

# **Guidelines for Organizing and Leading a State Assessment Consortium**

***John Olson  
Ed Roeber  
Barry Topol***



# **Guidelines for Organizing and Leading a State Assessment Consortium**

*John Olson  
Ed Roeber  
Barry Topol*

**Assessment Solutions Group**

**September 2010**





This report was written by the Assessment Solutions Group with support from the Bill & Melinda Gates Foundation.

Citation: Olson, J., Roeber, E., & Topol, B. (2010). *Guidelines for Organizing and Leading a State Assessment Consortium*. Boston, MA: Assessment Solutions Group.

The information in this report reflects the views of the authors. Publication of this document shall not be construed as endorsement of the views expressed in it by the Bill & Melinda Gates Foundation.

© 2010 Bill & Melinda Gates Foundation. All rights reserved.

## Table of Contents

|  |    |
|--|----|
| I. Introduction .....  | 1  |
| II. Context for State Collaboration and the Guidelines ..... | 3  |
| III. Practical Issues .....                                  | 4  |
| IV. Technical Issues .....                                   | 21 |
| V. Performance Standards .....                               | 34 |
| VI. Project Management and Governance Issues .....           | 37 |
| VII. Cost Issues .....                                       | 47 |
| VIII. Summary and Recommendations .....                      | 57 |
| References .....   | 60 |
| Appendices .....   | 61 |

This page intentionally left blank

# Guidelines for Organizing and Leading a State Assessment Consortium

## I. Introduction

Several recent initiatives are directly affecting states and their assessment programs, and may impact state assessment systems for many years to come. First is the Common Core State Standards (CCSS) that are being developed jointly by the National Governors Association and the Council of Chief State School Officers. The standards will lay the foundation for the state assessments since states will be required to adopt them (should they wish to qualify for Race to the Top funding) and there must be a close alignment of their future assessments to the content standards. Second are the requirements as laid out by the United States Department of Education (USED) in the Notice Inviting Applications (NIA) for the Race to the Top (RTTT) common assessments competition. The priorities and requirements for state consortia in designing, developing, and implementing high quality assessment systems are listed in the NIA. These priorities for the RTTT are addressed in more detail in a later section of this report. Suffice it to say, the priorities will strongly impact the direction of state assessments in the future. Third, and related to the RTTT common assessments, is the additional requirement that the new assessments measure College and Career Readiness (CCR) standards, as defined by the NGA, CCSSO, and the states. All of these recent initiatives will have a direct and significant impact on the planning and decisions that need to be made by a state assessment consortium.

Currently, much energy is being expended in the design of new common assessment systems. However, less effort is being directed toward guiding states in actually implementing these systems in the long run, as well as considering how key decisions that must be made by a consortium may affect individual states. In this report, the Assessment Solutions Group (ASG) has developed a set of guidelines for states to address as they develop assessment implementation plans. Guidelines for states are addressed and discussed in the following areas:

- How to manage a group of states and develop governance principles for the consortium;
- How states with variability in the experience and capacity of state assessment staff can effectively and collaboratively utilize their in-house expertise;
- How states with different assessment purposes can work together in a mutually beneficial manner;

- How to approach a variety of technical issues, such as equating and linking the new assessments to the current ones in order to report scores on the same scale;
- How to document validity, such as establishing content and construct validity or conducting studies of predictive validity;
- How to jointly establish performance standards for use by all states;
- How to approach security issues common across states;
- How to resolve ownership issues when a group of states share in the development of items and test forms;
- How different states' item release policies may affect technical design decisions such as field testing and equating models, and costs for continuous item development and replacement;
- How states can ensure the feasibility, practicality and sustainability of the project; and
- How states can monitor and control costs.

The *Guidelines for Organizing and Leading a State Consortium* is organized into eight sections that address the following topics and issues:

- I. Introduction
- II. Context for State Collaboration and the Guidelines
- III. Practical Issues
- IV. Technical Issues
- V. Performance Standards
- VI. Project Management and Governance Issues
- VII. Cost Issues
- VIII. Summary and Recommendations

Details of critical decision points in the work of a consortium and suggestions to states on important considerations are presented in the *Guidelines*. The Appendices provide additional information and examples of key tasks, schedules, budget reports, and other examples of resources that may be useful to a consortium.

## **II. Context for State Collaboration and the Guidelines**

The new initiatives noted in the Introduction all emphasize a collaborative approach taken by states working together to develop and implement common standards and assessments, and to share products and resources among states. Over the years, many states have participated in a variety of collaborations, such as the State Collaboratives on Assessment and Student Standards (SCASS) projects run by CCSSO since 1991 and the Enhanced Assessment Grants (EAG) projects funded by the USED since 2002. In addition, in recent years Achieve has conducted an end-of-course assessment development consortium project for states; the first product developed was an Algebra 2 assessment. The power of state collaboration lies in the ability to leverage the combined resources of many states and return a state's investment many times over through shared costs to create new products and tools, to promote program involvement, and to support and provide professional development. State collaboratives also help to bring career service professionals together to solve complex problems impacting states. For example, during the past 20 years, many different SCASS groups have been formed in response to specific state needs; projects undertaken have included addressing issues for students with special needs, policy issues, technical issues and psychometric solutions, and uses of technology to advance research and practice.

In this context, the common assessments being designed by state consortia in response to the RTTT competition will allow states to benefit from the power of collaboration. However, with this benefit come many challenges. Each state has its own unique characteristics, which may include demographic, economic, or geographic diversity. No one size fits all states when it comes to every aspect of an assessment system, so flexibility will need to be designed into the system. States will need to carefully consider all the issues involved while working with a multi-state consortium for the development of assessment components and the systems that will be used by states in the future. Costs for implementing, maintaining, and sustaining the system for many years after the work of the common assessment consortium is completed will need to be carefully considered.

In addition, many lessons have been learned from working with state collaboratives over the years. Two of the authors of these Guidelines (Roeber and Olson) were heavily involved in the SCASS program, from its inception in 1991 to 2003, initiating and directing the many SCASS projects while serving as director of state assessment programs for CCSSO at different times. Together, they have many years of experience in successfully starting and supporting state collaborations and consortia. In this document, the authors wish to share their knowledge with all states and others involved in the common assessment consortia and to provide advice on critical issues that will need to be addressed.

### **III. Practical Issues**

States will face a number of practical issues as they decide how to work together to build common assessments. The following issues are among the most important (see outline of topics below). For each, suggestions for consortia to consider are provided.

- A. Developing RFP(s) for the Consortium
- B. Staffing the Consortium
- C. Developing the Assessment Measures
- D. Assessment Administration
- E. Scoring and Analyzing the Common Assessments
- F. Standard Setting
- G. Reporting
- H. Other Issues

#### **A. Developing RFP(s) for the Consortium**

A key issue for any individual state or group of states is the process of selecting the vendors that will perform important work for the group. It is anticipated that a consortium of states will need assistance with the development of the assessment measures that will be used; the preparation of the assessment administration materials (which might be for paper-based as well as online administration); scoring of multiple-choice, constructed-response, and other types of assessment items; analysis of the assessment results; and the reporting of the assessment information in paper-based and electronic manner. Additionally, a consortium will most likely want assistance in developing and/or procuring technology systems for both online test administration and other components of the assessment system. A consortium might select one vendor that can carry out multiple pieces of these activities, a group of vendors that works together as a team, or a series of individual vendors that is coordinated by the state team and/or a management vendor.

*Is It Possible to Get Competitive Bids?* Because the time that was permitted in the NIA for a consortium of states to put together its bid to the U.S. Department of Education (USED) was short (only 75 days), there was not time for the states to competitively bid their assessment designs to determine which design might be most cost-effectively carried out by which bidder or combination of bidders. The next best thing was for the group of states to develop its assessment design as thoroughly and thoughtfully as possible and to determine the costs of the design.

In the past, such cost determination would be highly speculative, because experts would be trying to compare the costs to comparable programs run by single states.

Until now, there had been no comparable benchmark, however, for states to use to cost out a program with 10, 20, 30 or more states participating. Recently, the Assessment Solutions Group carried out a study in conjunction with Stanford University and the Nellie Mae Foundation in which it estimated the costs of a new, high quality assessment program and then examined a series of cost savings measures, through which such a program could be made affordable<sup>1</sup>. Subsequently, the Bill & Melinda Gates Foundation decided to fund ASG to provide cost estimation services for the various consortia of states organized in response to the RTTT common assessment competition.

Thus, consortia of states needed to arrive at tentative designs as soon as practical in order to permit cost estimates to be prepared and any adjustments to the assessment design to be made. This made it possible for consortia of states to get the next best thing to a competitive bid, and use these estimates in their grant proposals to the USED. The consortia also will be able to use this information when negotiating with the vendors who are selected through competitive bidding after award of the grants from the USED. In addition, in the ensuing period of time between the award from USED to implementation of the assessment four years later, a number of the design features might change and the consortium will need to aggressively manage the change process to ensure the assessment can be delivered at a price close to the amount funded by the grant and ultimately an amount member states can afford on an ongoing basis.

*Should One Vendor, Multiple Vendors, or a Team of Vendors be Used?* This is a decision that each consortium will need to make. There are multiple vendors that can do each type of work (i.e., test development, test administration, analysis, scoring, and reporting). States will need to consider the capacity of different vendors to carry out the work in the volume that will occur when a large group of states decides to assess students at the same time. Can one vendor do all of the work for the consortium? The answer is more likely to be “yes” for a test development contractor, since the size of the consortium will not seriously affect the volume of work for this sort of contractor.

In the case of a test administration contractor—the vendor contracted to do the work associated with the administration of the assessment (such as printing, packaging, shipping materials, etc.)—the volume of students to be assessed in a short period of time could be a serious issue. For example, it is possible that

---

<sup>1</sup> Topol B., Olson, J., and Roeber, E. *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*. Stanford Center for Opportunity Policy in Education, March 2010.

millions of students might be assessed using paper-based booklets that need to be printed, packaged, and shipped to thousands of locations. Can one vendor, even the largest, handle this volume? If, however, a team of different vendors is used to print, package, and ship all of the test books, how does the consortium ensure coordination and comparability among the vendors? Saying that one vendor is “in charge” does not necessarily make the coordination occur well. Thus, the consortium will need to be staffed sufficiently so that actual coordination among the vendors is ensured. In addition, including a liquidated damages penalty clause in the contract for poor performance may help motivate vendors to make sure all deliverables are delivered according to specifications (also note discussion in Section VII Part E).

The same issue is present for scanning and scoring of constructed responses from paper-based assessments. The volume for a large consortium of states is substantial. Can one vendor handle this well? Can multiple vendors coordinate well enough to ensure that student work is scored in the same fashion regardless of which vendor is leading the scoring? Again, the solution to this is to have sufficient staff for the consortium that comparability of scoring and accuracy of score assignment across scoring sites and states is determined by quality assurance checks during the scoring preparation time, as well as during actual scanning and scoring operations.

*If Multiple Vendors are Used, How will Work be Split Among Them?* As mentioned in the previous section, there are several reasons why a consortium might use more than one vendor. This decision could lead to different ways in which work is split among the vendors. Some of these arrangements would be relatively easy to manage, and others could make it far more challenging. Having a division by task might make it easier to make sure that there is internal consistency within the task (i.e., test development) but could create issues at the hand-off points to a subsequent vendor (i.e., the test administration contractor). Using multiple vendors for a task may make the workload lighter for each vendor, but increase the need for cross-vendor coordination.

There are some challenges inherent in an arrangement where one vendor does test development work and another handles test administration. For example, the test development contractor must complete its work to meet delivery deadlines established by the test administration contractor. One way to ensure a smooth transition is to develop a very detailed, comprehensive project schedule that spans both sets of activities, and this schedule is used to monitor the completion (and quality) of the work periodically – no less often than quarterly. Such joint project meetings can help identify issues before they become unsolvable by adjusting the work and the schedule as needed. This will require, at a minimum, that contractor

staff and consortium staff be dedicated to project management. If the number of vendors is large, it may also mean that one vendor or management organization is used to manage the other vendors on a full-time basis.

However, far more complex would be the case where two or more vendors are used to handle the assessment administration activities. In this case, duplicate systems might need to be created – to handle tasks such as scanning, distributed scoring, scorer training, quality control activities, and analysis and reporting, just to name a few. This will place a burden not only on the different vendors to ensure comparability, but it will also burden the states to assure that all states' students are processed in the same manner. A good way to handle such work that needs to be absolutely parallel is to develop and use comprehensive quality assurance materials and detailed steps (e.g., a full system demonstration) to show comparability before the systems are used to process real student results. This suggests that the vendor that received the highest point totals in the competitive bid situation may be awarded the largest volume of the work, with vendors receiving lower point totals being awarded lower percentages of the task(s) to be carried out. Regardless, however, consortium staff will need to monitor all work closely to ensure both accuracy and comparability.

*Can the Same Vendor Do Both Online and Paper-Based Assessments?* The answer to this is "maybe;" one vendor *could* do both types of assessment. Of course, this means that the vendor has to have the capacity to do both types of work on a very large scale. Selecting a vendor that has not done online testing with more than a few thousand students may not be wise for a program in which tens of millions of students have to be assessed online, perhaps at the same time. On the other hand, selecting an online vendor that has limited experience creating, printing, packaging and shipping printed materials but considerable experience in online assessment may not be a good bet, either.

It will be best for the consortium to select a vendor that can do each of these tasks well. It may be best to separate these two types of programs into separate bids, but be open to one vendor that demonstrates the capacity to handle both types of programs. If such a vendor cannot be located, then the consortium would be able to pick the vendor who can do each program well. However, it will be important to monitor and evaluate that the assessment experience of students is comparable regardless of how they are assessed. Should multiple vendors be used, it may be necessary to have one that serves as prime contractor or that leads the work of all of the other vendors.

*How Should Vendor(s) be Selected?* Ideally, the vendors should be selected by competitive bidding carried out by the consortium. Although, as noted earlier, the

time to do this was not available before consortium proposals were due to the Federal government in June 2010, states do have time after award of the grants in September 2010 to carry out such competitions and select the vendor(s) offering the best combination of technical capability and price. To assist in this effort, a standard, comprehensive set of financial data tools should be used to collect and analyze cost data from potential vendors.<sup>2</sup> Using this cost collection software will permit a consortium to gather cost data from potential vendors that are directly comparable. Thus, each consortium should be able to tell which vendor's price is best for the quality that was bid.

It is anticipated that the governing states, as well as key advisors, will review the vendors' proposals and analyze them for technical quality and responsiveness to the states' needs. These analyses can then be combined with the analyses of cost data to permit the contractor or contractor team to be named. This contractor or contractor team should be acceptable to the majority of the states participating in the consortium.

*How will the Contractor(s) be Procured?* The manner in which the contractor or team of contractors is procured will depend on how the consortium is organized and perhaps more importantly, how states plan to pay for the necessary assessments. If states decide that a central organization such as a non-profit professional organization will be the lead entity, it may be possible for states to contract with that entity and not have to use competitive bidding (especially if the central organization has done so on behalf of the states). For example, this approach was done by Achieve for the American Diploma Project Algebra II consortium, where legal precedent was set related to contract law for cross-state procurement. In some cases, one state might conduct the competitive bidding in such a manner that a purchase order can be used by other states to add on to the first state's contract. This was done in the case of the first examination developed collaboratively by Achieve. One state served as lead state and other states added on through a purchase order to the original bid. Another model is for the states to individually release essentially the same RFP and have bidders bid on all of them at the same time. The New England Common Assessments Program (NECAP) is an example of a multi-state consortium that has successfully used this approach to procure a service provider.

This is an important issue for states to work out as they form a consortium. The final approach described above – that of states essentially issuing the same RFP at the same time in order to make a common award – does not seem practical in the

---

<sup>2</sup> Similar to the cost modeling analysis performed by ASG in the study conducted for the Stanford Center for Opportunity Policy in Education (2010), as noted earlier.

case of a consortium of 25 or more states. One of the other approaches may be more feasible if state contracting rules permit them to be used.

## **B. Staffing the Consortium**

As states join a consortium in order to develop required state assessments, the issue of staffing the consortium will arise. There are several ways in which states can organize to see that the required work is carried out. This section will describe some of the ways that a consortium might choose to organize to complete the needed work.

*Recommended Consortium Advisory Structures* – This may be the key issue in the success or lack thereof of a group of states working together. Several functions must be coordinated by the organization that manages a consortium. First, the group of states will need to receive sound policy advice. The consortium must identify exactly what the givens and the prohibitions for the work of the states are and must ascertain the extent to which these givens and prohibitions are the same for all states in the consortium. Policy direction must be provided that is relatively impervious to elections, terms in office, or turnover of key state officials and staff.

Second, technical design and oversight will be needed. Some of this expertise can be provided at the staff level, while advice and review should also be provided by a technical advisory committee (TAC) for the consortium. Third, practical operational assistance will be needed to carry out assessment tasks on such a large scale. Each consortium will need to determine how much of the policy, technical, and practical needs will be met by staff, how much by external advisory committees or groups that will serve to provide advice periodically, and how much by staff hired directly by the consortium or through a management company hired to manage the work of the consortium. These committees might be comprised of individuals from the states being served or others with the needed expertise.

It will be best for a consortium to separate these roles wherever possible. For example, each consortium needs to have policy direction. This should be from state-level policymakers (i.e., governors, chief state school officers, state board of education members, and state legislators and/or their staff) along with a few key individuals with assessment expertise. Such a policy advisory group should avoid detailed discussions of technical issues, dealing with these only when there is a policy implication arising from the technical issue, or more likely, the converse, when a technical issue arises from a policy decision. This group should be appointed from across the entire group of participating states with multiple representatives from each group and with staggered terms. This will permit some new individuals to serve but also assure enough continuity for the group to function effectively. The types of issues that this policy advisory committee might address are presented in Section V.

Each consortium should also have a TAC. This committee should have representation from those with both a technical and practical assessment advisory experience. Ideally, this group would be drawn from individuals who currently serve on state technical advisory committees to assure understanding of state technical issues and how they are related to those same considerations which must be addressed in a consortium. TAC members often are university-based measurement specialists, former or current state assessment staff (assessment directors or psychometricians), or others with measurement or psychometric expertise. Currently, across the entire country, state TACs employ just over 100 different individuals total.

A third function, which might be handled by another consortium committee, or folded together with the consortium TAC, is a group that represents *users and stakeholders* of the assessments and the information they produce. This could include groups such as teachers, elementary/middle school and high school administrators, district administrators, content organizations (i.e., mathematics, reading, and writing), and curriculum groups, as well as organizations that represent students with disabilities (SWDs) and English learners (ELs). Such a group could be excellent at helping the consortium to establish and maintain assessment policies and implementation practices so as to minimize intrusion and maximize the value of the assessment information.

*Consortium Staffing and Roles* – Several types of individuals will be needed by the consortium to carry out the activities of the consortium. In some sense, these roles parallel those of a well-staffed state assessment program. Staff will be needed in the following areas: test design and development; assessment administration; online assessment administration; psychometrics, research, and data analysis; data use and reporting; information technology; assessment of special needs populations (SWDs and ELs); professional development, especially for formative assessment initiatives; and contracting and procurement. Staffing should be set up with an overall staff director and deputy director, along with persons in charge of the functions listed above, and consultants and program assistants in each area listed above. It is assumed that the bulk of the work will be carried out by vendors, so much of the work of the consortium staff will be to determine state member priorities and preferences, and then to work with the advisory committees and vendors to put these decisions into practice.

### **C. Developing the Assessment Measures**

As previously mentioned, it is assumed that the assessments to be used by the consortium of states will be developed by one or more vendors, working under the direction of the states that are members of the consortium and with the advice of the

policy, technical, and procedural advisory groups charged with advising on the development and implementation of the consortium’s assessment programs. As these assessments are being created, several important considerations must be kept in mind. First, the purpose of the assessment system and what is going to be measured by each assessment component needs to be carefully considered. In addition, because the full spectrum of students must participate in the assessment system, the assessment developers must be mindful of the ways in which SWDs and ELs can most fully participate in the assessments.

*Purpose of the Assessment* – The purpose for which an assessment will be used needs to be considered early in the process, as it impacts the design of the assessment and many other important decisions that will be made as plans are developed. Thus, it is crucial to carefully delineate the purpose, or purposes, for the assessment. Typically, there are a number of reasons why students are assessed, and it will be important to determine the purpose(s) for the assessments as they relate to these reasons. Among these reasons are to:

- Inform individual students about their achievement
- Determine whether students in the school are making adequate progress in learning
- Determine whether students have achieved enough to be promoted or to graduate from school
- Predict whether students will be successful in the next stage of school or outside of school (such as college and career readiness)
- Provide diagnostic information on a student’s strengths and weaknesses
- Determine the adequacy of the school’s instructional program
- Determine the effectiveness of teachers

There are many other purposes for which student assessment can be used in an educational setting. It is important to note that no single assessment can be designed for all the different purposes. Also, it is essential for the sake of clarity about the assessments to determine the purposes and specify for users what these purposes are, so that users understand how the results of the assessment will be reported, interpreted, and applied.

*Universal Design Considerations* – Access to the assessment tasks begins with the design and development of the tasks and continues with decisions about how the assessments are to be administered, scored, and reported. Several organizations, such as the Center for Applied Special Technology (CAST), National Center on Educational Outcomes (NCEO), have developed principles for *universal design* of assessment tasks. The design principles developed by NCEO (Thompson, Johnstone,

Anderson, & Miller, 2005) are good examples of best practice. These principles, when thoughtfully applied, permit students to more fully participate in the assessments by increasing the accessibility of the measures. Thus, it is recommended that universal design principles be used by state consortia in the creation of all assessments and component tasks to be used by the consortium.

*Assessment of Students with Disabilities* – Federal law is quite clear that all students with disabilities (SWDs) must participate in state assessment systems. Students with the most severe disabilities participate in an alternate assessment of alternate achievement standards (AA-AAS), which is beyond the scope of this document (and the current work of the consortia designers, although the USED issued a separate NIA for a state competition on the design and development of new alternate assessments). Thus, the new assessment systems now being designed must accommodate 99% of the student population, including approximately 85%-90% of SWDs. Some of these students currently participate in states' alternate assessments of modified achievement standards (AA-MAS) programs, although not all states have implemented these assessments. The consortium will want to consider how these students will participate in the overall common assessment program. Because the best approaches to use for assessing SWDs can be a very divisive and technically complex topic, the consortium also should refer to key documents and guidance from the USED's Office of Special Education Programs.

One of the best ways to ensure that most SWDs participate adequately in the assessment system being designed is to ensure that a full range of accommodations is available to students. States that have developed custom assessment programs typically offer SWDs a total of 50, 60, 70 or more possible accommodations. These include flexibility in assessment administration time, setting, presentation, response, and materials. Many of these changes do not change the nature of the assessment, but simply serve to level the playing field for these students (Olson & Dirir, 2010). However, not all states permit use of the same collection of accommodations. Among the accommodations that states may offer and others do not (and thus are the most controversial) are 1) read alouds of the reading or English language arts tests, and 2) using a calculator on all items of a mathematics tests (even those items for which other students are not permitted to use calculators).

The assessment consortium should clearly use universal design principles in developing and administering its assessment measures. In addition, it needs to come to a common agreement about which accommodations are made available to all SWDs across all participating states. This is essential so that comparable types of students participate in the assessments using comparable accommodations and thus the overall results are comparable across the participating states. Associated technical issues related to the valid reporting of results and the interpretation of test scores

also will need to be addressed. Finally, the consortium needs to determine whether it will create an AA-MAS assessment for students with persistent academic disabilities, assuming that this will be allowed by the USED in the future.

*Assessment of English Learners* – Federal law also requires that ELs participate in the assessments states use for accountability purposes. The first of these is an annual assessment of ELs’ English proficiency during their years being served by programs designed to improve their English proficiency. The USED has indicated that a separate design competition will be held for this program in the future.

The second requirement is that ELs participate in the academic assessments in the language that permit the best assessment of the ELs’ proficiency in that content area, if at all possible. A number of states do not translate their tests, for reasons of cost, philosophy, or state law. Other states do so, at least for the largest language group(s) other than English in their states. Typically, Spanish is the language into which assessments are translated, although individual states can choose additional languages (for example, Michigan also translates its tests into Arabic as well as Spanish). Many states, however, do provide a wide range of other accommodations for ELs.

States that do carry out translations often do oral translations (that is, EL students see the test in English but hear it in another language), whereas other states may translate the tests and publish the translation in a bilingual test booklet, often side-by-side with the English version. Other states design a completely separate native-language test, designed to meet the same test blueprint but with items developed completely in the context of the native language.

The challenge for the consortium will be to make a decision about translations and the use of translated versions of the assessment across the consortium. Given the size of the consortium, resources could be expended to create translations in a number of languages, something that individual states usually cannot afford to do. However, it is essential that the consortium as a whole decide this issue so that if translations are carried out, they are used in a consistent manner across the states so that cross-state assessment reporting is not compromised.

#### **D. Assessment Administration**

*Assessment Administration Window(s)* – Another key decision that states participating in an assessment consortium will need to determine is when during the school year the assessments are administered. First, they will need to decide if there is a consistent time across all states, such as in October or April. Second, they will need to determine how long this assessment window will last. For example, an assessment

window for a mathematics test may be a single day, some day during a set week, at some point during a month, or last for months depending on the school calendars within a state. It may be possible to find an example of each of these test windows among the states participating in an assessment consortium. Third, some states in a consortium might prefer fall versus spring, and with appropriate statistical work to ensure comparability, it may be possible for states to choose from several choices when they administer the tests, so that more than one assessment window might be offered.

Using a common assessment window will make it easier to prepare tests, since only one set of instruments will be required. However, assessing millions of students at the same time will certainly tax any online or paper-based scoring systems that even a consortium of test vendors assembled for this purpose could provide. The perception of fairness (as well as statistical equating) may weigh in on this decision as well; for some states, state policies specify testing is to be conducted only in the spring; for other states, fall testing may be perceived as better and more relevant to measure what students have retained over the summer than test results from the end of instruction the previous spring. A related matter is the start and end dates of school in different states; some states start several weeks before others, so testing during a particular week, for example, may unfairly advantage some states over others.

*Ensuring Security of Assessment Measures* – A key issue, related to the establishment of a test window, for a consortium of states is how to ensure the security of the assessment measures used, especially in the present day world of the Internet where news (and test items) can spread across the country in minutes. States participating in a consortium will need to think carefully about maintaining test security, and thus preserving the validity of test score interpretations. The consortium must collectively determine how long the assessment window will be open for each content area and grade level, and will have to balance the understanding that the longer a test window is open, the more likely that the assessment materials will be compromised. Some solutions might be for the consortium to develop multiple forms of the assessment to be used, and develop an administration plan to offer or distribute the multiple forms. For example, students in adjacent states might take the same test form if testing is done on the same days but different forms if administered on different days (or weeks).

Online administration and even computer adaptive testing is no assurance of security, either. Students who are not monitored closely can copy down or take photos of test questions during or following testing and then share them widely on social media such as Facebook, Twitter, and so forth. It will matter little whether students were tested on paper or online.

The consortium will need to carefully consider its policies and planfully determine how to protect its assessment items and instruments. The solution may well be a combination of a large bank of items for use online and multiple paper-based forms to use both within and across states. In addition, test policies and the length of test windows also will have a bearing on the solution to this dilemma.

*Using a Single Delivery System Common Across All States* – The states participating in a consortium will need to determine if a single delivery system is to be used in administration of the assessment instruments. This might be either an online system or a paper-based one. Using a single system across all states will save the consortium a significant amount of money, since only one system (with the fixed costs inherent in it) will need to be prepared. The challenge for the consortium will be to decide whether to use an online or paper-based delivery system, or both.

*Using Both Paper and Computer-Based Administration* – Of course, in discussions about whether the delivery system should be paper-based or online, some states that desire for the delivery system to be online may not be ready to make the transition completely in 2014-15. States will differ in the extent to which online testing will be feasible when the assessments are first rolled out. Hence, the consortium will probably need to plan for both types of administration, within and across states, and conduct the necessary bridge studies to determine the comparability of the two types of assessment programs.

*Using Shared Standards for Test Material Production, Packaging, and Delivery* – The consortium and/or the vendor(s) will need to develop specifications for the production of paper-based testing materials, including what types of assessment administration and assessment coordinator manuals to produce, and the formats for the production of these materials. States will need to provide specifications to the production contractor for packaging the testing materials by school, as well as determine the most efficient means for delivering the materials to schools (and the collection of answer documents and other used and unused testing materials after testing).

The consortium will also need to determine the formats to be used in an online assessment – the manner in which passages and test items are presented, how students can access various online mathematics tools, how to provide access to various reading tools (e.g., highlighters), and overall accessibility issues for SWDs. Questions that need to be addressed include how will students respond to the questions (e.g., can they return to questions skipped earlier in the test session); can students review items or materials once they are submitted; can students take the test more than once; how are computer-adaptive items administered to the students, can different state delivery platforms be used, how will source codes (open source, vendor’s proprietary systems, etc.) used for item development and delivery impact

the online assessment, and so forth. Specifications for all of these types of test administration policies and considerations will need to be made by the consortium.

## **E. Scoring and Analyzing the Common Assessments**

*In-State or Cross-State Scoring of Constructed-Response Items, Centralized or Distributed Scoring Model* – There are a number of ways in which constructed-response items might be scored. Each of these is briefly described below:

- The states in a consortium pay a vendor or group of vendors to score all of the student responses to constructed-response items. The advantage of this approach is that it does not involve any educators from the states, so presumably scoring could be done very efficiently and quickly. The disadvantage of this approach is that teachers do not get the professional development benefit of reviewing student responses and seeing how students responded to each of the questions. In addition, given the volume of a 25-state consortium, it will tax even the largest vendor to score so many students on so many different items at the same time. It may also be the most costly method of scoring student responses.
- Teachers in a state score all of the student responses provided by the students in their state. This model maximizes the value teachers can derive from scoring student responses. Because many teachers will be needed in this model, school may need to be suspended for several days (perhaps a week or more) and some mechanism provided for training teachers across the state. The disadvantage of this approach, beyond having numerous teachers absent from school for several days, is that teachers do not need to score hundreds or thousands of papers to derive value from scoring. Also, if teachers (or substitute teachers) need to be paid for this work, additional funding will need to be provided. In addition, an auditing system of teacher scoring would need to be implemented by the consortium.
- Computerized artificial intelligence (AI) software is used for scoring some types of student responses. This has the advantage of permitting students' responses to be scored almost as soon as they are provided, which could certainly speed the return of overall results to educators, parents, and students. However, it does not currently save much money, and the professional development opportunities for teachers would be missed entirely. AI scoring has not been widely implemented in states because it still is a relatively new technology. In addition, AI software can not accurately and reliably score all types of constructed-response items; these will have to be hand scored.

- Use a blend of these strategies. There are several ways in which these methods could be combined. First, teachers could score some of the papers and the vendor could score the remainder. Second, AI software could be used for the first (or the second) score, and either teachers or a vendor-supplied team could provide the other scoring. Because of the costs and benefits of hand scoring, this is an area that the consortium should carefully plan.

When scoring includes classroom teachers whose students were assessed, there are multiple approaches to ensure accurate and reliable scores:

- Teachers in each state would score the papers from within their own state and perhaps even from their own district. Samples of the papers from each participating school district would be adjudicated at the state level, and then papers from each state would be adjudicated nationally to determine the comparability of the scoring across the states. States would be encouraged to “adjust” their grading difficulty over time to eventually arrive at comparability through this social moderation process. Teachers could be asked to score papers from other districts and other states as a mechanism to make sure that all papers are being scored fairly.
- Using image-based methods now common in paper-based systems (and that is also applied to online assessments), the images of student work from state X could be sent to teachers in state Y (as well as states Z, A, and B) for scoring. Each teacher would be required to score student work for a certain period of time (e.g., three days) or a certain number of student papers (e.g., 100), but these would not be students from their state. Teachers could be trained online and the reliability of their scoring could be determined as they score. This would have the advantage of exposing teachers to high quality (and low quality) student work collected across the country, thus permitting teachers to develop more robust internal standards about quality student work.
- Actually mixing up student papers from different states and sending the paper-based responses to teachers in different locations would prove to be inefficient and a logistical nightmare. This approach is not recommended.

*Analyzing the Results and Handling of the Psychometric Work* – Analysis of the data from the common assessments would include conducting item analyses, DIF, equating, and scaling, at the least. Item Response Theory would likely be used, although the particular model to be used could be a source of contention within a consortium. Depending on the decisions made by the consortium regarding reporting, status and growth measures may need to be developed and analyzed. In addition, a standard setting will need to be done so that these results can be reported in the first

year of the operational program. Thus, these are activities that states will undoubtedly want their vendor(s) to handle for them, although the consortium may well have psychometric staff that can be used to verify the accuracy of the work. With the right level of staffing, this is work that the consortium staff could carry out, but it is recommended that at least initially, this work be done by the vendor(s) and verified by consortium staff and the technical advisory committee.

## **F. Standard Setting**

It is assumed that the new common assessments will have performance standards. Therefore, the consortium will need to plan for conducting a standard setting to establish common performance standards. This is a critical issue for a consortium to consider – the establishment of a common performance continuum and cut scores that will apply to all states. In order to report results that are comparable across all states in a consortium, and to interpret the performance of students in a meaningful way, states will need to create a common set of achievement or performance standards that describe what students know and can do at various levels on the reporting scale. Psychometricians assigned to the consortium will need to develop a plan and specify the procedures to be used in the standard setting and TAC members will need to review and approve it. The topic of common performance standards is discussed in more detail in Section V.

## **G. Reporting**

*Common Score Reports for All States* – The consortium is required to provide common reports of results – at least for overall performance on the assessment of each content area. State and federal laws also impact decisions about reporting. Beyond describing what proportion of students achieved a score of proficient or above, the consortium may wish to develop other types of reports, including those designed for students, parents, teachers of the students assessed (individual student, classroom rosters, performance on the skills assessed), all educators in the building (school summaries by skills assessed and student roster), item analysis reported, and so forth.

States also might have regional educational units (such as Intermediate School Districts, Regional Educational Service Agencies, or Boards of Cooperative Educational Services) for reporting, and will want statewide summaries of performance by skill assessed, and lists of performance of the districts within the state and buildings as well. Electronic files of student, school, and district performance will also need to be prepared. The vendor(s) hired for this purpose can prepare the reports and distribute these to schools and/or districts and to the state. Distribution in most cases will be electronic, such as through the use of Adobe Acrobat PDFs and the use of a secure

website. This will save both time and money over the preparation, packaging, and shipment of paper-based reports to schools in all districts and states in the consortium.

The consortium will need to work to create the formats for these reports and to make sure that they are readable and useful to the intended audiences. One way to do this is for draft reports to be shared with focus groups of individuals in the intended audience (e.g., parents) and incorporate the comments from the focus groups on how to increase the utility and usability of the information contained in the reports.

*Unique Reports for Individual States* – In addition to the common reports provided to member states, individual states or sub-groups of states might want to have specialized reports. For example, in addition to a general report of assessment results for parents, some states might want to provide parents with a detailed report on how they could help their child improve in the skills tested or in the content area. Each state will need to prepare both standard and unique accountability reports. Individual states should be free to pursue such specialized reporting so long as the cost of developing and implementing such specialized reports is borne by the state(s) pursuing such reporting, not the consortium as a whole, and that the programming for such reports and their production does not impede the production of the common reports for the consortium.

## **H. Other Issues**

*Forging a Common Item Release Policy and Schedule* – This can be a thorny issue to groups of states working in a consortium. Some states need to keep the measures secure during and after testing, choosing not to release any items. Other states release some (e.g., 10%, 20%, 50%) or all of the items after testing to aid in test score interpretation. Some states are legally mandated to release items each year. How can these differences be accommodated? One solution is for the states that want to release some or all of the assessment following testing to use a different form of the assessment than those who desire to keep the assessment secure after testing. The form that is released might be one that has been used previously (once or twice) and thus has come to the end of its “useful life.” Another solution, of course, is for the consortium to adopt rules that prohibit keeping the assessments secure after use, or ban the release of tests or test items following administration. This, however, will violate sunshine laws in some states, so a more flexible and accommodating policy is undoubtedly better. A final option is to plan for the release of one or more test forms annually and to build this into the test development budget for the project.

*Assessments in Other Content Areas Beyond ELA and Mathematics* – Because many states assess content areas beyond English language arts and mathematics, the

consortium may want to eventually plan for the creation and use of measures in other content areas. The most popular of such areas will undoubtedly be science and social studies (or component elements of social studies such as U.S. history, world history, economics, and/or geography). It will be natural for such additional work to be delayed until the assessments in ELA and mathematics are completed and being used. One method for pursuing these other content areas is to determine which states are interested in such measures and having only these states “chip in” to develop them. From simulation studies carried out by the Assessment Solutions Group, an assessment developed by a 10-state consortium can be delivered at substantial savings over what it would cost an individual state to do so.

## IV. Technical Issues

When designing and developing a state assessment program, many issues need to be addressed concerning the technical characteristics of the system. The AERA/APA/NCME Joint Standards (1999) spell out many of the technical issues in detail, as do the NCLB peer review requirements issued by the USED. However, the focus of those documents is typically on assessments used in individual states or districts, and not collaboratively developed assessments. In this section of the Guidelines document, the technical issues as they pertain to states working together in a consortium to design, develop, and implement common assessments are addressed.

As noted earlier in this report, the CCSS, RTTT common assessment competition, and CCR standards initiatives are currently impacting states and their assessment programs and will likely have a major effect for many years to come. All of these things are having a direct and significant impact on the planning of the state assessment consortia, both in designing the assessment system as well as in implementing it in the future. In particular, there are many technical issues related to the design and implementation of an assessment system that a consortium must address.

In this section, the following technical issues are discussed.

- A. Designing the Assessment System for a State Consortium
- B. Multiple Purposes for the Common Assessment
- C. Types of Assessment Measures that Can Be Used by the Consortium
- D. Creating Assessment Blueprints and Item/Test Specifications
- E. Adding Individual State Content to the Assessment System
- F. Comparability, Equating, and Scaling Issues
- G. Using Technology in the New Assessment System
- H. Establishing Content, Predictive, and Other Types of Validity
- I. Providing Technical Guidance to Individual States and the Consortium
- J. Steps for a State Consortium in the Development and Implementation of a New Assessment System

### A. Designing the Assessment System for a State Consortium

One of the initial and most important steps that will need to be taken by a state consortium is to agree on the purpose(s), scope, and design of the assessment system. In the ongoing deliberations among the group of states, it will be crucial that the common assessment meet the most important needs of the states participating in the consortium. Since each state may have its own unique needs, a

consensus will need to be reached among the group on how best to design the final product(s) from the consortium. Decisions will need to be made on a number of critical issues, including the types of assessments that will be part of the system (i.e., summative, interim/benchmark, or formative), the variety of item types to be used (such as multiple choice, computer enhanced, short constructed response, extended constructed response, performance events, or performance tasks), the length of the tests, the scores and subscores to be reported, the types of scales and subscales to be used (vertical, horizontal, or both), the types of reports and their audiences, and many other issues related to the design of the assessment. The final product(s) from the consortium will need to be carefully designed so that all states can use them as is, or be flexible enough that they can be easily modified by individual states if needed, since it is likely that the design may change depending on key decisions made by the consortium as well as possible changes required by a reauthorized ESEA.

A number of key characteristics of high quality assessment systems have been identified in the announcements for the Race to the Top competition or described by various experts that have provided input to the USED on this regard. As described in these documents, the assessments should:

- Measure college and career readiness
- Provide actionable feedback throughout the year
- Measure individual student growth
- Contain various item types including innovative items
- Reflect and support good instructional practice
- Be accessible to the broadest possible range of students
- Include appropriate accommodations for SWDs and ELs
- Produce results that can be aggregated at the classroom, school, LEA, and state levels
- Produce reports that are accurate, relevant, and timely, and displayed in ways that are clear and understandable for target audiences
- Make effective and appropriate use of technology
- Be valid, reliable, and fair
- Include an appropriate level of test security
- Have fast turnaround time on scoring
- Involve teachers in scoring certain CR items and performance tasks
- Can be maintained, administered, and scored at a cost that is sustainable by states over time

The common assessment will need to be designed and developed so that it is technically sound. It is essential that accurate information be collected for all students and at all levels in the educational system: individual, classroom, school,

district, and state. The assessment system will need to meet all standards for technical quality such as reliability and validity. The system also must provide dependable data that accurately reflect student learning, for both individuals and groups of students. The reporting scale that is used for the assessment needs to be able to adequately differentiate student performance across all levels of ability. A common metric and scale that all states can use will need to be developed for the common assessment.

In addition, it will be important to document the design aspects and technical quality of the common assessment. Each of the assessments that will be developed needs to have written reports describing how technical requirements will be met so that users can be ensured that adequate quality is built into the system from the start. After the assessments are developed, documentation of quality should be done via written technical reports, peer review of the programs, and/or external reviews of the assessment system. The consortium should have a comprehensive technical report written of the procedures used in developing the common assessments and include statistical data on all the items and forms created. All of these approaches will help ensure that the assessments are technically sound and that decisions to be made based on the assessments are likewise sound.

## **B. Multiple Purposes for the Common Assessment**

As discussed earlier, the purposes for which an assessment will be used need to be considered early in the process, as they impact the design of the assessment and many other important decisions that will be made as plans are developed. It will be important to carefully delineate the possible purposes for the assessment and how the different components will be used to let the users know so they fully understand how the results of the assessment will be reported and used.

For the common assessments, it may be important to accommodate several different assessment purposes within the consortium. For example, all states will need summative assessments in ELA and mathematics that will be used for accountability purposes, but some states may wish to use these for promotion or graduation purposes as well. Not all states may need or want interim or formative assessments and thus would not use these approaches if they are developed by a consortium. States may want to be able to administer the assessments online, but some may not be able to fully implement this by 2014-15. Therefore, the consortium will need to allow some flexibility in how individual states will use the assessments that are developed. Components may need to be designed in modules so that states can use some and not others. Test forms may need to be provided in both paper and pencil and online delivery modes. Such variations in the system need to be built in from the beginning so that different state needs and uses can be

accommodated and all test results can be interpreted in the same way. Finally, all assessment uses will need to be validated and substantiated in case of any legal defensibility challenges.

### **C. Types of Assessment Measures that Can Be Used by the Consortium**

In recent discussions on the different types of assessments that could be used for different participants in the educational system, a systemic approach has been described in which the needs of different users are balanced and a variety of assessment types are used to meet these needs. In creating an assessment system, the uses of different types of assessments will need to meet the needs of different participants in the educational system (such as school superintendents, principals, teachers, students, and parents). The needs of different users can be met by a variety of assessments for the different purposes of users.

For the common assessment consortium, a comprehensive assessment system will need to be created that includes various components and approaches. This system may include the following types of assessments:

- Summative – Assessments that will serve to assess student proficiency at the conclusion of each grade or other periods of instruction. They would include end of course or end of domain assessments used in high school.
- Interim/Benchmark or through-course assessment – Assessments that can be used during the school year at the conclusion of units of instruction or periodically throughout the year.
- Formative – Assessment approaches used by teachers during instruction to ensure on a daily or weekly basis that students are learning the expectations denoted in each common core state standard as instruction is occurring. Formative assessment opportunities should be connected to learning progressions and are part of a teacher’s professional development and practice.

The consortium will need to decide which components to include in its system. At the least, a common summative assessment will need to be developed for use by all states. The use of interim/benchmark and formative assessments, although they promise to add value to the overall system, may not be wanted by all states. However, if they are included in the system, detailed information will need to be specified on exactly how they will be used, how often, and how they are connected (aligned) to the summative assessment.

The use of a variety of assessment methods and item types by the consortium can help ensure that a range of information is gathered on whether students have learned the important aspects of each standard and exhibit the knowledge and skills required to perform at the designated levels. At the very least, assessments should include many if not most of the following types of items:

- Multiple-choice (selected response) items to measure the knowledge required by the types of assessments listed above
- Gridded response items that permit responses to carefully designed machine-scorable items that require students to make a graph or plot or draw a diagram within certain constraints
- Short constructed response items to permit students to show their understanding of concepts
- Extended constructed response items to permit students to show how they can apply their understandings in novel situations
- Performance events (on-demand performance assessments) as appropriate for students to show that they have procedural knowledge beyond just conceptual knowledge
- Performance assessments, consisting of projects and tasks that are completed over time, to permit students to demonstrate their understanding in a rich authentic manner

It is strongly recommended that the assessments use the full range of item types. In addition, portfolio assessment also may be considered by the state consortium since it can encourage educators to collect an even wider range of information on student achievement as well as allow a way for students to take more ownership for their own learning.

A variety of assessments should be used in the assessment system to determine if students are able to carry out the tasks implied by the Common Core State Standards and can exhibit an adequate level of knowledge and skills in the content and domains being assessed (e.g., the procedural knowledge necessary to perform the tasks). Developing these different types of assessments together in a coherent and systematic manner will help ensure that all the various parts fit together and provide useful information to classroom teachers, students, and others.

#### **D. Creating Assessment Blueprints and Item/Test Specifications**

After the content standards have been developed and approved by participating states and a general design has been created for the assessment, one of the next steps to be done by the consortium is to create an assessment blueprint for each of the components. The blueprints will set up a framework for measuring the

constructs; identify which standards will be selected for assessment; identify the target population(s); describe the intended inferences to be made; provide guidance in constructing test forms and include information on the percentages of items that will measure each of the content standards and domains; and contain information on the numbers of items, the types of items, etc. Blueprints also will need to address the proposed performance level descriptors and performance standards that will be set for the assessment (see the section on performance standards for more information).

In addition, a related document needs to be created that provides details of the specifications for the test items and forms that will be developed. These specifications should address technical and statistical requirements for developing the items and test forms, include information that includes the total number of score points that will be used, scoring rubrics, anticipated amount of time it will take students to respond to each item type, amount of time that will be needed to take the entire test (e.g., one or two class periods), and targeted difficulty levels of the items and forms. It should provide clear examples of different item types and test forms that will help inform the states and other users.

In addition to the summative assessments in mathematics and ELA, blueprints and specifications also should be created for the through course or interim/benchmark assessments, if they will be part of the system. In addition, specifications for the professional development needed for teachers to learn to use formative assessment as they teach should also be developed. All of these documents are critical for the state consortium (or its vendor) to develop early in the process, and it will be critical to reach agreement by the group on these technical issues as they will serve as guides for all development work that will be done for the common assessments.

### **E. Adding Individual State Content to the Assessment System**

States and/or the consortium may want to measure some content or attributes beyond those covered in the CCSS. States were informed that they will have some flexibility in the design of the common assessments. As was stated by the NGA/CCSSO, a state may include up to 15% of state-specific content standards in addition to the CCSS in order to measure additional standards that are specific to their state. The common assessments may need to be designed so that a state can assess this additional 15% in the state content standards, but without jeopardizing the ability to report comparable score information across all participating states. Some states may want to have unique components included in their assessment, such as additional items that measure certain state standards that are not part of the CCSS. Thus it may be important to design an assessment that contains a core section of test items that are common across all states and a noncore section that

allows for an individual state to add other items or unique components (if any) in the test form. These additional items would need to be placed separately from the core set so they did not impact the ability to report results that can be compared across all states.

## **F. Comparability, Equating, and Scaling Issues**

The goal of having test scores that are comparable across all states using the common assessment is a critical aspect of the summative assessment. However, there are a number of technical and logistical issues that are related to this goal. The equating process is used to ensure comparability of scores on the scale used for reporting across different forms of an assessment. In order to have comparability and be able to equate, a host of other requirements will need to be addressed. In a letter report to the USED on the RTTT (BOTA, NRC, 2009), it was stated:

“If states indeed adopt fully common standards and develop common assessments, these concerns would be reduced, but even seemingly small deviations from fully common standards and assessments will introduce important limitations in the degree of comparability of outcomes. For instance, to be fully comparable, the assessments would need to be administered under standardized conditions across the country. This means that time limits, the length of testing sessions, the instructions given to test takers, test security provisions, and the use of manipulatives and calculators would need to be the same everywhere. The test administration would need to be scheduled at an appropriate time during the school year such that all the students who will take the test have been exposed to and have an opportunity to learn the content covered on the test.” (p. 12).

So, in order for states to achieve a comfortable level of comparability, plans for test administrations will need to be carefully coordinated across states. The consortium also will need to address issues related to equating different forms of the assessment, within and across testing years. Standard procedures exist for doing this type of work and they should work well with the summative and interim assessments. Additional work may need to be done to examine procedures for ensuring the comparability of the performance tasks across forms and years, if that is desired, since these types of assessments and items can be more challenging to equate. States also will need to examine how well the professional development being used to teach educators to use formative assessment has occurred across states.

Other technical issues related to equating are decisions on whether to use an internal embedded item set, an external equating model, a linking model, or some other approach; to use a pre-equating or post-equating model, etc. There are advantages and disadvantages to each of these approaches, and the consortium will need to decide on what is best for its system and all participating states. A design will need to be developed that allows for continued equating of new test forms in years to come, and that is easily sustainable by individual states if needed. Depending on the equating design that is chosen, the test forms may need to allow for use of common (anchor) items to be embedded, which affects the design that will be used by the consortium for the forms.

Another issue is whether an individual state will want to link their new common assessment to the existing ones currently being used. This may be important if a state wants to maintain its reporting scale in order to provide results that are comparable to past years and to continue the reporting of trends or AYP. Methods exist for linking different tests and different scales, and these may need to be considered and implemented by states on a case by case basis in the first year of operational use of the common assessment (2014-15). The consortium may want to help states with linking designs and an appropriate methodology that would work with different states.

One other issue for consideration by the consortium concerns development and use of a horizontal or vertical scale, or both. A horizontal scale will allow for within grade reporting and comparisons of results for a given grade from year to year. Given the interest in measuring growth, the summative assessment also should support cross-grade comparability of scores. There are a variety of cross-grade or vertical scaling models that have been implemented in the country, and they should be examined by the consortium for possible use with the assessments for grades 3 through 8, and possibly high school. However, there are many technical challenges associated with establishing a vertical scale, and some would argue that it is impossible to do well. Therefore, the plans for development of a vertical scale will need to be carefully considered and evaluated by the consortium. A good overview of vertical scales can be found in *Educational Measurement, 4th Edition* (NCME & ACE, 2006). Although measuring growth at the high school level may be more of a challenge, a model could be designed to do this, too. The consortium will need to look at the vertical articulation of the standards and how they relate to both the proposed horizontal and vertical scales.

Concerning the issue of measuring growth, since this is a crucial aspect of the requirements of the RTTT common assessments, it will be important for the consortium to define exactly what it wants to have measured, how the measurement will be done, and what information will be reported from the

measure(s). Various models and approaches exist for getting measures of student growth, for example within grade or cross-grade scales, and value added models, and a number of states have implemented their own growth models for use with their current assessments for NCLB or state-level accountability. States' models vary in their data requirements and statistical complexity. The pros and cons of using a specific growth model for a common assessment will need to be weighed by all the states in the consortium, and consortium members should seek the help of their TAC members or psychometric advisors when dealing with these various issues.

## **G. Using Technology in the New Assessment System**

An important aspect of the new common assessment system is how technology will be used by all states and to what extent. The RTTT competition encourages increased innovation and expanded use of technology by states. However, for the implementation of the assessment system across participating states, a variety of issues related to technology will need to be addressed by the consortium. Some of the areas are:

- *Use of computer-based testing (CBT)* – Decisions will need to be made by the consortium on the use of a linear-administered CBT. Some states currently are using CBT for their assessments; many others are not. It is likely that future assessments will be delivered totally online. A number of issues will need to be addressed, such as having enough PCs in a given state for administering the assessments online during a tight testing window. This will be important to address on a state by state basis as well as deliberated by the entire consortium. States may want to consider various models using CBT or computer-adaptive testing, or both, for implementation in states.
- *Use of computer-adaptive testing (CAT)* – CATs are useful for a number of reasons and the summative assessments may benefit from this approach as CATs may allow for a shorter testing time than linear testing and the use of item pools that cover a wider range and more rigorous standards. They could possibly be used to identify standards that are typically more difficult for students and can be used in extended testing windows while helping to maintain higher security. However the use of CATs is relatively new in states and use of this approach across an entire consortium will need to be carefully designed and implemented.
- *Factors affecting the decision to accommodate both online assessment and paper and pencil testing (PPT) testing modes* – In some states it may be difficult to build sufficient capacity for *all* districts and schools to implement a fully online assessment by 2014-15. Factors that need to be considered include student-to-PC ratios, access to broadband/high-speed Internet,

network and bandwidth capacity, infrastructure in schools, and financial resources. Another issue is whether all parts of the assessment system, particularly some types of performance tasks or through course or interim/benchmark assessments, can be administered by computer. The consortium should do what it can to move toward a 100% CBT or CAT system, but it should also take steps in planning what options may be needed to help states that are not ready to achieve this goal. Therefore, the planning may need to consider having the common assessment administered in both CBT and PPT modes for a limited period of time, which means test forms would need to be printed and distributed to some schools as well as delivered online. This may also require states to study whether any score adjustments are needed for students who are assessed using the PPT format.

- *Use of a common test administration platform across states* – For optimal use of all components in the common assessment system, a technology platform that has a common user interface/look and feel across the states is crucial. The platform should be designed to be used to deliver the summative assessment, through course or interim/benchmark assessments, and/or curriculum-embedded (formative) assessments to students. It also should be able to deliver performance tasks of higher-order abilities online. A comprehensive platform also could be used by item writers for submitting new items, reviewing and editing of items, and the storage of items along with classification and statistical information for future use. In addition, the platform should be designed for use in scoring all multiple choice and some constructed response items, to deliver student responses for other tasks to trained scorers or teachers for scoring, for use in training and calibration of scorers, and for reporting purposes. For the latter, the platform should allow for efficient aggregation of scores and results that support reporting to different audiences.
- *Use of artificial intelligence (AI) systems for scoring* – The use of automated or artificial intelligence systems for scoring has not yet been widely implemented in state assessments, but as the programming and technology continue to improve, the potential exists for more widespread use of these systems in the future. As discussed in the previous section, the consortium should consider the use of AI for scoring the common assessments, and work toward developing and implementing such systems, since the benefits include much faster score turnaround and lower costs than having all constructed responses scored by humans. The use of AI for scoring does not preclude having educators do some of the scoring for professional development reasons to improve their understanding of the types of responses made by students to the assessment tasks.
- *Planning for the use of other types of technology, like handheld devices, tablets, or smart phones* – If possible, the consortium should examine the

possibility of using other types of technology for the common assessment, since new tools like smart phones and handheld computers are becoming more widespread and students are becoming increasingly familiar with the latest technological inventions. It may be possible to integrate the use of these devices into the future assessment. The system also must be sufficiently flexible to adapt to new technologies that may not have been invented yet.

- *Selection of an online vendor and key evaluation points* – A number of issues will need to be considered by the consortium in selecting the vendor(s) that will provide the online system and platform for the common state assessments. The consortium should work with its technology advisors and others to decide on the details of the requirements, specifications, and capabilities of the technology platform. Some of the key points that will need to be evaluated are
  - How the platform supports assessment and item development; how it will be used by states, districts, schools and possibly teachers; how scalable the system is.
  - How the platform supports test administration, scoring, and reporting for all states in the consortium; how it will work with a distributed scoring model if that is desired; how flexible it is for possibly different needs in states for reporting unique data for an individual state.
  - How adaptable the platform will be for use with various through course or interim/benchmark assessment approaches; how capable it will be in delivering many different types of innovative performance items.

## **H. Establishing Content, Predictive, and Other Types of Validity**

As the different components of the assessment system are developed, piloted, field tested, and readied for operational use by states, the consortium should have plans in place for validating each of the tests. There are a variety of possible validity studies that could be conducted by the consortium. For example, alignment studies will need to be done to determine the match of the items to the Common Core State Standards and to the established performance standards. After field testing, it may be possible to conduct factor analysis or structural equation model studies to examine the constructs that the tests are measuring. It also will be important to conduct studies of the CCR standards and how the assessment results relate to what is defined as success in college or career. These may be alignment studies of the common assessment to college and career standards, or predictive validity studies in which results from the common assessments are used with college students or new workers. Other predictive validity studies may look at the relationship of the common assessments to other tests that may be given in states, for example the ACT or SAT, or even NAEP. It will be important to have additional

information that can be used to compare results and progress over time. Many other types of studies can be done, for example to examine predictive validity or the intended and unintended consequences of the assessment system.

In order to implement a validity studies research agenda, the consortium may want to establish an independent advisory group or organization to design the studies, conduct the analyses, and report its findings. Validating the type of information that will be reported and the interpretations of the test scores from the summative, interim, and formative assessments should be incorporated as part of the plans of the state consortium.

### **I. Providing Technical Guidance to Individual States and the Consortium**

As described in the previous section, a Technical Advisory Committee should be convened to guide the design, development, and implementation of the common assessment system. Most, if not all states have their own TACs, and some of the members serving on them may be recruited to serve on a "Super-TAC" that would advise the entire consortium on technical issues. These TAC members should have a cross-state perspective and represent the areas of psychometrics and measurement, policy and practice, special populations, and large-scale assessment in general. This Super-TAC would need to begin work as soon as the grant/cooperative agreement from the USED is awarded (if not already started in the proposal development phase) and continue to meet periodically throughout the entire length of the four-year period of development work.

In addition, the Super-TAC may need to interact at times with individual state TACs to coordinate plans concerning the myriad of issues where the common and state assessments intersect, many of which have been described previously. Good communication, sharing of information, and joint problem solving across all states and advisors will be important as the project moves forward.

The consortium will likely need to include other advisors in the process, such as experts in the areas of technology, assessing ELs and SWDs, formative assessment, professional development, higher education, and other areas. The input of leaders in the field will only strengthen the design, development, and implementation of the common assessment system.

### **J. Steps for a State Consortium in the Development and Implementation of a New Assessment System**

A list of possible tasks and activities that states will likely need to conduct as they develop a new common assessment system is provided in Appendix A. The tasks

include a variety of development, technical, and operational activities. Note that the steps included here are to help provide general guidance to the state consortium as it begins down the path of planning for and implementing new assessments in 2014-15.

## **V. Performance Standards**

In this section, issues related to determining performance standards and setting common cut scores for the consortium of states are discussed. Although this topic has been mentioned in other parts of this report, it is an important issue for a consortium to consider, thus a summary of some of the crucial issues is presented separately in this section. The following topics are addressed.

- A. Issues and Challenges for Common Performance Standards
- B. Performance Level Descriptors
- C. Methodologies for Setting Cut Scores
- D. College and Career Readiness Standards
- E. Transitioning from Current State Performance Standards to Common Ones

### **A. Issues and Challenges for Common Performance Standards**

One of the most important issues for a consortium to consider is the establishment of common performance continuum and cut scores that will apply to all states in the consortium. In many ways, this is at the heart of the purpose of establishing a consortium. It also may be one of the thorniest issues, at least politically, that the consortium will encounter. In order to report results that are comparable across all states in a consortium, and to interpret the performance of students in a meaningful way, states will need to create a common set of achievement or performance standards that describe what students know and can do at various levels on the reporting scale.

The common performance standards will likely need to be benchmarked to international standards. The internationally benchmarked standards could possibly be used as key descriptors of goals in a standard-setting process. Reviews and evaluations of the common performance standards and their relation to international best practice, as well as the alignment of the assessment to the standards will be required. The consortium may want to make comparisons to international assessments like the Trends in International Mathematics and Science Study (TIMSS) and Progress in International Reading Literacy Study (PIRLS).

### **B. Performance Level Descriptors**

For each performance level selected by the consortium, the group will need to create a common definition of the performance levels. This starts first with the selection of the number of performance levels to be used for each content area.

Although four performance levels is the most common number currently in use in states, some states have fewer (three, which is the current Federally mandated minimum number of levels) or five or more.

Second, the consortium will need to determine the labels to give each performance level. This can be a trickier issue than it first appears, because some labels currently in use may not be acceptable to all consortium members. Safe labels such as “proficient” or “advanced” should probably be used.

Third, a representative panel of content area specialists and others should meet to define the performance levels. Then, since these levels and labels will be used by the consortium when the common results are released, it is important that all consortium members have an opportunity to review and approve the descriptions of each of the performance levels, and the labels associated with them.

These PLDs will help the individuals who will engage in standard setting to determine what levels of achievement of students should be assigned to each performance level. The consortium should agree on a common set of PLDs before they start standard setting so that the definitions for “proficient” performance have the same meaning across all states.

### **C. Methodologies for Setting Cut Scores**

Standard setting needs to be considered when designing an assessment. Using appropriate methods and sources of information to set standards will be critically important. The validity of arguments based on results from the assessment cannot be any stronger than the performance standards permit. Therefore, assessment designers should ensure that all crucial evidence is collected and reviewed regarding key topics, such as what types of text materials should students be able to use in order to succeed at the next grade level and what high performing students from around the world know and can do in various grades.

To develop a common set of performance standards, test data from the first operational administration of the common summative assessment done in 2014-15 will need to be aggregated across all states and standard-setting meetings for mathematics and ELA (reading and writing separately) be conducted. It may be possible to set performance standards using field test data from the previous year, but this will need to be carefully considered and implemented in order to get standards that are appropriate for reporting in the operational years. Regardless of when the standard setting is conducted, the process will need to include large panels of people with representation from all states and other stakeholders in the process and data that are accurate of student performance.

There are many methodologies that could be used to conduct a standard setting. It is not the intent of these guidelines to present the merits of different approaches that could be used by a consortium. An excellent summary and discussion of setting performance standards is included in *Educational Measurement, Fourth Edition* (NCME & ACE, 2006).

The consortium will need to decide in advance how many performance levels they will want to have, whether they be “NAEP-like” or not, and what they will be called (such as Below Basic, Basic, Proficient, Advanced or Levels I, II, III, IV). Clear examples of items and student performance at each of the levels will need to be shown. Consensus will need to be reached by all states on this critical aspect of reporting the results for the common assessments.

#### **D. College and Career Readiness Standards**

One aspect of the common assessments will be their need to be able to predict student readiness for college and/or career. This implies that the performance levels of the high school assessment will need to be statistically linked to other assessments of college readiness, be they college entrance tests, college placement tests, or workplace skills assessments. This will certainly add challenges to setting performance levels for the high school assessment as well as validation of these performance levels through external validity studies.

#### **E. Transitioning from Current State Performance Standards to Common Ones**

Since it is likely that the newly created performance standards will be different from any given state’s current ones, the consortium may also need to help states in the development of communications that can be used to explain the differences to parents, teachers, and others, as well as transition from one set of standards (and accompanying performance level labels and descriptors) to a new set of performance standards. A similar process will be required to help states explain the scaled scores and results for the common assessment since it is likely that the reporting scale will be different than the state’s current one.

## **VI. Project Management and Governance Issues**

In this section, issues related to the management of the consortium and governance of the groups of states are discussed. The following topics are addressed.

- A. Ensuring Strong Leadership for the Consortium
- B. Creating and Using a Policy Advisory Committee to Set Policy for the Consortium
- C. Selecting and Using Committees and Advisory Groups
- D. Ongoing Decision Making
- E. Use of Third Parties
- F. Indicators of Success After 1, 2, 3, or More Years
- G. Membership Fees in the Consortium
- H. Ownership of the Assessments
- I. Consortium Development and Implementation Schedule

### **A. Ensuring Strong Leadership for the Consortium**

Strong leadership is necessary for the consortium to succeed. This means several things in the context of a consortium involving a large number of states that have not worked together previously. First, it means that a small number of individuals, representing key attributes of all of the states, must decisively lead the group. Ideally, this small group will represent the diversity of the states (in terms of size, population, and other demographic characteristics) as well as the range of uses of the assessment information that the consortium will produce. This group needs to provide decisive leadership and be willing to make crucial and timely decisions so that the project remains on schedule.

Second, the leaders need to be inclusive. This means that wherever and whenever possible, decisions for the consortium need to be made based on the consensus of the entire group of states. The leadership group should seek advice from the entire group of states so that every state feels that its concerns have at least been heard if not heeded.

Third, the leadership group needs to communicate often, and in detail, what is happening with the project. Since the leadership team will be working with the vendor(s) to carry out the tasks on a week-by-week basis, they need to communicate frequently to the rest of the states so that every state feels like it is “in the loop.”

Fourth, the project leaders need to commit the time and energy to remain actively in charge of the project, so that they do not permit the vendor(s) to run the project nor

make key decisions without the consortium's input and consent. The time commitment may include not only routine project management but also emergency times when crises arise, or some step in assessment development does not proceed as planned. Thus, those individuals who volunteer to be leaders of the project need to be able to commit on-going and emergency time to the project. If individuals are unable to make such a commitment (and it would be understandable if this is the case), then the consortium may need to hire full-time staff directly or via a management contract.

Finally, a most critical position of leadership is the project manager hired by the project to oversee the work of the project's vendor(s). This individual or organization needs to be tireless in actively managing the contract(s) for the consortium, ensuring that things are proceeding on schedule because of frequent, in-person inspections of contractor work.

## **B. Creating and Using a Policy Advisory Committee to Set Policy for the Consortium**

As discussed in Section III, it is essential that the project have a policy advisory committee to help guide its work. Because a change in leadership in one state could change the direction that a particular state participating in the project might take, it is essential to buffer against such changes possibly affecting the entire consortium adversely by using a broader, more representative group that represents the policy perspective across all the participating states. Issues that such a group might take up include

- Overall assessment model – summative, interim, and formative assessments
- Manner in which the assessment results will be reported
- Appropriate and prohibited uses of the assessment results
- Professional development plans and activities
- Time(s) of the year for assessment
- Fees for late-entering states
- Item release policies
- Other content areas and/or grades to be assessed
- Overall project budgets and expenditures
- Selection of key consortium staff and contractors
- Monitoring the technical quality of the assessment program

This group should be representative of state-level policy, content, and assessment groups.

*Which Groups Should be Represented?* The following types of groups should be represented on the policy advisory groups:

- Policy – state boards of education, chief state school officers, state legislators, and governors
- Curriculum – national and state mathematics councils, reading, writing, and English associations
- Special Populations – experts and advocates for SWDs and ELs, as well as any other historically disadvantaged student groups
- Professional Development – state-level staff development councils or groups
- Assessment – state assessment directors
- Technical Advisory Committee – a representative from the project technical advisory committee

By forming and using such a group, overall leadership can be provided to the project that will permit it to operate without sudden changes in direction, but also to maximize the value and utility of the information that the consortium’s assessment system will provide.

### **C. Selecting and Using Committees and Advisory Groups**

It is strongly advised that the consortium use cross-state project advisory committees. There are several that immediately come to mind, and others might be formed as the need arises. Those advisory groups that should be formed at the outset include

- *Content Advisory Committee* – This group of about 6-8 individuals should serve as advisors to the project, project staff, and contractors on content-related issues. These issues include what content to assess, how to assess the content, what types of reports of results are desired and that are feasible, how the assessment results can be used, the professional development needed by teachers on instruction and formative assessment, and so forth.
- *Assessment Advisory Committee* – This group of about 6-8 assessment directors or their staff should serve to advise on the technical and procedural aspects of item development, editing, review, field-testing, and selection. They can also advise on assessment administration procedures, analysis procedures, and the psychometric work needed to validate the assessments. This committee would serve to provide input to all assessment directors on the technical issues inherent in a large-scale assessment program of this scope.
- *Professional Development Committee* – This group of about 6-8 individuals should plan the awareness and training procedures and materials necessary to prepare local educators to understand the Common Core State Standards, how

to teach the standards (including implications of the standards for local district curricula, instruction, and learning), preparing students to take part in the assessment program, using the assessment results to improve the instructional program, and the preparation of teachers to formatively assess students during instruction.

- *Technical Advisory Committee* – As noted earlier, a group of about 12-15 outside technically-trained individuals who may currently serve as advisors to states should be asked to serve as a “Super-TAC” for the project. Because many of these individuals already serve on state TACs, they are mindful of the myriad of issues that can occur when assessments are designed, developed, and implemented. Thus, they will be the ideal advisors for comparable work across a large number of states.

By using these advisory committees, more work can take place simultaneously in preparation for a successful launch of the new consortium assessment program. A secondary benefit of including more state-level staff in planning for the implementation of the common assessment program is that more state staff will be committed to the successful launch of the new assessments.

#### **D. Ongoing Decision Making**

The project manager will work with the project vendor(s) on a day-to-day basis to make sure that all project activities are being carried out successfully.

The project leadership team should meet frequently – as often as monthly or more often initially, as many key decisions will need to be made. The project manager should convene this group and keep them informed of activities between meetings. This group can meet in person or via webinar as issues arise.

The above-listed advisory groups should be used to provide the policy, technical, and practical advice to the overall project. They should also be under the direction of the project manager. These committees and groups should be convened on a regular, planned basis (such as semi-annually or quarterly), as well as when serious issues arise that require more immediate attention.

Finally, all of the states in the consortium also need to be convened periodically (as the state advisory committee), more often perhaps at the outset of the project and less often as the project activities are being successfully carried out in subsequent years. States will need to select one or two people to attend these meetings when the consortium determines the number of representatives from each state. These individuals should strive to attend every meeting in order to provide consistent advice

and to carry out any work assigned to them (or agreed to by them) between meetings.

By using the four-level structure described above, states have several ways in which to be involved and have a say in decision making. States can choose to be involved in the project leadership team, one of the advisory groups (content, assessment, or professional development), as well as the state advisory committee. Thus, it should be possible for state department of education staff to continue to work for their state education agency and still be involved in consortium decision-making. Of, course, several of the state education agency staff may want to consider staff positions with the consortium as a whole, since the consortium will need to hire several types of individuals to successfully carry out its activities.

There are a number of issues in managing a multi-state consortium for a large-scale common state assessment. Among these are

*What happens if more money is needed?* If additional resources are needed, states participating in the consortium will either need to reduce the scope of the project or provide funds from state, Federal, or other sources. The consortium will need to tightly monitor the cost of the project and budget status of all participants throughout the process.

*What happens when there is turnover in policymakers and other key individuals in states?* Ideally, by staggering membership on the policy advisory group, and using such a group in advising the overall project, turnover in governors, chief state school officers, legislatures, and state boards of education will not have an immediate impact on the consortium or its activities. Hopefully, the policy advisory group will be shielded from dramatic changes in leadership, so the policies, operations, or membership of the consortium do not radically change as the result of policymaker changes.

*What happens when the project schedule is not maintained?* This is a critical issue. The schedule needs to be managed actively by consortium staff, contractor staff, and the various advisory committees. Ideally, extra time should have been built into the schedule to prepare for the eventuality of slippage in the schedule, since this is not uncommon in newly created assessment programs. If such slippage does occur, revised schedules will need to be prepared. It will be critical to keep the schedules up to date.

*What happens if the Common Assessment program is not federally approved?* If a Peer Review process is used in the re-authorized Elementary and Secondary Education Act (ESEA), the assessment developed and used by the consortium will need such

approval. One way to help ensure this is for consortium and contractor staff, as well as advisory groups, to remain vigilant about the types of changes that may occur when ESEA is re-authorized, and what implications they may have for the approval of the assessments by the USED. If despite this, the assessments are not approved, the consortium will need to take corrective action to make sure that the assessments are approved for use.

*What happens if the standards in the Common Core are adjusted before use or as the assessments are being used?* If, or perhaps more accurately, when adjustments are made to the standards, states participating in the consortium will need to make adjustments to the assessments to maintain alignment with the overall Common Core State Standards. If the adjustments are minor, the trend line for the assessment program will not need to be broken. However, if the “adjustments” are more major and pervasive, a new trend line will need to be used, which could be upsetting to policy makers, although required by the statistical reality of the “revised” assessments.

*What if the new assessment turns out to be too costly for some states to use and continue in the future?* Hopefully, this will not occur, because ASG has been tasked with examining costs as the consortium assessment designs are developed. These costs included those for development as well as implementation of the assessments by states in the years following 2014. Should it turn out in the long run that states are not able to afford the new assessment, the consortium will have a couple of choices. First, it can raise funds to pay the added costs. These might be state funds, Federal funds, or grant funds. Second, the consortium can look at a number of ways to lower the operational costs of the assessments without impacting the quality and content of the assessments. Again, cost modeling services can be helpful in this regard. Third, the consortium can look at various ways of changing the content of the assessments themselves such as reducing the number of the more expensive items used – the constructed responses and performance assessments – and replacing them with less expensive items like multiple choice. These changes to the assessment content could also be limited to those states with funding issues, although comparability would need to be evaluated.

*How will consistency in direction be maintained?* It is the task of the advisory groups and the consortium staff to ensure the consistency in direction for the consortium and to make any changes in direction needed after careful consideration is given to the alternatives, the advantages of each, and the costs of each. Ultimately, consistency will be maintained if the individuals involved include contingency plans and build in safeguards from the outset.

*How will risk be managed? What are some risk mitigation strategies?* Risk management is something that all consortium and contractor staff should practice and the advisory groups are mindful of as the work is being done. Where possible, redundancy should be built into the system – for example, extra items, extra test forms, extra scorers, extra time for development, extra field test participants, redundant data analyses, careful checking of scoring and reporting before release, and so forth. Many of these steps are ones that individual states might carry out, but now need to be done on a much larger scale, relative to the much greater consequences for failure.

### **E. Use of Third Parties**

Third parties will be used by the consortium in several ways. First, the project manager will likely be a person or entity from outside of the project. Second, special content, assessment, IT, or psychometric specialists might be employed as project needs arise. These may be occasions when an issue or an idea needs to be pursued where one or more specialists might provide one-time advice to the group. Third, specialized organizations might be used to deal with some of the project's thorny issues, such as the rollout of online assessments or preparing exceptional materials for providing student results to parents. These groups might be used to provide advice and information that is not required on an on-going basis. Fourth, specialists might be employed to monitor assessment design changes, technical issues, and the budgets for the overall consortium and each of the vendors. Individuals who specialize in any of these functions might be helpful to the consortium to keep it on track and on budget.

### **F. Indicators of Success after 1, 2, 3, or More Years**

Each consortium will ultimately have to establish a detailed multi-year schedule for its activities, as well as a higher-level schedule for those not involved in the daily activities of the consortium to use to follow developments. A high-level schedule is provided in Appendix B as an example of key milestones during the four-year development period.

By the end of the first grant year (September 30, 2011), the consortium should have hired its development vendor(s) and should have the item development process underway or about to commence. At the end of the second grant year (September 30, 2012), all of the assessments should be under development, with some being edited, and informally pilot tested. By the end of the third year (September 30, 2013), states should have completed the development and editing of the measures, and be preparing for field testing of the new assessments. During the final grant year, states will be conducting the consortium-wide field testing, analyzing the field-test results in

order to select items for the final forms, assembling the final test forms and computer delivery system, and preparing districts for the administration of the assessments in the spring or the following fall.

As mentioned above, the consortium will need to develop detailed plans and a timeline for its activities; these dates are merely suggested benchmarks.

## **G. Membership Fees in the Consortium**

All fees in the consortium will need to be determined by the members in it. This is an area where the policy advisory committee may need to weigh in. Among decisions that will need to be made are

- One-time or annual membership fee applied to all state members in the consortium. Such fees could be used to support the activities of the consortium (e.g., advisory committee meetings, staff salaries, travel for members, and so forth).
- Fees for new states that wish to join the consortium after it is under way. Such states might join at the conclusion of the development period (2013-14), before the new assessments are first administered, or at any point before or after this time. As part of this decision, the consortium will need to determine if late-joining states need to pay any one-time or extra on-going charges in order to join.
- Fees (if any) for states that do not want to join the consortium but just want to use the tests produced by the consortium. This will meet the requirement of USED that the materials produced by the consortium be available.

## **H. Ownership of the Assessments**

A related topic is what resources are owned by each state that is a member of the consortium. If a state chooses to leave the consortium, a policy needs to define any “individual” ownership over any of the materials produced by the consortium, such as the tests to be used for the next two years as the state prepares to transition to a different assessment or system. Another issue to be considered is who will own and maintain the copyright permissions on materials used in the tests?

The states that are members of the consortium own the measures created by the consortium. However, the USED requires that each consortium share its materials at no cost. An outstanding question to be resolved by the USED and the consortium is the appropriateness and legalities of the consortium charging other users reasonable fees to use the materials that the consortium has developed. This is a particular

concern in the eventuality that state resources as well as Federal funds have been used to develop the consortium's assessment program.

*State Leaving the Consortium* – When a state leaves the consortium, it may wonder what resources it owns and that it can take with it. One strategy is to indicate that departing states leave with nothing other than good wishes. Another strategy is to provide a departing state with access to older forms, ones that have been used but not released. Another variant of this is to provide a departing state with access only to released forms. The consortium may wish to establish a price schedule for a departing state that wishes to purchase forms from the consortium in the future. If so, these fees should not act as an incentive for the state to leave the consortium but purchase the consortium's tests at lower cost.

*New States Join the Consortium* – It is anticipated that states not part of the consortium when it was formed may wish to join the consortium at a later time. The consortium's advisory groups should consider what to do in such instances. Several options are available. These include charging a "membership fee" for a newcomer to cover past non-recovered expenses, permitting the new state to join at no cost but pay its share of future fixed and variable costs (which could be lower due to yet another state across which to spread the fixed costs), or not to permit states to join for a period of time. The latter may not be a viable long-term strategy, but might be advisable during the early few years as the consortium sets up shop and builds the initial assessment measures. There are many policies and procedures to work out and having new members involved will only complicate matters.

## **I. Consortium Development and Implementation Schedule**

The consortium will need to develop high-level assessment development and assessment administration schedules, as well as highly detailed timelines and schedules listing every step in the process on a daily/weekly basis. It will need to be in an electronic format that is easily accessible by all key project personnel, preferably on the Web. These schedules also will likely undergo revisions and updates on an ongoing basis. The comprehensive detailed schedule for a consortium over the four years that it has to develop the assessment measures might be several hundred pages long. More important than a lengthy schedule, however, is the use of the schedule to manage the overall program. This means frequent meetings between consortium staff and vendors, with ongoing references to the project schedule. This should be a consistent activity during each project meeting, which should occur at least monthly during the life of the project.

The consortium may need to determine ways to streamline the number of meetings that are held. In addition to management meetings, states will need to consider ways

to deal with the many other meetings that will need to be scheduled, including those of all the committees listed in Part C of this section. A wise consortium will look to the use of remote meeting technologies such as webinars in addition to face-to-face meetings.

## **VII. Cost Issues**

There are several financial functions related to managing the consortium and its common assessment system that will need to be handled by the consortium. This section of the Guidelines document reviews some of the key financial areas associated with the work of the consortium and the assessment system that will be designed, developed, and implemented, and provides recommendations as to how a consortium can manage them effectively. The following topics are discussed in this section.

- A. Determining the Cost of the Assessment System
- B. Developing the Consortium Budget
- C. Allocating Costs Across Participating States
- D. Monitoring Costs
- E. Dealing with Vendor Cost Overruns or Failure to Perform
- F. Handling Scope Change Requests and Budget Modifications
- G. Vendor Reporting Requirements
- H. Consortium Budget Reporting Requirements
- I. Suggestions on Procurement Strategy

### **A. Determining the Cost of the Assessment System**

One of the primary goals of the consortium is to procure a high quality assessment at a price that is affordable to all states in the group. Such a high quality assessment system would likely not be affordable for an individual state acting on its own. However, the ability of the consortium to spread the fixed assessment costs of development and vendor overhead over the states participating in the consortium results in each state's required funding for the assessment system being significantly less than it would otherwise be. That said, as the assessment system is being designed, the consortium must still obtain a reasonable and robust estimate of the total and per state assessment costs (both development and ongoing administration) and get state approval for each. States must know their share of the consortium assessment budget in advance of the consortium actually committing to any significant development costs to make sure the costs are affordable in future years of administering the assessment and that costs are provided for in their budget authorizations and procurement plans.

Each consortium has developed and costed an assessment design prior to submitting its grant proposal to the Department in June 2010. As noted, ASG was brought on board to assist the consortia with this activity. Additionally, each consortium was required to comment on the sustainability of the assessment, after

the development period, in its NIA response. Operational costs of the proposed assessment must be compared to existing state assessment expenditures and any gaps in costs must be explained. Therefore, it is expected that the assessment design will undergo some changes during the first project year as design and cost estimates are further refined and states buy into each.

It is best to have the final design and all cost estimates completed *prior* to letting the RFP for vendor services. Significant time, money, and effort are wasted if the cost of a proposed assessment is greater than the consortium or member states can afford or can be sustained in years to come. If a state cannot afford its share of the assessment system cost and drops out of the consortium, remaining members must pick up the additional costs. Therefore, it makes sense to obtain accurate cost estimates at the outset of the formation of the consortium. Determining the assessment system's costs up front also puts the consortium in a better position to negotiate with potential vendors. The difficult part is determining what the assessment costs should be.

In the past, a single state might have been able to make approximate comparisons of the components of its current assessment program to those of a prospective program in order to estimate the total cost for a new program, provided no major changes to the assessment system were being proposed. However, a consortium will, almost by definition, procure an assessment that is significantly different (and hopefully of higher quality) than that of any individual state in the group. In order to calculate the proper amount the consortium should pay for a new assessment, one must be able to build up the cost of the assessment by the basic elements of the assessment system. This must be done function by function and element by element. Complicating matters is that many new technologies, without established pricing history, might be desired to be used by the consortium. Examples of these new technologies are online systems, artificial intelligence engines for scoring of constructed response questions, and systems that can deliver innovative, interactive item types (simulations). The consortium will need to conduct its own research, or work with someone that has conducted this research, in order to properly estimate these prospective assessment costs.

Finally, the consortium will most likely want to make use of higher quality items such as performance events and performance tasks that do a better job of measuring a student's higher order and critical thinking skills. Development and administration costs for these items can be significantly greater than those for traditional selected response and constructed response items. Again, the consortium will need to conduct its own research or work with someone that has already gained this knowledge in order to arrive at a good estimate of the cost for the new assessment system. In the past year, the ASG has done much of this

research, has worked with states<sup>3</sup> and Stanford University<sup>4</sup>, has developed and refined a proven cost model, and was selected by the Bill and Melinda Gates Foundation to assist each state consortium in determining its proposed assessment system costs.

## **B. Developing the Consortium Budget**

The consortium budget for the assessment will be made up of three different parts each with development and ongoing cost components. First, the actual assessment development and ongoing costs will be major budget items for the group. The development of the assessment system will be a large undertaking that will require outside assessment design, development, and psychometric expertise; consist of new item types that must be developed to the new Common Core State Standards; and generally be of higher quality than existing state large scale assessments. The ongoing assessment administration costs generally dwarf the development costs and are far and away the largest budget item for states, and the one that all states must make sure they can afford on a continuing basis. Currently, the federal government funds about 40% of state assessment costs via NCLB Title VI funding although it is not known what funding will be provided in the future given the reauthorization of the ESEA (the current NCLB), current budget issues being addressed by the federal government and, potentially, increased costs of higher quality assessments. Consortium members will need to anticipate potential resources from both the federal and state governments when determining affordable assessment costs.

The development costs for the technology system(s) to deliver the assessments will be the second major element of the consortium budget. As new systems are implemented to deliver the assessment, score constructed response and other items, and report information to students, teachers and other constituents, the costs for the development of these systems will need to be estimated and budgeted. Once the systems are developed, the ongoing support costs, including staffing, will need to be managed. The third area of costs for which the consortium must budget are the expenses of managing the consortium itself. Costs for the various working teams, state personnel, subcontractors, policy groups, advisory committees, as well as all the meetings needed to manage the project, need to be

---

<sup>3</sup>Assessment Solutions Group. *Cost Estimate for the New Kentucky Assessment System*. Prepared by ASG for the Kentucky Department of Education: Author. 2009.

<sup>4</sup>Topol B., Olson, J., and Roeber, E. *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*. Stanford Center for Opportunity Policy in Education, March 2010.

captured and monitored both during the project period and beyond. The consortium budget will flow from the consortium governance structure and assessment committee membership. An allocation of personnel time spent on consortium matters will need to be estimated and budgeted, as well as all the indirect expenses associated with consortium management activities. Staff hired by the consortium will need to be paid and their expenses covered. A proper time and expense tracking system will need to be implemented to insure that time and other costs are properly charged to the consortium. The financial representatives from the consortium should be involved early in the process to ensure that all items are properly budgeted and tracked.

### **C. Allocating Costs across Participating States**

Once total assessment system costs and the composition of those costs are known, they can be allocated to the states participating in the consortium. The best way to delineate costs for allocation purposes is on a fixed/variable basis. Fixed costs of the assessment (functions like development, program management, IT, QA, corporate overhead and consortium overhead) can be allocated to participating states in a few different ways.

- *Divided equally across all states* – This methodology assumes that each state would incur similar costs if it were acting on its own and therefore an equal amount of fixed costs are allocated to each state. The advantage of the methodology is that it is fair to all states and that fixed costs are treated as fixed among all states. A disadvantage is that small states will pay a greater cost per student for these expenses than larger states.
- *Allocated on a per student basis* – Charging fixed expenses to all states based on the number of students in each state can be another reasonable allocation methodology. If the consortium uses this allocation methodology, small states will have an added benefit of paying a lesser portion of fixed overhead than they would under the first methodology. A potential benefit to large states is that more small states may be able to afford to participate in the consortium which ultimately reduces to cost of the assessment for every state. A potential disadvantage of this approach may be that larger states will pay a larger share of fixed expenses than in the methodology above.
- *Hybrid methodology* – There are several hybrid approaches to allocating the fixed assessment costs. One would be to create a factor for per-student funding from state and/or federal sources and allocate fixed costs based on a formula using these funding ratios and student counts for the states in the consortium. The formula can vary based on negotiations within the consortium and the

weights assigned to the various factors. Another hybrid model is to charge a portion of the fixed costs (such as the first half of them) to each state equally, and the remaining fixed and all of the variable costs on a per-student basis.

Functions with costs that vary with the number of students taking the assessment (scoring costs, printing costs, online system per student administration costs, shipping costs, editing costs, reporting costs, etc.) should be allocated to each state based on the number of students taking the assessment in each state relative to the total number of students taking the assessment in the consortium. Additionally, some states may be interested in certain aspects or features of the assessment system (e.g., through course assessment or formative assessment professional development) that others are not. These costs will need to be separately tracked and allocated to the states using these services using one of the above methodologies.

#### **D. Monitoring Costs**

It goes without saying that all of the costs for development and ongoing management of the assessment need to be carefully monitored. If fixed cost contracts are let for the technology, assessment development, and ongoing assessment administration, cost overruns will be the responsibility of the vendor. However, it is never to a state's or consortium's advantage to have a vendor experience significant cost overruns on a project as the vendor will usually try to regain its lost profitability elsewhere in the project. This can result in continuing problems in the management of the program and considerable customer displeasure. Therefore, it is best to work with the vendor to proactively manage the program as efficiently as possible and develop the reporting and control processes required to do so.

The vendor and/or consortium program management function should keep the consortium abreast as to cost, schedule, and quality issues, as well as potential risks, via the normal program review process. Budget issues should be reviewed with both the vendor and the consortium leadership team at least quarterly and, preferably, on a monthly basis. Additionally, a policy will need to be implemented as to how shortages and surpluses in the consortium budget are shared (or not) among the participating states. It would be wise to set aside a contingency fund each year to handle possible cost overruns and/or scope changes. The contingency fund can be refunded to states at periodic points during the life of the program or maintained at the consortium level to offset future state charges.

## **E. Dealing with Vendor Cost Overruns or Failure to Perform**

As stated above, the consortium should work with and manage the vendor to deliver the program on budget, on schedule, and with the required quality. A fixed price contract with liquidated damages is the typical method for penalizing the vendor for a failure to perform. Liquidated damages can be assessed at an individual item level, a program level, or a combination of both. It is recommended that liquidated damages for *major* program deliverables (both schedule and quality) with a cap on annual liquidated damages at 10% of annual program costs be included in the contract. Liquidated damages clauses in excess of this amount will likely result in the vendor charging a higher price to cover its excess risk and will likely not lead to improved performance. The best way for the consortium to minimize schedule risk is to work with the vendor to make sure all schedule deadlines, both consortium and vendor, are met. A contract with a reasonable liquidated damages clause as opposed to one with excessive liquidated damages will be more helpful in establishing a cooperative working relationship with the vendor.

## **F. Handling Scope Change Requests and Budget Modifications**

Most assessment programs, particularly new ones, will go through some scope changes as the program is rolled out and plans reviewed and refined. It is important to minimize these scope changes and implement them in the proper time period if the consortium is to maximize the efficient delivery of the current program. If at all possible, scope change requests for operational programs should be scheduled for implementation in the FOLLOWING program year in order to minimize the potential of the scope change negatively impacting the current year's program schedule. Additionally, the program budget at the consortium level should be adequately funded to account for scope changes. An approval system for scope changes that reflects different levels of approval for different size scope changes will need to be implemented at the proper consortium approval level. It is recommended that contingency funding be established for both the grant and post grant periods in order for the consortium to be able to adapt the assessment program as required.

Finally, it is important to recognize that the shoe is on the other foot when negotiating a scope change with a vendor, as the customer has no other practical options to using the existing vendor to deliver the change in program scope. As a result, costs for scope changes that increase the scope of a program are typically quite high while cost reductions resulting from decreases in program scope are generally relatively low. For large changes in contract scope (both increases and decreases), it is recommended that cost modeling services be used as a way of

estimating the actual vendor change in cost. This data will provide reliable information that can be used to negotiate the price of the scope change with the vendor. Better yet, if appropriate cost sheets with appropriate program metrics are included with the RFP and required with the vendor RFP submission AND the contract requires relatively close adherence to the cost and program metrics included in the cost sheets in the event of a scope change, the consortium can protect itself from improper pricing. The Assessment Solutions Group has developed cost input worksheets with appropriate metrics the consortium can use in its assessment services RFP.

### **G. Vendor Reporting Requirements**

Again, it is essential that a positive working relationship be established between the consortium and its contractor(s). The better the working relationship, the more open the contractor will be in sharing important program details. The project management team should hold monthly meetings with the vendor at the outset to monitor cost, schedule, and quality of the program. Later on, it may be possible to hold project meetings less frequently (i.e., every other month or quarterly). Although many of these meetings should be held face to face, some may be held electronically via WebEx or teleconference to help control costs for the program. The status of all key deliverables should be reviewed during these sessions. Sufficient detail should be covered to satisfy the consortium program management function that the program is proceeding as planned or that proper corrective changes are being implemented to fix any problems that may have arisen. Cost and schedule can be reviewed at the functional level on a current month basis, program year-to-date basis, and calendar year-to-date basis. The ideal way to review these data is to combine the consortium/vendor program reviews with the regularly scheduled vendor internal program reviews if the vendor is agreeable.

### **H. Consortium Budget Reporting Requirements**

As stated earlier in this document, active and open communication between and among the consortium management and member states is important for the successful functioning of the consortium as a whole. This communication practice extends to the reporting of cost, schedule, and quality issues. Monthly or quarterly management and budget reports should be issued to all member states providing status on schedule, cost, and quality issues with potential risk areas highlighted. Additionally, a commentary on the status of all key deliverables should be included in the report. Costs should be detailed by function on a monthly, quarterly, and program year-to-date basis. Some states may also have a need to see costs vs. budget on a fiscal year basis so the reporting system that the consortium develops will need to be flexible enough to accommodate both. Additionally, the USED will

have project reporting requirements during the development period of the grant. The consortium will need to determine what the federal reporting requirements may mean to the design of its project information and cost reporting system. Sample consortium budget reports are listed in Appendix C.

## **I. Suggestions on Procurement Strategy**

After the USED common assessment award is granted, an intensive period of work needs to occur to refine and cost the assessment design, develop detailed systems requirements for the various technology components of the assessment system, and further refine and establish the consortium working teams. We believe outside vendors should be hired as soon as possible (i.e., in the first few months of the project – October to December 2010) to manage the overall project, consult on assessment design, evaluate existing technology vendors and make build/augment/buy recommendations, and refine the assessment and technology cost estimates and budget. Additionally, the working teams will need to move forward on development of the various modules of the assessment system (assessment design, research and evaluation, reporting, professional capacity and outreach, technology, collaboration with higher education, etc.). This work should be completed prior to issuing RFPs for the actual assessment system development.

This will be the largest and most complex assessment procurement undertaking ever and probably needs to be handled in a different manner than how assessments are procured and managed by individual states today. Because of the complexity of the common assessments and the large technology component that will be an integral part of the new assessment system, ASG recommends that the management structure and work effort in creating the assessment system be divided into two major categories – traditional assessment development and technology development. The entities selected to lead the two areas would operate as prime contractors, each choosing the best subcontractor(s) to deliver the required service. Additionally, the technology lead partner would act as the systems integrator for the project, making sure that all the technology pieces seamlessly work together as a unified system.

Both lead managers should report to either the consortium project management partner or to the consortium itself. (Note that it will be interesting to see how/whether assessment and technology vendors team up to offer services to the consortium. Once this is determined, the consortium can decide if it wants to pull out certain functions, because of their critical nature or the entity providing them, and have them report directly to the consortium.) The potential assessment functions and technology applications under each category are as follows:

## Assessment Functions

1. Item development and test creation – Responsible for the actual development of the test items and assessment, production of forms or computer-ready items, printing of test materials, and processing and scoring activities. Note that with the advent of new content standards and the need for higher quality items, the currently available pool of item developers may be hard-pressed to do all the work that will be required. However, the large testing vendors may have enough good item writers for this work. Additionally, some of the research-oriented centers, national research labs, and other organizations may be able to provide innovative assessment prototypes for this work. Several options are available with regard to hiring various item development partners, including separate vendors by content area or domain, separate agreements to develop performance events and performance tasks, and/or separate agreements to develop computer enhanced or technology enabled items.
2. Quality Assurance – Conducting quality control checks of critical psychometric activities, such as equating, scaling, and reporting of scores from the common assessment. We recommend that a fully redundant model be used to independently check all assessment results before they are reported. Other quality assurance work could focus on quality of items, scoring, or technological applications.
3. Alignment studies – Conducting alignment studies of the assessment items to the Common Core State Standards, as well as examining the relationship of the items to the college and career readiness standards. Typically, this work is done by a different group than the item development vendor(s) to ensure the integrity of the process and the results of the alignment study.
4. Test Validation – Examining validity issues for the new assessments. A research-oriented organization under the direction of the prime assessment vendor can assist the consortium in conducting a variety of validity studies and evaluating items and test forms to help validate the assessment instruments.

## Technology Functions

1. Online Test Platform – This procurement will be for the basic test delivery and selected response item scoring and reporting functions. While a separate reporting system and developer may be necessary, we believe the online test vendor will most likely be able to provide this capability.
2. Artificial Intelligence Scoring – AI scoring will be such an important component in the affordability of the new system that we recommend a separate vendor for this scoring service. This procurement RFP can also include the automated speaking and listening systems that may be a component of the assessment system.

3. Other Technology Applications – The lead technology entity will also provide management of other technology applications (web-based item development platform, central database and information repository, etc.) the consortium may want to develop/procure.

As stated above, the consortium may wish to exert greater management or have an area have greater independence by having it report directly to the consortium itself or the consortium's project management partner.

Finally, an experienced and skilled organization should be hired by the consortium that can do cost modeling, budget forecasting, and various analyses to examine costs for the many work activities and identify areas for improved efficiencies. This group could assist the consortium in analyzing RFP bids and cost proposals from prospective vendors to help get the best deal for the consortium.

Finally, the nature of the bidding process itself must be determined. Current bid processes include such things as a single evaluation of all aspects of the proposal (cost and technical), separate cost and technical proposal evaluation, requirement to select the low cost bidder, cost and technical proposals evaluated separately, and using a preliminary bid and then a Best and Final Offer (BAFO) process. It is recommended that a proposal be evaluated in its entirety (technical and cost) and that a detailed scoring rubric be developed for the evaluation of the various bids. The best overall proposal should be selected. The process should allow for negotiation of the final price with multiple bidders. These recommendations should help the consortium procure the best product at the best price.

## VIII. Summary and Recommendations

In developing the guidelines presented in this report, the authors acknowledge that there are likely other issues that will need to be addressed by a state consortium. The process of designing and developing a high quality common assessment system that can be used by many states is daunting, and states will face many challenges in doing so. In the *Guidelines for Organizing and Leading a State Assessment Consortium*, we have targeted many of the most critical issues that the consortium will need to make decisions about – in the planning stage and during the development and implementation phases. These issues have been discussed in detail in the previous sections of the report.

Overall, the following areas will certainly require a great amount of attention by the states participating in a common assessment development consortium. In our opinion, the areas that absolutely must be addressed are

- States will need to share a common vision for the common assessment system, one that is coherent and meets the needs of individual states. This vision should be defined and described in advance by the consortium.
- Strong leadership will be needed to allow the consortium to be successful. The leadership will need to come from the states as well as from others who are working closely with the states in the consortium. The project will need good management, both in the personnel assigned to the project and in the procedures used. Clear governance policies need to be agreed on by all states.
- In all key decisions made by the consortium, it is important that a consensus is reached across all states. This should be the goal for all crucial decisions that are made. If consensus cannot be reached, then the consortia will need to have a formal protocol for action.
- The organizational structure of the consortium needs to be identified and clearly laid out early on, with specific information on how all the players in the group will interact (states, committees, partners, advisors, contractors/vendors, and others). The roles and responsibilities of all parties need to be specified.
- The plans for use of the common assessments should allow for a degree of flexibility so that states can adapt different aspects to their specific needs or capabilities. For example, the consortium may need to be flexible in some areas, such as the use of the interim and formative components, teacher scoring, and/or administration on computer, so that individual states can either implement the consortium's preferred approach or possibly use another option.

- The consortium will need to determine how technology will be used by the consortium, such as various systems for administering, scoring, maintaining longitudinal data, reporting, item banking, etc.
- A number of important decisions will need to be made in regards to technical issues that impact the common assessments. These include
  - The design of each assessment component, the types of measures that will be used, and the numbers and types of items required in the math and ELA assessments
  - The administration mode to be used, such as CAT, CBT, and/or PPT
  - Comparability of scores across all states, especially if different states have different schedules in administering the assessments
  - Reporting of results, such as by standard or strand, types of reports, and their audiences
  - Psychometric issues, such as equating methodology, scaling, use of a vertical scale, types of validity studies, etc.
  - Other critical issues, such as specifying performance level descriptors that apply to all states and setting common performance standards.
- Other areas where important decisions also will need to be made include
  - Use of data (and perhaps restrictions in use of data) from the common assessments by different states, districts, schools, teachers, parents, and the public
  - Professional development for teachers
  - Common policies and procedures for participation of SWDs and ELs in the assessments, accessibility needs, and use of accommodations
  - Test security
- Another important area that the consortium will need to consider is dealing with possible changes in the future to the design of the common assessments, to the standards, in uses of technology, or even in federal policy impacting state assessments (such as the reauthorization of ESEA).
- Finally, it will be critical to address the issue of costs – the total cost for all the work of the consortium needs to be determined in advance. Obviously, the total amount for the four-year project must be under the \$150,000,000 that the USED will fund for Level 1 budgets. These costs need to be analyzed and summarized to ensure that the assessment system that is proposed is affordable. In addition, costs for individual states to maintain and sustain the common assessment in the years that follow the completion of the development work also need to be determined in advance. As noted in the previous section of this report, the data can be estimated and obtained by the use of good comprehensive cost models.
- As the consortium does its work and moves forward with development, ongoing costs during Years 2, 3, and 4 will need to be closely monitored. The

project budgets will need to be evaluated at least quarterly to make sure that unforeseen expenditures and cost overruns do not cause budget deficits.

- A well-written RFP (or multiple RFPs) will be needed to procure the vendor(s) for the project. The RFPs need to be very specific, clear, and comprehensive, and should include a good way to analyze costs that are proposed by vendors.

The Assessment Solutions Group is pleased to be able to provide states with these *Guidelines* for organizing and leading a consortium, with the hopes that the collaborative work will be successful and the common assessments that are developed will provide states with the best in innovative products and valuable resources that can be used for many years. The important decisions that are made early in the process will directly affect the outcomes of the work.

## References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association. 1999.

Assessment Solutions Group. *Cost Estimate for the New Kentucky Assessment System*. Prepared by ASG for the Kentucky Department of Education: Author. 2009.

Board on Testing and Assessment. *Letter Report to the USED on the RTTT Fund*. National Research Council. October 2009.

National Council on Measurement in Education and American Council on Education. *Educational Measurement, Fourth Edition*. (Edited by Robert Brennan). 2006.

Olson, J. F. & Dirir, M. *Technical Report for Studies of the Validity of Test Results for Test Accommodations -- Establishing the Validity of Test Accommodations and Score Interpretations for Students with Disabilities: A Collaboration of State-based Research* (Connecticut Enhanced Assessment Grant). Council of Chief State School Officers, June 2010.

Thompson, S. J., Johnstone, C. J., Anderson, M. E., & Miller, N. A. *Considerations for the Development and Review of Universally Designed Assessments* (Technical Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. 2005.

Topol B., Olson, J., and Roeber, E. *The Cost of New Higher Quality Assessments: A Comprehensive Analysis of the Potential Costs for Future State Assessments*. Stanford Center for Opportunity Policy in Education, March 2010.

## **Appendices**

A – Sample List of Key Tasks and Activities for a State Consortium

B – Sample Project Calendar: Major Milestones in Years 1-4 and 1<sup>st</sup> Operational Year

C – Sample Budget Report Formats

D – About the Assessment Solutions Group

## Appendix A

### Sample List of Key Tasks and Activities for a State Consortium

- ✓ Select management organization
- ✓ Hire management staff
- ✓ Create policy, technical, and practical advisory groups
- ✓ Develop final draft assessment designs
- ✓ Review assessment designs with policy, technical, and practical advisory groups and state members
- ✓ Finalize assessment designs
- ✓ Establish general performance standards and content/grade specific performance standards
- ✓ Develop educator and other citizen awareness materials for use during each phase of development and implementation
- ✓ Translate assessment designs into bid specifications
- ✓ Develop Request for Proposals
- ✓ Issue RFPs
- ✓ Conduct pre-bid meetings
- ✓ Select vendors
- ✓ Negotiate contracts
- ✓ Sign contracts
- ✓ Begin assessment development work
- ✓ Develop new assessment measures
- ✓ Edit the new assessment measures
- ✓ Create pilot test forms
- ✓ Informally pilot test the new measures with a couple hundred students
- ✓ Analyze pilot test data (student results and teacher comments)
- ✓ Review pilot test data and revise the items as necessary
- ✓ Prepare for formal field test
- ✓ Load items in online assessment systems
- ✓ Conduct field testing using paper and electronic means
- ✓ Collect field test information

- ✓ Analyze field test data
- ✓ Review field test data
- ✓ Revise/drop items as necessary
- ✓ Conduct standard setting
- ✓ Prepare technical report on item development
- ✓ Select items for final test forms
- ✓ Package the final items in the test forms
- ✓ Print the paper-based forms
- ✓ Load the electronic forms into online systems
- ✓ Develop assessment administration materials
- ✓ Package paper-based materials by school
- ✓ Ship paper-based materials to schools
- ✓ Open online assessment system(s)
- ✓ Conduct statewide assessments
- ✓ Collect student responses on paper
- ✓ Scan student paper responses
- ✓ Analyze all student responses
- ✓ Produce reports of results
- ✓ Conduct confirmatory standard setting to validate cut scores
- ✓ Post reports of results
- ✓ Conduct various statistical results
- ✓ Prepare technical report on assessment administration

## **Appendix B**

### **Sample Project Calendar: Major Milestones in Years 1-4 and 1<sup>st</sup> Operational Year**

#### **Major Test Development Activities by Year**

##### Year One (2010-2011)

- Refine/finalize and cost assessment design; develop, issue, and award RFP for assessment vendor(s)
- Hire technology consultant/systems integrator to refine technology requirements and write system requirements
- Develop, issue, and award RFP for systems vendors
- Develop assessment blueprints and test/item specifications
- Determine performance level descriptors and general performance standards

##### Year Two (2011-2012)

- Based on item specifications, develop a limited number of innovative items (CR, PE, PT) for pilot/tryout
- Conduct pilot test
- Score items
- Gather preliminary statistics, analyze data, and refine items
- Begin technology system builds

##### Year 3 (2012-2013)

- Continue with more significant item development
- Conduct field test (online)
- Score items
- Compute statistics, analyze items, and refine/adjust
- Review functionality of online system
- Limited tryout of technology systems

##### Year 4 (2013-2014)

- Develop remaining items for use in first operational year
- Conduct field test #2
- Score items
- Compute statistics, analyze items, and refine/adjust
- Select all items for operational use and prepare forms or computer delivery system
- Full tryout of technology systems

Year 5 (2014-2015) – 1<sup>st</sup> Operational Year

- Administer operational assessment in every state
- Score items
- Conduct standard setting
- Provide data to all states
- Report results

## **Appendix C**

### **Sample Budget Report Formats**

### Sample Consortium Cost and Unit Reporting Worksheet

**Budget vs. Actual Costs - Current Month**

Fiscal Year Month:

Fiscal Year:

|  | Prior Month \$ |        | Current Month \$ |        |          | Prior Month YTD \$ |        | Current Month Year to Date \$ |        |          | Total at Completion |        |              | Comments |
|--|----------------|--------|------------------|--------|----------|--------------------|--------|-------------------------------|--------|----------|---------------------|--------|--------------|----------|
|  | Actual         | Budget | Actual           | Budget | Variance | Actual             | Budget | Actual                        | Budget | Variance | Estimate            | Budget | \$ Remaining |          |
| <i>Content Development:</i>              |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Ancillary Development                    |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Art Procurement                          |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Forms Construction                       |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Item Development                         |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Permissions                              |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Translations                             |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Total Content Development</b>         |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <i>Labor Support:</i>                    |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Information Technology                   |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Program Management                       |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Psychometrics                            |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Quality Assurance                        |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Total Labor Support</b>               |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <i>Production/Mfg/Distribution:</i>      |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Accommodations                           |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Design & Composition                     |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Film, Prep & Plate                       |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Freight                                  |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Paper, Print & Binding                   |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Warehousing                              |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Total Production/Mfg/Distribution</b> |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <i>Receive/Scan/Edit/Report:</i>         |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Receiving/Document Staging               |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Scanning                                 |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Editing                                  |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Reporting                                |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Total Receive/Scan/Edit/Report</b>    |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Scoring</b>                           |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <i>Travel:</i>                           |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Development                              |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Program Management                       |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| Scoring                                  |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Total Travel</b>                      |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |
| <b>Total Costs</b>                       |                |        |                  |        |          |                    |        |                               |        |          |                     |        |              |          |

## Sample Consortium Cost and Unit Reporting Worksheet

**Budget vs. Actual Units - Current Month**

Fiscal Year Month:

Fiscal Year:

|  | Prior Month Units |        | Current Month Units |        |          | Prior Month YTD Units |        | Current Month Year to Date Units |        |          | Total at Completion |        |                 | Comments |
|--|-------------------|--------|---------------------|--------|----------|-----------------------|--------|----------------------------------|--------|----------|---------------------|--------|-----------------|----------|
|  | Actual            | Budget | Actual              | Budget | Variance | Actual                | Budget | Actual                           | Budget | Variance | Estimate            | Budget | Units Remaining |          |
| <i>Content Development:</i>              |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Ancillary Development                    |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Art Procurement                          |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Forms Construction                       |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Item Development                         |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Permissions                              |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Translations                             |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Total Content Development</b>         |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <i>Labor Support:</i>                    |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Information Technology                   |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Program Management                       |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Psychometrics                            |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Quality Assurance                        |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Total Labor Support</b>               |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <i>Production/Mfg/Distribution:</i>      |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Accommodations                           |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Design & Composition                     |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Film, Prep & Plate                       |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Freight                                  |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Paper, Print & Binding                   |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Warehousing                              |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Total Production/Mfg/Distribution</b> |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <i>Receive/Scan/Edit/Report:</i>         |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Receiving/Document Staging               |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Scanning                                 |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Editing                                  |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Reporting                                |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Total Receive/Scan/Edit/Report</b>    |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Scoring</b>                           |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <i>Travel:</i>                           |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Development                              |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Program Management                       |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| Scoring                                  |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Total Travel</b>                      |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |
| <b>Total Units</b>                       |                   |        |                     |        |          |                       |        |                                  |        |          |                     |        |                 |          |



### Sample Consortium Schedule Variance Reporting Worksheet

**Schedule Attainment - Current Month**

**Fiscal Year Month:**

**Fiscal Year:**

|  | Prior Month % Complete |        | Current Month % Complete |        |          | Prior Month YTD % Complete |        | Current Month YTD % Complete |        |          | Completion Date |           | Comments |
|--|------------------------|--------|--------------------------|--------|----------|----------------------------|--------|------------------------------|--------|----------|-----------------|-----------|----------|
|  | Actual                 | Budget | Actual                   | Budget | Variance | Actual                     | Budget | Actual                       | Budget | Variance | Estimated       | Scheduled |          |
| <i>Content Development:</i>              |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Ancillary Development                    |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Art Procurement                          |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Forms Construction                       |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Item Development                         |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Permissions                              |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Translations                             |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Total Content Development</b>         |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <i>Labor Support:</i>                    |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Information Technology                   |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Program Management                       |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Psychometrics                            |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Quality Assurance                        |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Total Labor Support</b>               |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <i>Production/Mfg/Distribution:</i>      |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Accommodations                           |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Design & Composition                     |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Film, Prep & Plate                       |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Freight                                  |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Paper, Print & Binding                   |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Warehousing                              |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Total Production/Mfg/Distribution</b> |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <i>Receive/Scan/Edit/Report:</i>         |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Receiving/Document Staging               |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Scanning                                 |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Editing                                  |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Reporting                                |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Total Receive/Scan/Edit/Report</b>    |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Scoring</b>                           |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <i>Travel:</i>                           |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Development                              |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Program Management                       |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| Scoring                                  |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Total Travel</b>                      |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |
| <b>Total % Complete</b>                  |                        |        |                          |        |          |                            |        |                              |        |          |                 |           |          |

## **Appendix D**

### **About the Assessment Solutions Group**

ASG has assembled a skilled and knowledgeable team. The team includes Dr. John Olson, Dr. Edward Roeber, Mr. Barry Topol, CPA, and Mr. Patrick Hennon. This team has extensive expertise in educational assessment and measurement, assessment costing, finance and budgeting, fiscal and data management, content standards development, and assessment of students with special needs, such as special education and English learners, as well as management of state consortia. Team members have experience working with the USED, CCSSO, and SCASS, as well as many different state education agencies, districts, and schools. ASG has developed a comprehensive assessment cost model that can be used to determine the appropriate price for any assessment. The model has hundreds of variables, over 70 data tables, and is used extensively in ASG's assessment cost work.

Over the past 20 years, ASG team members have successfully worked with states on a wide variety of cost-related projects in the areas of budget analysis, and comprehensive cost evaluations of state assessment systems. For example, team members have worked in state agencies, at contractors who work with state agencies, on external design committees in several states, on technical advisory committees in several states, on external evaluation groups for state assessment programs, have developed numerous RFPs, written responses to even more numerous RFPs, and have developed and reviewed assessment costs on more than 100 state RFP responses.

Most recently, the ASG team has worked with states, CCSSO, and Stanford University, the latter to carry out a study to estimate the costs of new, higher quality assessments, and then to determine if whether through a series of cost savings measures, such a program could be made affordable. This work directly led to activities funded by the Gates Foundation to provide cost estimation services for the consortia of states organized in response to the RTTT common assessment competition. This report also was funded by the Gates Foundation.