Ph.D.

Vice President, Research
ETS K-12 Works

Dr. Michael Yoes

NCS Pearson

Philip B. Young, Ph.D.

Director, State Programs
Riverside Publishing

Victoria Young

Director of Instructional Coordination
Texas Education Agency

Jodie Zalk

Assessment Specialist
Massachusetts Department of
Education

Molly Zebrowski

Manager of Business Development
NCS Pearson

Charles Zogby

Secretary of Education
Pennsylvania Department of Education